



OPEN

Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national European cohort of 5594 patients

Espen Jimenez-Solem^{1,13,15}, Tonny S. Petersen^{1,13}, Casper Hansen², Christian Hansen², Christina Lioma², Christian Igel², Wouter Boomsma², Oswin Krause², Stephan Lorenzen², Raghavendra Selvan², Janne Petersen^{11,12,15}, Martin Erik Nyeland¹, Mikkel Zöllner Ankarfeldt^{11,15}, Gert Mehl Virenfeldt¹¹, Matilde Winther-Jensen¹¹, Allan Linneberg¹¹, Mostafa Mehdipour Ghazi², Nicki Detlefsen^{2,3}, Andreas David Lauritzen², Abraham George Smith², Marleen de Bruijne^{2,14}, Bulat Ibragimov², Jens Petersen², Martin Lillholm², Jon Middleton², Stine Hasling Mogensen⁴, Hans-Christian Thorsen-Meyer⁵, Anders Perner⁵, Marie Helleberg⁶, Benjamin Skov Kaas-Hansen⁷, Mikkel Bonde¹⁰, Alexander Bonde^{9,10}, Akshay Pai^{2,8}, Mads Nielsen² & Martin Silleesen^{9,10,13}✉

Patients with severe COVID-19 have overwhelmed healthcare systems worldwide. We hypothesized that machine learning (ML) models could be used to predict risks at different stages of management and thereby provide insights into drivers and prognostic markers of disease progression and death. From a cohort of approx. 2.6 million citizens in Denmark, SARS-CoV-2 PCR tests were performed on subjects suspected for COVID-19 disease; 3944 cases had at least one positive test and were subjected to further analysis. SARS-CoV-2 positive cases from the United Kingdom Biobank was used for external validation. The ML models predicted the risk of death (Receiver Operation Characteristics—Area Under the Curve, ROC-AUC) of 0.906 at diagnosis, 0.818, at hospital admission and 0.721 at Intensive Care Unit (ICU) admission. Similar metrics were achieved for predicted risks of hospital and ICU admission and use of mechanical ventilation. Common risk factors, included age, body mass index and hypertension, although the top risk features shifted towards markers of shock and organ dysfunction in ICU patients. The external validation indicated fair predictive performance for mortality prediction, but suboptimal performance for predicting ICU admission. ML may be used to identify

¹Department of Clinical Pharmacology, Copenhagen University Hospital, Bispebjerg and Frederiksberg, Copenhagen, Denmark. ²Department of Computer Science, University of Copenhagen, Copenhagen, Denmark. ³DTU Compute, Denmark's Technical University, Lyngby, Denmark. ⁴Danish Medicines Agency, Copenhagen, Denmark. ⁵Department of Intensive Care Medicine, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. ⁶Department of Infectious Diseases, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. ⁷Clinical Pharmacology Unit, Zealand University Hospital, Roskilde, Denmark. ⁸Cerebriu A/S, Copenhagen, Denmark. ⁹Department of Surgical Gastroenterology, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark. ¹⁰Center for Surgical Translational and Artificial Intelligence Research (CSTAR), Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. ¹¹Center for Clinical Research and Prevention, Copenhagen University Hospital, Bispebjerg and Frederiksberg, Copenhagen, Denmark. ¹²Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. ¹³Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. ¹⁴Department of Radiology and Nuclear Medicine, Erasmus MC – University Medical Center Rotterdam, Rotterdam, The Netherlands. ¹⁵Copenhagen Phase IV Unit (Phase4CPH), Department of Clinical Pharmacology and Center for Clinical Research and Prevention, Copenhagen University Hospital, Bispebjerg and Frederiksberg, Copenhagen, Denmark. ✉email: Martin.Silleesen@Regionh.dk

drivers of progression to more severe disease and for prognostication patients in patients with COVID-19. We provide access to an online risk calculator based on these findings.

The COVID-19 pandemic has put severe strains on hospital systems around the world and as of October 1st, 2020, the World Health Organization (WHO) estimated that more than 34 million patients are affected worldwide, and that the pandemic is the direct cause of more than 1 million deaths—a number that will likely rise as the pandemic progresses.

The unknown clinical features coupled with the speed of viral spreading creates an unfortunate situation where health care providers are lacking important diagnostic adjuncts, such as accurate prediction models and data-driven insights into the drivers of disease progression. Such data-driven tools are not only pivotal for health care providers in terms of risk prediction. They would also serve an important role for policy makers in terms of resource allocation strategies. Finally, with the emergence of vaccines with limited initial availability, such tools could assist in identifying high-risk patients for first-tier vaccination efforts.

Several studies have now proposed prediction models based on a variety of clinical features^{1–3}, indicating good predictive ability of machine learning (ML) models, including combinations of Support Vector Machines, Logistic Regression, Gradient Boosted Decision trees, Neural Networks and Random Forests, on mortality prediction^{3,4}. Tasks such as respiratory decompensation⁵, X-ray and clinical feature detection^{6,7} and SARS-CoV-2 detection optimisation^{8,9} have also been addressed. The majority of these are, however, trained and validated on national datasets from hospital admitted COVID-19 patients. While these may be of local value, whether the classification ability transfers to other healthcare systems is questionable.

As such, a recent review of ML models¹⁰, aimed at risk prediction in SARS-CoV-2 positive patients found that the majority of studies were constructed on Chinese data with a high risk of bias as assessed by the Prediction model Risk of Bias Assessment tool (PROBAST)¹¹. Furthermore, models often utilize data from hospital admitted SARS-CoV-2 positive patients only³, which may skew results due to the lack of data from patients with milder disease trajectories. Finally, models are often developed for predictions at hospital admission^{1–3}, providing little insight into the effects of in-hospital management.

This, in turn, hinders identification of optimal biomarkers and prognostic features of adverse outcomes, as these may change as the patient advances through the trajectory of the disease and the objective of this study was thus to construct and validate an ML model for SARS-CoV-2 adverse outcome risk prediction on a European dataset from Denmark, with external validation in a United Kingdom (UK) dataset. Secondly, we sought to use the constructed models for identification of disease risk factors, including comorbidities, biomarkers and vital signs as the patient moves through the disease trajectory, thus enabling the identification of important clinical predictors of adverse outcomes at four timepoints or timeframes, based on accumulating available data: On diagnosis, on hospital admission (where applicable) and immediately before and after admission to the intensive care unit (ICU), where applicable. Finally, we sought to create an online prediction tool to support rapid risk assessment upon COVID-19 diagnosis based on the most relevant data points.

We hypothesized that ML could be leveraged to provide accurate outcome predictions for COVID-19 patients, and that including accumulating data points from available sources from Electronic Health Record (EHR) repositories in a combined model could improve risk prediction as well as identification of relevant disease drivers at specific time points.

Methods

All methods were carried out in accordance with relevant guidelines and regulations. The study was approved by the relevant legal and ethics boards, including the Danish Patient Safety Authority (Styrelsen for Patientsikkerhed, approval #31-1521-257) and the Danish Data Protection Agency (Datatilsynet, approval #P-2020-320) as well as the UK Biobank (Application ID #60861) COVID-19 cohort. Under Danish law, approval from these agencies are required for access to and handling of patient sensitive data, including EHR records, whereas legal approval for the study was furthermore obtained from the Danish Capital Region (Region Hovedstaden). Patients from the UK biobank have provided informed consent prior to enrolment in the biobank. Under Danish law, informed consent for patient chart access for research purposes can be waived, provided approval from the Danish Patient Safety Authority (see approval number above) is obtained prior to data access.

We conducted a prospective study by including all individuals undergoing a SARS-CoV-2 test (nasal and/or pharyngeal swab subjected to Real-Time Polymerase Chain Reaction testing) in the Capital and Zealand Regions (approximately 2.6 million citizens) of Denmark between March 1st, 2020 and June 16th 2020. Data inclusion was censored on June 16th. Patients were identified through their Central Person Registry (CPR) number, a unique numerical combination given to every Danish citizen, enabling linking of electronic health records (EHRs) with nationwide medical registry data.

During the study period, all SARS-CoV-2 tests were performed at regional hospitals, with patients referred for testing based on presence of symptoms albeit test strategies shifted towards the end of the inclusion period to include a wider screening indication.

For cases with at least one positive tests, we extracted data from the bi-regional EHR system, including demographics, comorbidities and prescription medication. In-hospital data included laboratory results and vital signs.

Supplementary Table S1 lists extracted comorbidities with their definitions, Supplementary Table S2 extracted laboratory values, and Supplementary Table S3 extracted temporal features (vital signs).

For the purpose of external validation of the ML models, we extracted data from the UK biobank COVID-19 cohort. The UK biobank contains detailed healthcare information on 500,000 UK citizens, of which 1650 have been tested SARS-CoV-2 positive. This cohort has recently been made available for the purpose of COVID-19

research by the UK biobank consortium¹², and contain COVID-19 diagnostic test data, in hospital data as well as general practitioner data. For further information on the UK biobank COVID-19 cohort, please see Ref.¹³.

The rationale behind choosing the UK biobank dataset as an external validation cohort was primarily the need for model validation on an international dataset from a comparable health care system. The specific choice of the UK biobank was due to the curated and high-quality nature of this dataset, encompassing data from 500,000 UK citizens.

Prediction models. ML models were trained and validated on the Danish dataset. A subset of models sharing identical data fields (e.g. age, comorbidities etc.) between the Danish and UK cohorts were subsequently externally validated on the UK biobank dataset.

We constructed ML prediction models by including available data for patients up to and including the selected time frames or time points. These time frames or points were.

- At time of SARS-CoV-2 positivity (all patients, Diagnosis model)
- The first 12 h of hospital admission (Admission model)
- 12 h up to ICU admission (Pre-ICU model)
- 12 h after ICU admission (Post-ICU model).

Models were trained to predict one of four events, where applicable:

- Hospital admission (SARS-CoV-2 positive patients)
- ICU admission (Diagnosed and hospital admitted patients)
- Mechanical ventilation (Diagnosed, hospital and ICU admitted patients)
- Death (all patients)

For each task, we trained with different feature sets to study how incrementally adding data affects model performance as well as to gain insight into drivers of disease progression:

- Base models: Age, sex and body mass index (BMI).
- Comorbidities: Base model and comorbidities (Table 1 and Supplementary Table S1)
- Temporal features: Comorbidities model and temporal features (Supplementary Table S3)
- In-hospital laboratory tests: Temporal features model and in-hospital lab tests (Supplementary Table S2).

For the purpose of external validation, data points were available in the UK biobank matching those of the base and comorbidities models. In-hospital models could not be externally validated due to lack of availability of these data points in the UK biobank.

ML models. We used *random forests (RFs)*¹⁴, implemented in the open-source machine learning library *scikit-learn*¹⁵. Because each individual tree was trained on a bootstrap sample, there were out-of-bag (OOB) samples from the training data that could be used to estimate the performance of the RF, a method considered superior to nested cross-validation¹⁶.

All models were evaluated on the Danish set using fivefold cross-validation. The folds were stratified to ensure that the splits were representative of the full cohort. For each split, we conducted grid search on the available training data fold to tune the hyperparameters of the RF models for each prediction task. As selection criterion, we computed the *Receiver Operating Characteristics Area Under the Curve (ROC-AUC)* on the OOB samples. Each model used a 1000 decision trees, while we varied the maximum number of features considered in each split (all features or square root of all features) and either the maximum depth (5, 10 or unlimited) or the minimum samples for a split (2, 5 or 10). For the full model evaluation, we combine the outputs for all test folds resulting in predictions for the entire data set on which we report the ROC-AUC and the precision/recall AUC (PR-AUC).

For evaluation on the UK data, models were trained on the entire Danish data set. As before, for each model, a grid search based on the OOB ROC-AUC was performed on the same parameter grid. Each model was then evaluated on the entire UK cohort.

For each task, we also applied standard logistic regression as a baseline model, but leave out the results here, since logistic regression exhibited sub-par or equal performance compared to the RF in most cases, the only exception being prediction of death when not including in-hospital tests.

Post-hoc analysis of the use of the predictive variables across all decision trees in the RF allowed us to derive a measure of feature importance. Feature importance was calculated by the *mean decrease in impurity (MDI)*. This measure considers how often a feature is used when classifying the training data points and how well it splits the training data points when being used. The predictive variable importance was computed for the models trained on the entire data set. The top-10 or top-20 (depending on the model) were extracted and visualized. The correlations of these features were then computed across the entire dataset.

Missing data. Missing data was considered missing at random.

Percentages of available data in the Danish cohort are presented in supplementary Table S4. Missing values for BMI were imputed by using k-nearest neighbour imputation using age and sex¹⁷, with k = 100. Other missing

	All SARS-CoV-2 patients (n = 3944)	Non-hospitalized (n = 2585)	Hospitalized patients (n = 1359)	Hospitalized patients without ICU admission (n = 1178)	Hospitalized patients with ICU admission (n = 181)	Survivors (n = 3620)	Non-survivors (n = 324)
Body Mass Index	25.9 (22.7–29.8)	25.8 (22.6–29.7)	26.0 (22.7–29.9)	25.7 (22.6–29.4)	27.3 (23.5–31.4)	26.0 (22.7–29.9)	25.2 (22.0–28.8)
Age	53.0 (36.0–69.0)	44.0 (31.0–58.0)	70.0 (55.0–80.0)**	71.0 (54.0–81.0)	69.0 (58.0–75.0)	50.0 (34.0–64.0)	81.0 (73.0–87.0)**
Male Sex	41.9	35.6	53.9**	51.4	70.2**	40.6	56.2**
Diabetes	13.4	8.0	23.5**	23.4	24.3	12.0	28.7**
Ischemic heart disease	9.1	5.2	16.5**	16.0	19.3	8.0	21.0**
Heart failure	4.2	1.5	9.4**	9.5	8.8	3.1	16.7**
Arrhythmia	10.4	4.9	20.9**	21.0	20.4	8.5	32.1**
Stroke	6.7	3.1	13.5**	14.1	9.9	5.7	17.9**
COPD or asthma	11.4	9.2	15.7**	15.4	17.1	10.9	17.0**
Sleep apnoea	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Arthritis	1.6	1.4	2.2	2.1	2.8	1.5	3.1
Osteoporosis	6.0	2.9	12.1**	12.0	12.7	4.9	18.5**
Dementia	2.9	1.3	5.9**	6.7	0.6**	1.8	14.8**
Severe mental disorder	1.7	1.5	2.0	2.1	1.1	1.5	4.3**
Immunodeficiencies	0.4	0.3	0.5	0.5	0.6	0.4	0.6
Neurological manifestations	18.1	15.0	24.1**	25.1	17.1	16.8	33.0**
Cancer	10.3	6.3	18.0**	17.8	18.8	8.8	27.5**
Chronic kidney failure	2.4	0.7	5.7**	5.6	6.6	1.8	9.6**
Dialysis	0.5	0.1	1.3**	1.3	1.1	0.4	1.5
Hypertension	30.1	18.3	52.5**	48.6	78.5**	27.0	65.1**

Table 1. Demographic information on the group of SARS-CoV-2 positive patients, including information on pre-existing comorbidities. Supplementary Table S2 holds information on diagnoses codes included in the individual comorbidity classifications. The table presents information on the full cohort (admitted and non-admitted SARS-CoV-2 positive patients) as well as subgroups admitted to a hospital and Intensive Care Unit (ICU) respectively. Furthermore, differential demographics between survivors and non-survivors (in-hospital mortality) is presented. Continuous variables are presented as medians with (interquartile range). *COPD* chronic obstructive pulmonary disease. ** $p < 0.001$ when subgroups are compared (e.g. hospitalized vs. non-hospitalized, ICU vs. non-ICU, survivors vs. non-survivors).

data points were set to “not available” for the purpose of ML modelling and deleted by case wise deletion for group comparisons.

Data presentation and statistical testing. Continuous data is presented as medians (interquartile range) and compared using the Mann–Whitney U test. Categorical data is presented as percentages and compared using the Chi-square test.

ML model performances are presented as ROC-AUC for Positive and Negative predictive value and precision/recall. Model comparisons were performed by the deLong test¹⁸.

The p-values for comparisons of outcome groups are provided for reference only. As these comparisons are not part of the study hypotheses, p values are presented without post-hoc correction for multiple testing and should be interpreted as such.

In addition, calibration curves are presented for the combined test folds and the external validation data. For each calibration plot, the predictions were grouped using quantile-based binning.

Online models. The risk prediction model for SARS-CoV-2 positive patients admitted to hospital is available in an online version on <https://cope.science.ku.dk>.

Results

A total of 3944 individuals had at least one positive SARS-CoV-2 test in the two Danish regions and were included in the study. These were supplemented by the 1650 patients from the UK biobank used for external model validation. Figure 1 depicts patient identification and selection in a flowchart form for the Danish cohort.

Among the Danish cases, 1359 (34.5%) required hospitalization, and 181 (4.6%) intensive care. A total of 324 patients (8.2%) died.

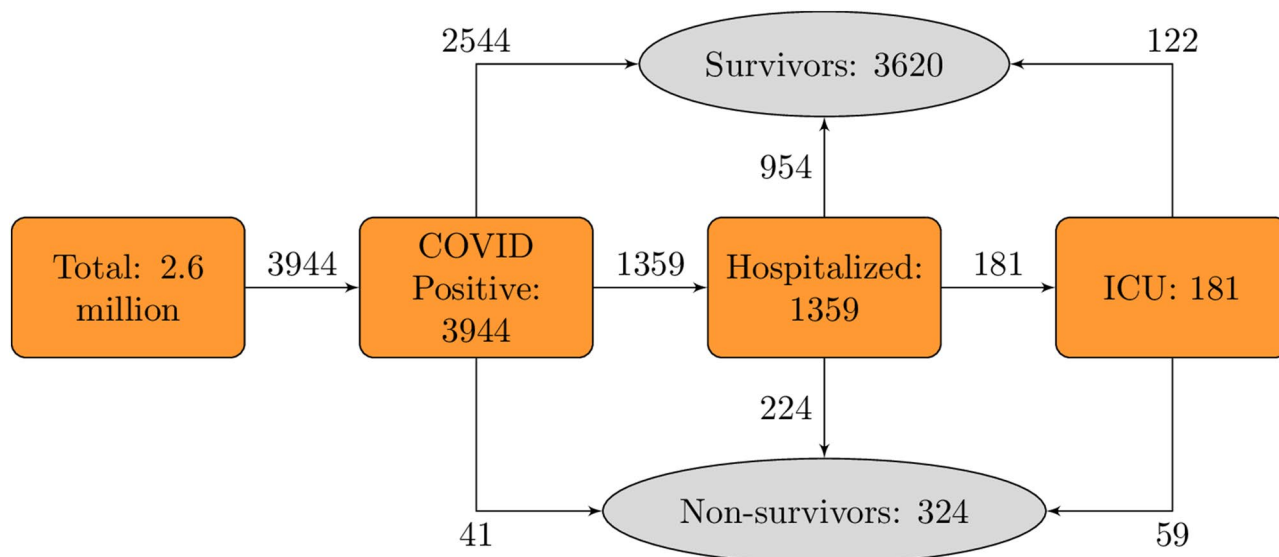


Figure 1. Flowchart of patient selection and identification for the Danish cohort. ICU intensive care unit.

	Unit	All SARS-CoV-2 patients (n = 3944)	Non-hospitalized (n = 2585)	Hospitalized patients (n = 1359)	Hospitalized patients without ICU admission (n = 1178)	Hospitalized patients with ICU admission (n = 181)	Survivors (n = 3620)	Non-survivors (n = 324)
CRP	mg/l	65.0 (29.0–122.0)	21.0 (8.7–46.5)	66.0 (30.0–124.0)**	58.0 (26.0–110.0)	142.0 (86.8–231.0)**	59.5 (24.3–119.0)	79.5 (42.0–160.0)**
Lymphocyte count	10 ⁹ /l	0.9 (0.6–1.4)	1.2 (0.8–1.7)	0.9 (0.6–1.3)**	10.0 (0.7–1.4)	0.8 (0.5–1.1)	1.0 (0.7–1.4)	0.8 (0.6–1.1)**
Lactic dehydrogenase	U/l	274.5 (210.0–376.8)	215.5 (176.5–234.3)	277.5 (214.0–381.5)**	263.0 (206.0–342.5)	430.0 (323.0–579.0)**	272.0 (207.0–366.5)	292.0 (226.0–411.5)**
Alanine aminotransferase	U/l	27.0 (18.0–45.0)	26.0 (17.5–35.0)	27.5 (18.0–46.0)**	26.0 (18.0–43.0)	36.0 (23.3–62.0)**	28.0 (19.0–46.3)	25.0 (17.0–41.0)**
Hemoglobin	mmol/l	8.0 (7.1–8.7)	8.3 (7.6–8.9)	8.0 (7.1–8.7)**	8.0 (7.1–8.8)	7.6 (6.9–8.4)	8.1 (7.2–8.8)	7.7 (6.7–8.4)**
White blood cells	10 ⁹ /l	6.8 (5.1–9.4)	6.3 (4.7–7.3)	6.8 (5.1–9.5)**	6.8 (5.1–9.3)	7.9 (5.3–10.6)	6.7 (5.0–9.0)	7.4 (5.6–10.7)**
Neutrophil count	10 ⁹ /l	5.0 (3.5–7.4)	4.15 (3.3–5.2)	5.0 (3.5–7.5)**	4.9 (3.5–7.1)	6.1 (4.2–8.8)	4.8 (3.4–6.9)	5.7 (4.1–8.7)**
D dimer	mg/l	1.0 (0.6–2.4)	0.7 (0.7–1.9)	1.0 (0.6–2.4)**	0.9 (0.5–2.1)	1.4 (0.8–4.3)**	0.9 (0.5–2.1)	1.8 (0.8–4.4)**
Blood urea nitrogen	mmol/l	6.3 (4.3–9.9)	4.3 (3.8–6.7)	6.4 (4.4–10.0)**	6.3 (4.4–9.9)	6.8 (5.1–10.8)	5.7 (4.1–8.4)	9.8 (6.8–14.8)**
Creatinine	umol/l	81.5 (65.6–108.0)	67.0 (58.0–85.0)	83.0 (65.0–109.0)**	82.0 (65.0–107.0)	85.5 (67.0–111.3)	79.0 (63.0–100.0)	99.0 (75.6–149.0)**
Ferritin	ug/l	557.0 (203.0–1125.0)	167.0 (99.0–613.0)	570.0 (219.5–1145.0)**	399.0 (176.3–913.3)	1100.0 (741.0–1910.0)**	516.0 (204.0–1140.0)	649.0 (201.0–1110.0)**
Base excess	mmol/l	0.8 (–1.4–3.0)	0.3 (–0.9–2.7)	0.8 (–1.4–3.1)**	0.7 (–1.4–3.0)	1.0 (–1.7–3.6)	0.8 (–1.1–3.1)	0.6 (–2.8–3.0)
HCO ₃	mEq/l	25.2 (23.6–27.0)	25.0 (24.3–26.8)	25.2 (23.6–27.0)**	25.2 (23.6–26.9)	25.4 (23.6–27.4)**	25.3 (24.0–27.0)	24.9 (22.2–27.0)**
Lactate	mmol/l	1.1 (0.8–1.5)	1.2 (0.8–1.8)	1.1 (0.8–1.5)**	1.0 (0.8–1.5)	1.3 (1.0–1.7)**	1.0 (0.8–1.4)	1.3 (0.9–1.7)**
Arterial O ₂ saturation	%	0.94 (0.92–0.99)	0.95 (0.92–0.97)	0.94 (0.92–0.97)**	0.95 (0.92–0.97)	0.92 (0.91–0.94)**	0.95 (0.92–0.97)	0.93 (0.92–0.96)**
pCO ₂	mmHg	4.6 (4.1–5.1)	4.7 (4.4–5.0)	4.6 (4.1–5.1)**	4.6 (4.1–5.1)	4.7 (4.1–5.2)**	4.6 (4.1–5.1)	4.7 (4.0–5.2)**
pH	–	7.46 (7.43–7.49)	7.47 (7.43–7.49)	7.46 (7.43–7.49)**	7.46 (7.43–7.49)	7.47 (7.42–7.49)**	7.47 (7.43–7.49)	7.45 (7.41–7.49)**
pO ₂	mmHg	9.1 (8.0–10.9)	10.0 (7.9–10.7)	9.1 (8.0–10.9)**	9.4 (8.2–11.1)	8.3 (7.5–9.6)**	9.3 (8.2–11.1)	8.8 (7.8–10.2)**
Pulse	/min	83.0 (72.0–94.0)	86.0 (74.0–102.25)	83.0 (72.0–93.0)**	82.0 (71.0–92.0)	86.0 (76.0–97.0)	82.0 (72.0–93.0)	85.0 (72.8–98.3)**
Temperature	C	37.4 (36.8–38.1)	37.3 (36.6–38.0)	37.4 (36.8–38.1)**	37.3 (36.8–38.0)	37.6 (37.1–38.3)	37.4 (36.8–38.1)	37.3 (36.7–38.0)**
Early warning score	–	3.0 (2.0–6.0)	2.0 (1.0–4.0)	4.0 (2.0–6.0)**	3.0 (2.0–5.0)	6.0 (5.0–8.0)**	3.0 (2.0–5.0)	5.0 (3.0–6.0)**
Respiratory rate	/min	20.0 (17.0–22.0)	18.0 (16.0–20.0)	20.0 (18.0–22.0)**	20.0 (17.0–21.0)	24.0 (20.0–30.0)**	20.0 (17.0–22.0)	20.0 (18.0–24.0)**
Saturation	%	96.0 (94.0–98.0)	98.0 (96.5–99.0)	96.0 (94.0–97.0)**	96.0 (95.0–98.0)	94.0 (92.0–95.0)	96.0 (95.0–98.0)	95.0 (93.0–97.0)**
Is smoking	%	4.8	0.5	13.1**	12.1	19.3	3.4	21.3**

Table 2. Temporal features (laboratory tests and vital signs). The table presents information on the full cohort (admitted and non-admitted SARS-CoV-2 positive patients) as well as subgroups admitted to a hospital and Intensive Care Unit (ICU) respectively. Furthermore, differential temporal features between survivors and non-survivors (in-hospital mortality) is presented. Continuous variables are presented as medians with (interquartile range). ***p* < 0.001 when subgroups are compared (e.g. hospitalized vs. non-hospitalized, ICU vs. non-ICU, survivors vs. non-survivors).

	Hospital admission		ICU admission		Ventilator treatment		Death	
	TPR/FPR	Pre/Rec	TPR/FPR	Pre/Rec	TPR/FPR	Pre/Rec	TPR/FPR	Pre/Rec
Diagnosis								
Age + Gender + BMI	0.820	0.705	0.802	0.173	0.815	0.184	0.902	0.412
+Comorbidities	0.822	0.705	0.844*	0.206	0.851*	0.192	0.906	0.412
+Temporal Features	–	–	–	–	–	–	–	–
+In-hospital Tests	–	–	–	–	–	–	–	–
Admission								
Age + Gender + BMI	–	–	0.685	0.226	0.675	0.200	0.785	0.435
+Comorbidities	–	–	0.752*	0.282	0.743*	0.238	0.794	0.445
+Temporal Features	–	–	0.763*	0.308	0.762*	0.289	0.796	0.444
+In-hospital Tests	–	–	0.805*#	0.418	0.786*	0.345	0.818*	0.540
Pre-ICU								
Age + Gender + BMI	–	–	–	–	0.598	0.892	0.733	0.575
+Comorbidities	–	–	–	–	0.567	0.869	0.735	0.548
+Temporal Features	–	–	–	–	0.563	0.871	0.738	0.567
+In-hospital Tests	–	–	–	–	0.502	0.867	0.721	0.567
Post-ICU								
Age + Gender + BMI	–	–	–	–	0.598	0.892	0.733	0.575
+Comorbidities	–	–	–	–	0.530	0.843	0.724	0.552
+Temporal Features	–	–	–	–	0.584	0.861	0.739	0.569
+In-hospital Tests	–	–	–	–	0.671	0.928	0.741	0.568

Table 3. Main results from the prediction models. Predictions were performed with data available from four different time frames in the patient disease trajectories (left column): On diagnosis (Diagnoses model), On hospital admission and 12-h into admission (Admission model), 12 h leading up to Intensive Care Unit (ICU) admission (Pre-ICU model) and 12 h after ICU admission (post-ICU model). Models were trained to predict risk of hospital admission, ICU admission, ventilator treatment and death (top row). All models were trained with incremental data, starting with age, gender and Body Mass Index, then adding comorbidity information, temporal features (e.g. vital signs) and finally by adding hospital laboratory tests where applicable. Please see supplementary tables S1 and S2 for data definitions. Performance metrics are presented as the Receiver Operating Characteristics Area Under the Curve (ROC-AUC) for True/False positive rates (TPR/FPR) and Precision/Recall (Pre/Rec). *Model is significantly ($p < 0.01$) better than the base prediction model (Age + gender + Body Mass Index, BMI). #Model is significantly ($p < 0.01$) better than the comorbidities model. §Model is significantly ($p < 0.01$) better than the temporal model. --: Insufficient data available at the time point, or prediction irrelevant (e.g. predicting hospital admission for patients already in the ICU).

Demographics and comorbidities are summarized in Table 1, selected laboratory values and vital signs in Table 2. Demographics information for the UK biobank external validation cohort is presented in supplementary Table S5.

When compared to non-hospitalized patients, hospital admitted patients were older and more likely to be male and a number of comorbidities were overrepresented in the admitted subgroup. These included hypertension, diabetes, ischemic heart disease, heart failure, arrhythmias, stroke, chronic obstructive pulmonary disease (COPD) or asthma, osteoporosis, neurological disease, cancer, chronic kidney failure and use of dialysis. Hospitalized patients were more likely to be smokers (Table 2).

For hospitalized patients requiring ICU admission vs. hospitalized patients without ICU admission, only male sex, Body Mass Index (BMI), dementia and hypertension differed between patients and ICU-admitted patients were furthermore more likely to be smokers, older and male (Table 1).

Non-survivors were furthermore more likely to suffer from hypertension, diabetes, ischemic heart disease, heart failure, arrhythmias, stroke, COPD or asthma, osteoporosis, dementia, mental disorders, neurological disease, cancer, chronic kidney failure and use of dialysis.

When compared to non-admitted, admitted patients differed significantly in all measured values (Table 2). Among those hospitalised, those admitted to the ICU had derangements in many variables (Table 2). The same was observed for non-survivors compared with survivors (Table 2).

ML models prediction. ML models are presented in Table 3 and graphically depicted in supplementary Fig. S1 (Diagnosis model), Supplementary Fig. S2 (Admission model), Supplementary Fig. S3 (Pre-ICU model) and Supplementary Fig. S4 (Post-ICU model).

Base models deployed on the time of diagnosis were able to predict hospital admission with a ROC-AUC of 0.820, ICU admission 0.802, ventilator treatment 0.815 and death 0.902 (Table 3 and Supplementary Fig. S1).

Adding information on patient comorbidities increased the predictive ability for all outcomes.

Models deployed at hospital admission achieved ROC-AUC scores ranging from 0.675 to 0.818 for the selected outcomes (Table 3 and Supplementary Fig. S2). Adding information on comorbidities, temporal features

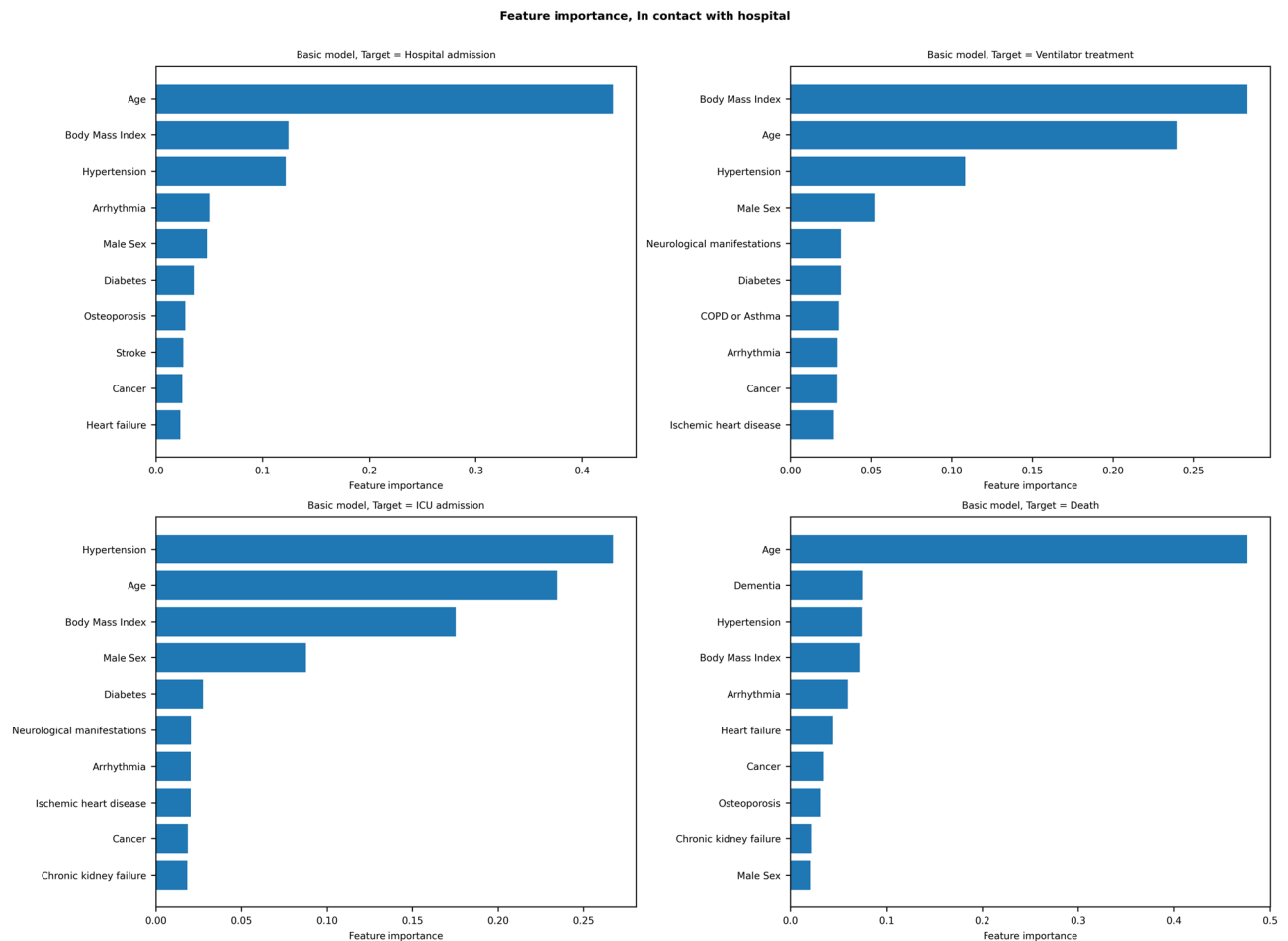


Figure 2. Feature importance for the basic (including age, sex, body mass index, and comorbidities) diagnosis models, predicting risk of intensive care admission (first row), hospital admission (second row), ventilator treatment (third row) and death (fourth row) on SARS-CoV-2 positivity.

and hospital laboratory tests increased model performance for ICU admission, use of mechanical ventilation and death.

Models deployed pre- or post-ICU admission achieved ROC-AUC's from 0.502 to 0.741 (Table 3 and Supplementary Figs. S3 and S4).

The calibration curves (supplementary Figs. S7 and S8) show that the models are well calibrated when looking at all diagnosed subjects and at patients admitted to the hospital. When restricted to patients admitted to ICU, the calibration gets worse as expected due to smaller sample size. Ventilator treatment could not be predicted accurately, and the calibration curves reflect this.

External validation results (Supplementary Table S6) on UK data indicated an overall reduction in model classification ability. For diagnosed patients, ROC-AUCs were 0.661 for predicting hospital admission, 0.529 for predicting ICU admission and 0.742 for predicting mortality. Inspection of the calibration curves (Supplementary Fig. S7) shows that the models are only slightly worse calibrated for the UK data, meaning that the model outputs approximately the correct probability for individual patients, despite the degradation in ROC-AUC.

As patients progressed through the disease severity trajectories, mortality prediction remained in the area of 0.617–0.722 (Supplementary Table S6).

Detection of important features and drivers of disease progression. Results of the drivers of disease progression feature detection analysis for each of the selected timepoints are depicted in Fig. 2 (diagnoses model) and Fig. 3 (admission model) as well as Supplementary Fig. S4 (pre-ICU model) and Supplementary Fig. S5 (post-ICU model).

For diagnosed patients (Fig. 2), age and BMI were among the most relevant features for all targets, and indeed the most important features for predicting hospital admission and ventilator treatment.

Hypertension was the most important feature for predicting ICU admission, and indeed an important feature for all models.

For admitted patients (Fig. 3) the most relevant drivers of disease progression were age, BMI, hypertension and the presence of dementia. When the full dataset was analysed, lab tests indicating aspects of cell dysfunction (Lactic dehydrogenase, LDH), kidney dysfunction (Blood urea nitrogen and Creatinine), the inflammatory

Feature importance, Admitted to hospital (+12h)

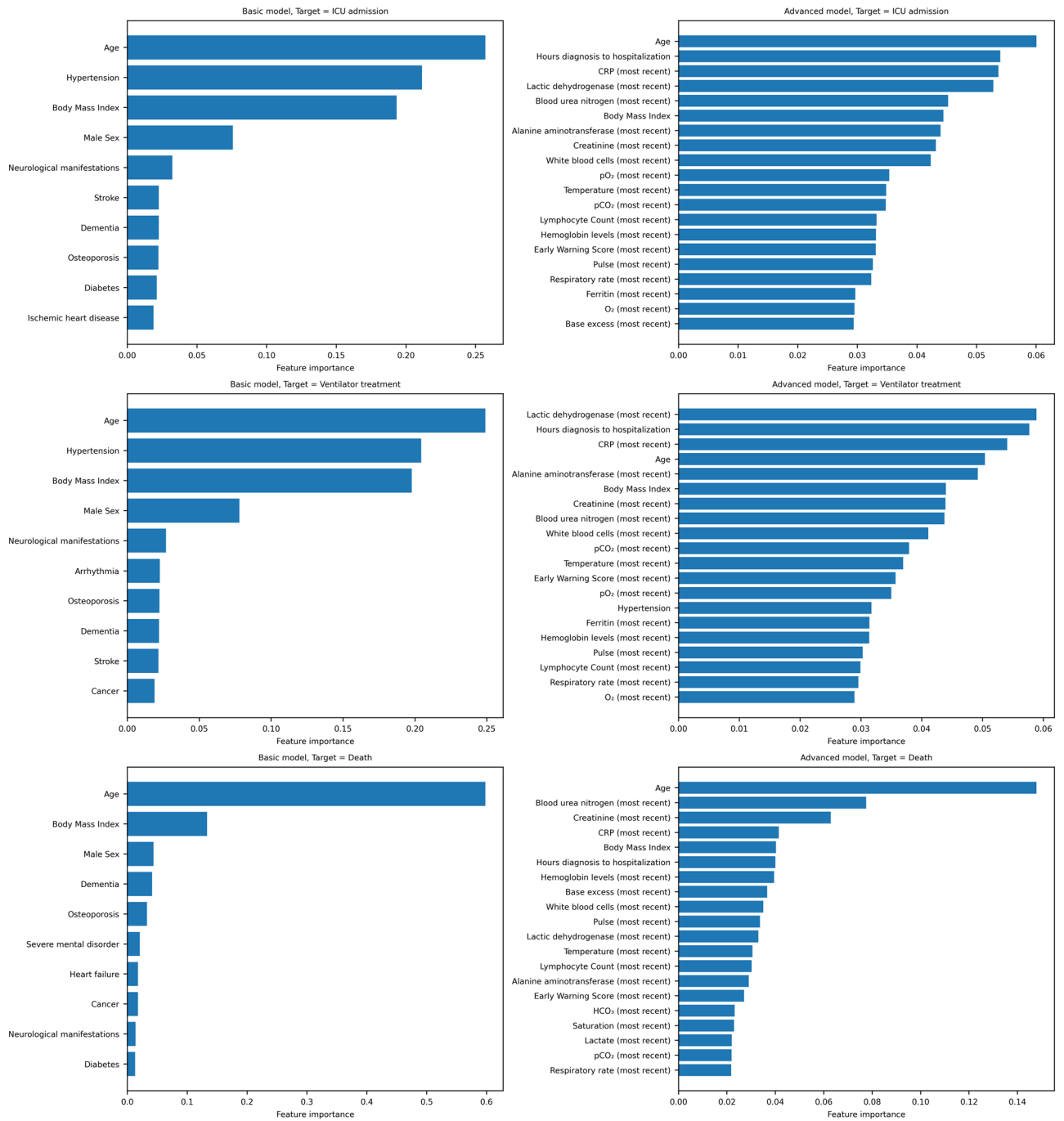


Figure 3. Feature importance for the basic (including age, sex, body mass index, and comorbidities) and advanced (all data) admission models, predicting risk of intensive care admission (first row), ventilator treatment (second row) and death (third row) on hospital admission.

response (C-reactive protein, CRP), and liver damage (Alanine Aminotransferase, ALT) were identified as important prognostic markers for disease progression.

For ICU models (Supplementary Figs. S5 and S6), relevant disease progression features again included age and BMI as well as comorbidities including hypertension, heart failure and neurological disease. When the full dataset was analysed, disease progression drivers included features reflecting insufficient respiration and metabolism (pulse oximetry oxygen saturation, pO₂ values and slopes over time), as well as BUN and CRP levels.

Discussion

In this study, we analyse prognostic and factors associated with disease progression in 3944 SARS-CoV-2 positive patients by constructing an interpretable ML framework. In contrast to previous studies^{1,2}, these included diagnosed patients outside hospitals, and thus included the entire spectrum of SARS-CoV-2 positive patients in the 2.6 million regional population.

Results indicate that by focusing on a limited number of demographic variables, including age, gender and BMI, it is possible to predict the risk of hospital and ICU admission, use of mechanical ventilation and death as early as at the time of diagnosis. Using these parameters only, our model achieved a ROC-AUC of 0.902 for mortality prediction, which is slightly inferior to a model reported by Gao et al. achieving a ROC-AUC of 0.962 using more complex clinical data points on admission³.

Adding information on comorbidities to the model increase performance, indicating that these features play a prognostic role in the outcome of patients as they progress through the disease trajectory.

As such, results from the ML feature detection indicate that comorbidity factors such as hypertension and diabetes are driving factors of adverse outcome, which is in line with reports from other cohort studies^{19–21}. The role of hypertension is further underlined by reports indicating a role of the angiotensin converting enzyme 2 (ACE2) receptor as an entry point for the SARS-CoV2²². Whether COVID-19 interacts unfavourably with hypertension per se, or whether this risk is simply a manifestation of reduced tolerance to severe infection and hypoxia is currently debated²³.

Furthermore, BMI was identified as a major feature of adverse outcome, as also reported by others^{24–26}. Whether this is due to a reduced respiratory capacity or chronic impairment of the immune system through alterations in tumour necrosis factor and interferon secretion associated with obesity, is also currently debated²⁷. Caution should, however, be taken when analysing these results, as the median observed differences between groups were minor and may not be clinically relevant. Furthermore, data imputation may have impacted on these results.

The addition of more data points, including temporal features and lab tests improved the model's predictive value for hospitalized patients, with group comparisons indicating alterations of a plethora of laboratory tests for admitted patients, including features of immune activation and organ dysfunction. Interestingly, laboratory tests differed to a lesser extent between ICU and non-ICU patients, except for CRP levels, lymphocyte counts, LDH, ALT, neutrophil, D-dimer and ferritin levels as well as arterial blood gas values. As expected, ICU admitted patients had lower oxygen saturation and higher respiratory rates, likely reflecting the acute respiratory distress from COVID-19 pneumonia.

Feature analysis indicated that strong prognostic markers expectedly included CRP levels, but also markers of organ damage, including kidney injury (creatinine and blood urea nitrogen), liver injury (ALAT), cell damage (LDH), anaemia (haemoglobin levels) as well as ferritin levels. These, as well as vital signs and arterial blood gas values superseded many of the comorbidities in feature importance once the patient progressed through hospital and ICU admission, which again indicates that drivers and prognostic markers of adverse outcomes represent a dynamic field affected by the patient's current point on the disease trajectory, and that differential values should be considered when risk-assessing COVID-19 patients depending on their current status (e.g. in hospital, in ICU etc.). A caveat is, however, that multiple comorbidities and advanced age may result in decisions by patients, relatives or clinicians limiting the use of life-support, and thus potentially precluding them from ICU admission and reducing the effect of comorbidities and age on model predictions.

Kidney injury has previously been reported in patients with COVID-19^{28,29} and our finding that markers of kidney injury may be important at hospital admission supports the notion that COVID-19 associated kidney injury plays an important pathophysiological role.

The importance of LDH for COVID-19 patients has previously been reported in other ML¹ as well as clinical studies³⁰ and these results are supported by the feature detection from this study, indicating that LDH levels serve as an important prognostic marker on hospital admission, although its value is superseded by other biomarkers when the patient advance to the ICU stage. As LDH can be seen as a general marker of cell and organ damage with a reported prognostic value for mortality in ICU patients³¹, these findings likely indicate a general organ affection associated with COVID-19 disease progression.

Abnormal liver function tests, including ALAT, has previously been associated with COVID-19 disease severity³² and these reports have indicated the presence of elevated liver enzymes in both severe and non-severe COVID-19 cases³³. Whether this is a function of viral infection, shock or a consequence of hepatotoxic pharmaceuticals deployed during treatment is still not clear³⁴.

Ferritin levels have previously been associated with COVID-19³⁵, presumably due to its role in immunomodulation and association with the cytokine storm response seen in critical illness³⁶.

Taken together, the feature importance of laboratory tests indicating affection of several organ systems indicates that COVID-19 disease severity follows a predictable pattern characterized by multi-organ affection (albeit not always dysfunction), which is in line with previous findings³⁷.

Once patients progress to the ICU stage, feature detection indicated a switch towards vital signs and biomarkers indicating that the severity of respiratory failure, shock and inflammatory markers were the most important features of risk of death (Supplementary Figs. S5 and S6).

When the feature importance of all models is analysed, the results indicate that COVID-19 outcomes are at the time of diagnoses largely predictable through a relatively limited number of features, dominated by age, BMI and comorbidities, effectively proxies for frailty.

As patients follow their disease trajectories, differential features supersede each other in prognostic importance, and prognostic models should thus consider the patient's place in the disease trajectory.

The results of the external validation did, however, show an overall reduction in the model's classification ability when the UK biobank cohort was analysed, thus impacting on the generalizability of the presented models, but results should be interpreted with caution.

As such, the UK cohort was assembled for the purpose of biobanking studies, and thus comprise a highly selected subset of patients, whereas the Danish cohort was population wide in the two analysed geographical regions. Demographic data also highlights differences in the two populations, including an age difference between groups. Actually, when predicting death for ICU patients, where demographics are similar, we do not observe a reduction in model performance.

The differences in results can be explained by the change of the underlying data distribution, demonstrating that caution should be exercised when evaluating whether ML models are useful for local health care practitioners if developed on other cohorts, especially when developed on early phase COVID-19 data. As such, significant variations in national factors such as isolation policies and triage for ICU and mechanical ventilation, population demographics etc. may impact on results. This notion is supported by the finding that our model retained reasonable classification ability for mortality in UK patients, but failed to predict ICU admission risk.

These results could thus indicate that potential users of ML models for COVID-19 patients should carefully examine the generalizability of the training cohort and healthcare infrastructure where patients originated from and compare these with local features prior to model usage.

Our study has several limitations. The number of patients available for this analysis was limited, and additional patient data could change the results. This is especially evident when performing predictions in the ICU setting, where the number of patients was limited. A larger and preferably multinational dataset would be required to address this issue.

Secondly, we have extracted a subset of clinical variables from the EHR system and analysing other features could affect the model. Furthermore, the changing criteria for SARS-CoV-2 testing associated with the course of the pandemic, likely also affects the results.

For external validation, our results are limited by the fact the UK biobank data did not offer datapoints allowing for external validation of advanced features models.

Even with these limitations, we may conclude that ML may be leveraged to perform outcome prediction in COVID-19 patients, as well as serve as a potential tool for identifying drivers and prognostic markers.

Data availability

Patient data from this study has not been made available to the public due to patient confidentiality constraints.

Received: 7 October 2020; Accepted: 12 January 2021

Published online: 05 February 2021

References

1. Yan, L. *et al.* An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.* **2**, 283–288. <https://doi.org/10.1038/s42256-020-0180-7> (2020).
2. An, C. *et al.* Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Sci. Rep.* **10**, 18716. <https://doi.org/10.1038/s41598-020-75767-2> (2020).
3. Gao, Y. *et al.* Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat. Commun.* **11**, 5033. <https://doi.org/10.1038/s41467-020-18684-2> (2020).
4. Abdulaal, A. *et al.* Prognostic modeling of COVID-19 using artificial intelligence in the United Kingdom: model development and validation. *J. Med. Internet Res.* **22**, e20259. <https://doi.org/10.2196/20259> (2020).
5. Burdick, H. *et al.* Prediction of respiratory decompensation in Covid-19 patients using machine learning: the READY trial. *Comput. Biol. Med.* **124**, 103949. <https://doi.org/10.1016/j.compbiomed.2020.103949> (2020).
6. Li, L. *et al.* Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology* <https://doi.org/10.1148/radiol.2020200905> (2020).
7. Izquierdo, J. L., Ancochea, J. & Soriano, J. B. Clinical characteristics and prognostic factors for intensive care unit admission of patients With COVID-19: retrospective study using machine learning and natural language processing. *J. Med. Internet Res.* **22**, e21801. <https://doi.org/10.2196/21801> (2020).
8. Cabitza, F. *et al.* Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. *Clin. Chem. Lab. Med.* <https://doi.org/10.1515/cclm-2020-1294> (2020).
9. Formica, V. *et al.* Complete blood count might help to identify subjects with high probability of testing positive to SARS-CoV-2. *Clin. Med. (Lond.)* **20**, e114–e119. <https://doi.org/10.7861/clinmed.2020-0373> (2020).
10. Wynants, L. *et al.* Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* **369**, m1328. <https://doi.org/10.1136/bmj.m1328> (2020).
11. Wolff, R. F. *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann. Intern. Med.* **170**, 51–58. <https://doi.org/10.7326/M18-1376> (2019).
12. Khanji, M. Y., Aung, N., Chahal, C. A. A. & Petersen, S. E. COVID-19 and the UK biobank—opportunities and challenges for research and collaboration with other large population studies. *Front. Cardiovasc. Med.* <https://doi.org/10.3389/fcvm.2020.00156> (2020).
13. Atkins, J. L. *et al.* Preexisting comorbidities predicting COVID-19 and mortality in the UK biobank community cohort. *J. Gerontol. A Biol. Sci. Med. Sci.* **75**, 2224–2230. <https://doi.org/10.1093/geron/glaa183> (2020).
14. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. <https://doi.org/10.1023/A:1010933404324> (2001).
15. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**(85), 2825–2830 (2011).
16. Janitza, S. & Hornung, R. On the overestimation of random forest's out-of-bag error. *PLoS ONE* **13**, e0201904. <https://doi.org/10.1371/journal.pone.0201904> (2018).
17. Troyanskaya, O. *et al.* Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520–525. <https://doi.org/10.1093/bioinformatics/17.6.520> (2001).
18. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
19. Gao, C. *et al.* Association of hypertension and antihypertensive treatment with COVID-19 mortality: a retrospective observational study. *Eur. Heart J.* **41**, 2058–2066. <https://doi.org/10.1093/eurheartj/ehaa433> (2020).

20. Richardson, S. *et al.* Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* <https://doi.org/10.1001/jama.2020.6775> (2020).
21. Guo, W. *et al.* Diabetes is a risk factor for the progression and prognosis of COVID-19. *Diabetes Metab. Res. Rev.* <https://doi.org/10.1002/dmrr.3319> (2020).
22. Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med.* **11**, 875–879. <https://doi.org/10.1038/nm1267> (2005).
23. Li, H. *et al.* SARS-CoV-2 and viral sepsis: observations and hypotheses. *Lancet* **395**, 1517–1520. [https://doi.org/10.1016/S0140-6736\(20\)30920-X](https://doi.org/10.1016/S0140-6736(20)30920-X) (2020).
24. Sattar, N., McInnes, I. B. & McMurray, J. J. V. Obesity a risk factor for severe COVID-19 infection: multiple potential mechanisms. *Circulation* <https://doi.org/10.1161/CIRCULATIONAHA.120.047659> (2020).
25. Palaioimos, L. *et al.* Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with COVID-19 in the Bronx, New York. *Metabolism* **108**, 154262. <https://doi.org/10.1016/j.metabol.2020.154262> (2020).
26. Simonnet, A. *et al.* High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity (Silver Spring)* **28**, 1195–1199. <https://doi.org/10.1002/oby.22831> (2020).
27. Huttunen, R. & Syrjanen, J. Obesity and the outcome of infection. *Lancet Infect. Dis.* **10**, 442–443. [https://doi.org/10.1016/S1473-3099\(10\)70103-1](https://doi.org/10.1016/S1473-3099(10)70103-1) (2010).
28. Cheng, Y. *et al.* Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int.* **97**, 829–838. <https://doi.org/10.1016/j.kint.2020.03.005> (2020).
29. Sise, M. E., Baggett, M. V., Shepard, J. O., Stevens, J. S. & Rhee, E. P. Case 17–2020: a 68-year-old man with covid-19 and acute kidney injury. *N. Engl. J. Med.* **382**, 2147–2156. <https://doi.org/10.1056/NEJMcp2002418> (2020).
30. Shi, J. *et al.* Lactate dehydrogenase and susceptibility to deterioration of mild COVID-19 patients: a multicenter nested case-control study. *BMC Med.* **18**, 168. <https://doi.org/10.1186/s12916-020-01633-7> (2020).
31. Montaner, J. S. *et al.* Multisystem organ failure predicts mortality of ICU patients with acute respiratory failure secondary to AIDS-related PCP. *Chest* **102**, 1823–1828. <https://doi.org/10.1378/chest.102.6.1823> (1992).
32. Cai, Q. *et al.* COVID-19: abnormal liver function tests. *J. Hepatol.* <https://doi.org/10.1016/j.jhep.2020.04.006> (2020).
33. Guan, W. J. *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoA2002032> (2020).
34. Feng, G. *et al.* COVID-19 and liver dysfunction: current insights and emergent therapeutic strategies. *J. Clin. Transl. Hepatol.* **8**, 18–24. <https://doi.org/10.14218/JCTH.2020.00018> (2020).
35. Ni, M., Tian, F. B., Xiang, D. D. & Yu, B. Characteristics of inflammatory factors and lymphocyte subsets in patients with severe COVID-19. *J. Med. Virol.* <https://doi.org/10.1002/jmv.26070> (2020).
36. Kernan, K. F. & Carcillo, J. A. Hyperferritinemia and inflammation. *Int. Immunol.* **29**, 401–409. <https://doi.org/10.1093/intimm/dxx031> (2017).
37. Zaim, S., Chong, J. H., Sankaranarayanan, V. & Harky, A. COVID-19 and multiorgan response. *Curr. Probl. Cardiol.* **45**, 100618. <https://doi.org/10.1016/j.cpcardiol.2020.100618> (2020).

Acknowledgements

The study was funded by grants from the Novo Nordisk foundation to MS (#NNF20SA0062879 and #NNF19OC0055183). Imaging for the online calculator was created by Fusion Medical Imaging.

Author contributions

E.J.S., T.S.P., M.N., S.H.M. and M.S. conceived the study and drafted the manuscript. J.P., M.E.N., M.Z.A., G.M.V. conducted data search, extraction, cleaning and validation. C.L., C.I., W.B., O.K., C.H., C.H., S.L., R.S., M.G., N.D., A.L., A.S., M.B., B.I., J.P., M.L., J.M., A.P. and M.N. performed data science analysis. H.C.T.M., A.P., M.H., B.S.K.H. and A.B. performed clinical data validation. All authors performed critical revisions of the manuscript draft and approved the final manuscript. M.S. assumes the role of guarantor, and attests that all authors meet ICMJE criteria for authorship.

Funding

The study was funded by grants from the Novo Nordisk Foundation to MS (#NNF20SA0062879 and #NNF19OC0055183) and MN (#NNF20SA0062879). The foundation took no part in project design, data handling and manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-81844-x>.

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021