Investigating phenotypes of pulmonary COVID-19 recovery – a longitudinal observational prospective multicenter trial

Thomas Sonnweber^{1*}, Piotr Tymoszek^{1,7*}, Sabina Sahanic¹, Anna Boehm¹, Alex Pizzini¹, Anna Luger², Christoph Schwabl², Manfred Nairz¹, Philipp Grubwieser¹, Katharina Kurz¹, Sabine Koppelstätter¹, Magdalena Aichner¹, Bernhard Puchner³, Alexander Egger⁴, Gregor Hoermann^{4,5}, Ewald Wöll⁶, Günter Weiss¹, Gerlig Widmann², Ivan Tancevski^{1#}, Judith Löffler-Ragg^{1#}

¹Department of Internal Medicine II, Medical University of Innsbruck, Innsbruck, Austria

²Department of Radiology, Medical University of Innsbruck, Innsbruck, Austria

³The Karl Landsteiner Institute, Reha Zentrum Münster, Münster, Austria

⁴Central Institute of Medical and Chemical Laboratory Diagnostics, University Hospital Innsbruck, Innsbruck, Austria

⁵MLL Munich Leukemia Laboratory, Munich, Germany

⁶Department of Internal Medicine, St. Vinzenz Hospital, Zams, Austria

⁷ present affiliation: Data Analytics As a Service Tirol, daas.tirol, Innsbruck, Austria

*Contributed equally

#Correspondence:

Prof. Dr. Judith Löffler-Ragg or Priv. Doz. Dr. Ivan Tancevski

Email: judith.loeffler@i-med.ac.at or ivan.tancevski@i-med.ac.at

Tel: +43 50504/81413 or +43 50504/81602

Address: Anichstraße 35, 6020 Innsbruck, Austria

Word count (Introduction, Methods, Results, Discussion): 3993, figure count: 10, table count: 3.

23 **Keywords**

24 computed tomography; cluster analysis; COVID-19; dyspnea; long COVID; lung damage; machine
25 learning; post-COVID-19 syndrome; recovery phenotype; respiratory sequelae; SARS-CoV-2

26 **Impact statement**

27 Protracted systemic and micro-vascular inflammation as well as high anti-SARS-CoV-2 antibody
28 levels are associated with high risk of persistent structural and functional lung deficits six months
29 after COVID-19.

Abstract

Background: The optimal procedures to prevent, identify, monitor and treat long-term pulmonary sequelae of COVID-19 are elusive. Here, we characterized the kinetics of respiratory and symptom recovery following COVID-19.

Methods: We conducted a longitudinal, multi-center observational study in ambulatory and hospitalized COVID-19 patients recruited in early 2020 (n = 145). Pulmonary computed tomography (CT) and lung function (LF) readouts, symptom prevalence, clinical and laboratory parameters were collected during acute COVID-19 and at 60-, 100- and 180-days follow-up visits. Recovery kinetics and risk factors were investigated by logistic regression. Classification of clinical features and participants was accomplished by unsupervised and semi-supervised multi-parameter clustering and machine learning.

Results: At the six-month follow-up, 49% of participants reported persistent symptoms. The frequency of structural lung CT abnormalities ranged from 18% in the mild outpatient cases to 76% in the ICU convalescents. Prevalence of impaired LF ranged from 14% in the mild outpatient cases to 50% in the ICU survivors. Incomplete radiological lung recovery was associated with increased anti-S1/S2 antibody titer, IL-6 and CRP levels at the early follow-up. We demonstrated that the risk of perturbed pulmonary recovery could be robustly estimated at early follow-up by clustering and machine learning classifiers employing solely non-CT and non-LF parameters.

Conclusion: The severity of acute COVID-19 and protracted systemic inflammation is strongly linked to persistent structural and functional lung abnormality. Automated screening of multi-parameter health record data may assist at the prediction of incomplete pulmonary recovery and optimize COVID-19 follow-up management.

Funding: The State of Tyrol (GZ 71934), Boehringer Ingelheim/Investigator initiated study (IIS 1199-0424).

Trial Registration: ClinicalTrials.gov: NCT04416100

Introduction

The ongoing COVID-19 pandemic challenges health care systems. As of December 2021, the John Hopkins dashboard (1) reports 276 million cases and 5.4 million COVID-19-related deaths worldwide (2). Although the vast majority of COVID-19 patients display mild disease, approximately 10-15% of cases progress to a severe condition and approximately 5% suffer from critical illness (3, 4). Similar to severe acute respiratory syndrome (SARS) (5–8), a significant portion of COVID-19 patients report lingering or recurring clinical impairment and cardiopulmonary recovery may take several months to years (9–17). This observation has led to the introduction of the term ‘long COVID’, defined by the persistence of COVID-19 symptoms for more than four weeks, and the ‘post-acute sequelae of COVID-19’ (PASC) referring to symptom persistence for more than twelve weeks (10, 18, 19). Evidence-based strategies for prediction, monitoring and treatment of post-acute COVID-19 sequelae are urgently needed (20).

We herein prospectively analyzed the prevalence of non-resolving structural and functional lung abnormalities and persistent COVID-19-related symptoms six months after diagnosis. Using univariate risk modeling as well as multi-parameter clustering and machine learning (ML), we investigated sets of risk-factors and tested the operability of ML classifiers at predicting protracted lung and symptom recovery. The classification and prediction procedures were implemented in an open-source risk assessment tool (<https://im2-ibk.shinyapps.io/CovILD/>).

Methods

Study design

The CovILD (“Development of interstitial lung disease in COVID-19”) multi-center, longitudinal observational study (9) was initiated in April 2020. Adult residents of Tyrol, Austria, with symptomatic, PCR-confirmed SARS-CoV-2 infection (21) were enrolled by: Department of Internal Medicine II at the Medical University of Innsbruck (primary follow-up center), St. Vinzenz Hospital in Zams and the acute rehabilitation facility in Münster (**Table 1**). The participants were diagnosed with COVID-19 between 3rd March and 29th June 2020. In course of the study, including the 2020 SARS-CoV-2 outbreak and follow-up visits, the regional health system was able to guarantee an unrestricted, optimal standard of diagnostics and care for all participants. Corticosteroids were not standard of care during the recruitment period of the study, thus were not administered as a therapy of acute COVID-19. Some participants with non-resolving pneumonia received systemic steroids beginning from week 4 post-diagnosis at the discretion of the physician (**Table 2**). The analysis endpoints were the presence of any, mild (severity score ≤ 5) and moderate-to-severe (severity score > 5) lung CT abnormalities, impaired lung function (LF) and persistent COVID-19 symptoms at the 180-day follow-up visit (**Table 3**).

In total 190 COVID-19 patients were screened for participation. Thereof, $n = 18$ subjects refused to give informed consent, $n = 27$ declared difficulties to appear at the study follow-ups. Data of $n = 145$ participants were eligible for analysis (**Figure 1**). All participants gave written informed consent. The study was approved by the institutional review board at the Medical University of Innsbruck (approval number: 1103/2020) and registered at ClinicalTrials.gov (NCT04416100).

Procedures

We retrospectively assessed patient characteristics during acute COVID-19 and performed follow-up investigations at 60 days (63 ± 23 days (mean $\pm$ SD); visit 1), 100 days (103 ± 21 days; visit 2) and 180 days (190 ± 15 days; visit 3) after diagnosis of COVID-19. Each visit included symptom and physical performance assessment with a standardized questionnaire, LF testing, standard laboratory testing and a CT scan of the chest. The variables available for analysis with their stratification schemes are listed in **Appendix 1 – table 1**.

Serological markers were determined in certified laboratories (Central Institute of Clinical and Chemical Laboratory Diagnostics, Rheumatology and Infectious Diseases Laboratory, both at the

University Hospital of Innsbruck). C-reactive protein (CRP), interleukin-6 (IL-6), N-terminal pro natriuretic peptide (NT-proBNP), and serum ferritin were measured using a Roche Cobas 8000 analyzer. D-dimer was determined with a Siemens BCS-XP instrument using the Siemens D-Dimer Innovance® reagent. Anti-S1/S2 protein SARS-CoV-2 immunoglobulin gamma (IgG) were quantified with LIAISON® chemoluminescence assay (DiaSorin, Italy), expressed as binding antibody units (BAU) and stratified by quartiles (22).

Low dose (100 kVp tube potential) craniocaudal CT scans of the chest were acquired without iodine contrast and without ECG gating on a 128 slice multi-detector CT (128×0.6 mm collimation, 1.1 spiral pitch factor, SOMATOM Definition Flash, Siemens Healthineers, Erlangen, Germany). In case of clinically suspected pulmonary embolism, CT scans were performed with a contrast agent. Axial reconstructions were done with 1 mm slices. CT scans were evaluated for ground-glass opacities, consolidations, bronchial dilation, and reticulations as defined by the Fleischner society. Lung findings were graded with a quantitative CT severity score (0 – 25 points) (9).

Impaired LF was defined as: (1) forced vital capacity (FVC) < 80% or (2) forced expiratory volume in 1 second (FEV₁) < 80% or (3) FEV₁:FVC < 70% or (4) total lung capacity (TLC) < 80% or (5) diffusing capacity of carbon monoxide (DLCO) < 80% predicted.

Statistical analysis

Statistical analyses were performed with R version 4.0.5 (**Figure 1**). Data transformation and visualization were accomplished by *tidyverse* (23), *ggplot2* (24), *ggvenn*, *plotROC* (25) and *cowplot* (26) packages. The recorded variables were binarized as shown in **Appendix 1 – table 1**. Acute COVID-19 severity strata were defined as presented in **Table 1**. P values were corrected for multiple comparisons with the Benjamini-Hochberg method (27), effects were termed significant for $p < 0.05$.

Variable overlap, kinetics and risk modeling

Overlap between the 180-day follow-up outcome features was assessed by analysis of quasi-proportional Venn plots (package *nVennR*) (28) and calculation of the Cohen's κ statistic (package *vcd*) (29). Kinetics of binary outcome variables in participants subsets with the complete longitudinal data record was modeled with mixed-effect logistic regression (random effect: individual, fixed effect: time, packages *lme4* (30) and *lmerTest* (31)). Analyses in the severity groups were done with separate models. Significance was assessed by the likelihood ratio test (LRT) against the random-term-only model. Univariate risk modeling was performed with fixed-effect

logistic regression (**Appendix 1 – table 2**). Odds ratio (OR) significance was determined by Wald Z test.

Cluster analysis

Clustering of non-CT and non-LF binary clinical features (**Appendix 1 – table 1**) was accomplished with PAM algorithm (partitioning around medoids, package *cluster*) (32) and simple matching distance (SMD, package *nomclust*) (33). Association analysis for the participants was performed with a combined procedure involving clustering of the observations by the self-organizing map algorithm (SOM, 4×4 hexagonal grid, SMD distance, *kohonen* package) followed by clustering of the SOM nodes by the Ward.D2 hierarchical clustering algorithm (Euclidean distance, *hclust()* function, package *stats*) (34–36). Clustering analyses were performed in the participant subset with the complete set of clustering variables. The selection of the optimal clustering algorithm was motivated by the highest ratio of between-cluster to total variance and the best stability measured by mean classification error in 20-fold cross-validation (CV) (**Figure 6 – figure supplement 1AB, Figure 7 – figure supplement 1AB**) (37). The optimal cluster number was determined by the bend of the within-cluster sum-of-squares curve (function *fviz_nbclust()*, package *factoextra*) and by the stability in 20-fold CV (**Figure 6 – figure supplement 1CD, Figure 7 – figure supplement 1DF**) (37, 38) as well as by a visual inspection of the SOM node clustering dendrograms **Figure 7 – figure supplement 1E**. Assignment of 180-day follow-up outcome features to the clusters of clinical parameters was accomplished with a k-nearest neighbor (kNN) label propagation algorithm (**Appendix 1 – table 3**) (10, 39). Cluster assignment visualization in a 4-dimensional principal analysis score plot was done with the *PCAprproj()* tool (package *pcaPP*) (40). To determine the importance of particular clustering variables, the variance (between-cluster to total variance ratio) between the initial cluster structure and the structure with random re-sampling of the variable was compared, as initially proposed for the random forests machine learning classifier (41). Frequencies of the outcome events in the participant clusters were compared with χ^2 test. In house-developed association analysis wrappers are available at <https://github.com/PiotrTymoszuk/clustering-tools-2>.

Machine learning

Machine learning (ML) classifiers: C5.0 (package *C50*) (42), random forests (*randomForest*) (41), support vector machines with radial kernel (*kernelab*) (43), neural networks (*nnet*) (44), elastic net

166 (*glmnet*) (45) were trained to predict the 180-day follow-up outcomes employing non-CT and non-
167 LF binary explanatory features (**Appendix 1 – table 1**). The ML training was performed in
168 the participant subsets with the complete set of explanatory and outcome variables. The training,
169 optimization and CV (20-fold, 5 repetitions) was accomplished by the *train()* tool from *caret*
170 package, with the Cohen’s κ statistic as a model selection metric (**Appendix 1 – table 4**)
171 (46). Classifier ensembles were constructed with the elastic net procedure (*caretStack()* function,
172 *caretEnsemble* package, **Appendix 1 – table 4**) (47). Classifier performance in the training
173 cohort and CV was assessed by receiver-operating characteristics, Cohen’s κ and accuracy
174 (packages *caret* and *vcd*, **Appendix 1 – table 5**) (29, 46). Variable importance measures
175 were extracted from the C5.0 (percent variable usage, *c5imp()* function, package *C50*) (42), random
176 forests (Δ Gini index, *importance()*, package *randomForest*) (41) and elastic net classifiers
177 (regression coefficient β , *coef()*, package *glmnet*) (45).

178 ***Pulmonary recovery assessment app***

179 Participant clustering and machine learning classifiers trained in the CovILD cohort were
180 implemented in an open-source online pulmonary assessment R shiny app ([https://im2-](https://im2-ibk.shinyapps.io/CovILD/)
181 [ibk.shinyapps.io/CovILD/](https://im2-ibk.shinyapps.io/CovILD/), code: [https://github.com/PiotrTymoszuk/COVILD-recovery-assessment-](https://github.com/PiotrTymoszuk/COVILD-recovery-assessment-app)
182 [app](https://github.com/PiotrTymoszuk/COVILD-recovery-assessment-app)). Prediction of the cluster assignment based on the user-provided patient data is done by the
183 kNN label propagation algorithm (10, 39).

Results

Patient characteristics

The CovILD study participants (n = 145) were predominantly male (57.8%), the age ranged between 19 and 87 years. 77.2% of participants displayed pre-existing comorbidity, predominantly cardiovascular and metabolic disease. The cohort included mild (outpatient care, 24.8%), moderate (hospitalization without oxygen supply, 25.5%), severe (hospitalization with oxygen supply, 27.6%), and critical (ICU treatment, 22.1%) cases of acute COVID-19 (**Table 1**). The majority of hospitalized participants received anti-infectives during acute COVID-19, anti-coagulative and/or anti-platelet treatment was introduced primarily in the ventilated patients. Systemic steroid administration was initiated at the discretion of the physician beginning from week 4 after diagnosis (**Table 2**).

Clinical recovery after COVID-19

Most patients, irrespective of the acute COVID-19 severity, showed a significant resolution of disease symptoms over time (**Figure 2A**). Persistent complaints at the six-month follow-up were reported by 49% of the study subjects (**Table 3**), with self-reported impaired physical performance (34.7%), sleep disorders (27.1%) and exertional dyspnea (22.8%) as leading manifestations. The frequency of all investigated symptoms declined significantly, even though the pace of their resolution was remarkably slower in the late (100- and 180-day follow-ups) than in the early recovery phase (acute COVID-19 till 60-day follow-up) (**Figure 2B**).

Impaired lung function (LF) was observed in 33.6% of the participants at the six-month follow-up (**Table 3**). Except for the critical COVID-19 survivors (60 days: 66.7%, 180 days post-COVID-19: 50%), no significant reduction in the frequency of LF impairment over time was observed (**Figure 3**). At the six-month follow-up, structural lung abnormalities were found in 48.5% of patients and moderate-to-severe radiological lung alterations (CT severity score > 5) were present in 19.4% of participants (**Table 3**). The majority of the participants with impaired LF displayed radiological lung findings. However, a substantial fraction of CT abnormalities, especially mild ones, were accompanied neither by persistent symptoms nor by LF deficits (**Figure 3 – figure supplement 1, Figure 3 – figure supplement 2, Figure 3 – figure supplement 3A**).

The frequency, scoring and recovery of CT lung findings were related to the severity of acute infection. Pulmonary lesions scored > 5 CT severity points at the 180-day follow-up were most frequent in the individuals with severe and critical acute COVID-19 (**Figure 3 – figure supplement 3**). Notably, the hospitalized group with oxygen therapy demonstrated the fastest recovery kinetics. As for the symptom resolution, LF and CT lung recovery decelerated in the late phase of COVID-19 convalescence (**Figure 3**).

Risk factors of protracted recovery

To identify risk factors of delayed recovery at the six-month follow-up, we screened a set of 52 binary clinical parameters (**Appendix 1 – table 1**) recorded during acute COVID-19 and at the 60-day visit by univariate modeling (**Appendix 1 – table 2**). By this means, no significant correlates for long-term symptom persistence were identified. Risk factors and readouts of severe and critical COVID-19 including multi-morbidity, malignancy, male sex, prolonged hospitalization, ICU stay and immunosuppressive therapy were significantly associated with persistent CT (**Figure 4**) and LF abnormalities (**Figure 5**). Persistently elevated inflammatory markers, IL-6 (> 7 ng/L) and CRP (> 0.5 mg/L), were strong unfavorable risk factors for incomplete radiological and functional pulmonary recovery. Additionally, the biochemical readout of microvascular inflammation, D-dimer (> 500 pg/mL) was significantly linked to LF deficits. Low serum anti-S1/S2 IgG titers at the 60-day follow-up and ambulatory acute COVID-19 correlated with an improved pulmonary recovery (**Figure 4, Figure 5**).

Clusters of clinical features linked to persistent symptoms and lung abnormalities

Employing the unsupervised PAM (partitioning around medoids) algorithm (32), three clusters of co-occurring non-CT and non-LF clinical features of acute COVID-19 and early convalescence (**Appendix 1 – table 1**) were identified (**Figure 6 – figure supplement 1, Appendix 1 – table 3**): (1) cluster 1 with male sex, hypertension, cardiovascular and metabolic comorbidity, (2) cluster 2 including characteristics of acute COVID-19 severity and inflammatory markers, and (3) cluster 3 consisting of acute and persistent COVID-19 symptoms (**Figure 6 – figure supplement 2, Appendix 1 – table 3**).

The six-month follow-up outcome variables were incorporated in the cluster structure using k-nearest-neighbors prediction (39). Long-term symptom persistence was associated with acute and long-lasting COVID-19 symptoms in cluster 3, whereas pulmonary outcome parameters grouped with cluster 2 features (**Figure 6A, Figure 6 – figure supplement 2, Appendix 1 – table 3**). Preexisting comorbidities such as malignancy, kidney, lung and gastrointestinal disease, obesity and diabetes were found the closest cluster neighbors of mild CT abnormalities (severity score ≤ 5). Moderate-to-severe structural alterations (severity score > 5) and LF deficits were, in turn, tightly linked to markers of protracted systemic inflammation (IL-6, CRP, anemia of inflammation) (48) (**Figure 6B**).

Risk stratification for perturbed pulmonary recovery by unsupervised clustering

Next, we tested whether subsets of patients at risk of an incomplete six-month recovery may be defined by a similar clustering procedure employing exclusively non-CT and non-LF clinical variables (**Appendix 1 – table 1**). Applying a combined self-organizing map (SOM) – hierarchical clustering approach, three clusters of the study participants were identified (**Figure 7, Figure 7 – figure supplement 1**) (34, 35). Prolonged hospitalization, anti-infective therapy, overweight or obesity, pain during acute COVID-19 and low anti-S1/S2 titers at the 60-day follow-up were found the most influential clustering features (**Figure 7 – figure supplement 2**) (41). The patient subsets identified by the SOM approach differed significantly in frequency of radiological lung abnormalities and substantially, yet not significantly, in the frequency of LF impairment at the 180-day follow-up. In particular, most of the individuals assigned to the largest, low-risk (LR) subset were CT and LF abnormality-free. The frequency and severity of radiological pulmonary findings were elevated in the smallest intermediate-risk subset (IR) and peaked in the high-risk (HR) group (**Figure 8A**). Despite a comparable frequency of long-term symptoms between the LR, IR and HR subsets (**Figure 8A**), the HR collective showed the lowest prevalence of dyspnea, cough, night sweating, pain, gastrointestinal manifestations and complete absence of hyposmia at the 180-day follow-up (**Figure 8B**). Although the LR subset primarily comprised mild COVID-19 cases and the HR subset ICU survivors, the cluster assignment (IR vs LR, HR vs LR) remained an independent correlate of persistent CT and LF

abnormalities after adjustment for the acute COVID-19 severity (**Figure 8 – figure supplement 1**).

Prediction of persistent symptoms and pulmonary abnormalities by machine learning

Finally, we investigated if the six-month follow-up outcome may be predicted by machine learning (ML) classifiers trained with a set of non-CT and non-LF variables recorded during acute COVID-19 and at the 60-day follow-up (**Appendix 1 – table 1**). To this end, five technically unrelated ML classifiers were tested (**Appendix 1 – table 4**) (46): C5.0 (42), random forests (RF) (41), support vector machines with radial kernel (SVM-R) (43), shallow neural network (Nnet) (44) and elastic net generalized linear regression (glmNet) (45). In addition, the single classifiers with varying outcome-specific accuracy (**Figure 9 – figure supplement 1**) were bundled into ensembles by the elastic net procedure (**Figure 9 – figure supplement 2, Appendix 1 – table 4**) (46, 47). Finally, the classifier and ensemble performance was investigated in the training cohort and 20-fold CV by receiver-operator characteristic (ROC) (**Appendix 1 – table 5**).

All tested ML algorithms and ensembles demonstrated good accuracy (area under the curve [AUC] > 0.78) and sensitivity (> 0.84) at predicting any lung CT abnormalities at the six-month follow-up in the study cohort serving as a training data set. Their efficiency in CV was moderate (AUC: 0.69 – 0.81, sensitivity: 0.69 – 0.78) (**Figure 9, Figure 9 – figure supplement 3, Appendix 1 – table 5**). In turn, moderate-to-severe structural lung findings were recognized with markedly lower sensitivity both in the training data set (> 0.43) and the CV (0.39 – 0.48). Even though impaired LF and persistent symptoms were common at the six-month follow-up in the training data set (**Figure 2, Figure 3**), nearly half of the cases were not identified by any of the tested ML algorithms and their ensembles in the CV setting (**Figure 9, Figure 9 – figure supplement 3, Appendix 1 – table 5**). The sensitivity of the ensembles and single classifiers at predicting CT and LF abnormalities was substantially better in severe and critical COVID-19 survivors than in ambulatory and moderate cases (**Figure 10, Appendix 1 – table 6**).

298 The most important explanatory variables for pulmonary abnormalities by three unrelated classifiers
299 (C5.0, RF and glmNet) included pre-existing malignancy, multi-morbidity, markers of systemic
300 inflammation (IL-6 and CRP) and anti-S1/S2 antibody levels at the 60-day follow-up (**Figure 9**
301 **– figure supplement 4, Figure 9 – figure supplement 5, Figure 9 –**
302 **figure supplement 6**). The highly influential parameters at prediction of symptoms at the
303 180-day follow-up encompassed symptom presence at 60-day follow-up, as well as obesity and
304 dyspnea during acute COVID-19 (**Figure 9 – figure supplement 7**).

Discussion

Herein, we prospectively evaluated trajectories of COVID-19 recovery in an observational cohort enrolled in the Austrian CovILD study (9). Despite the resolution of symptoms and pulmonary abnormalities at the six-months follow-up in a large fraction of the study participants, the recovery pace was substantially slower in the late convalescence when compared with the first three months after diagnosis (9, 13). Persistent symptoms and CT findings were detected in more than 40% and reduced lung function in approximately one-third of the cohort, which is in line with recovery kinetics and signs of lung lesion chronicity reported by others (11–16). By comparison, similar protracted pulmonary was reported for SARS (5–8) and for non-COVID-19 acute respiratory distress syndrome (ARDS) (49, 50). Of note, treatment approaches for hospitalized patients in our and in similar cohorts recruited at the pandemic onset in early 2020 (11–15) differ significantly from the current standard of care for acute COVID-19 which includes early systemic steroid use, anti-viral and various immune-modulatory medications. How improved standardized therapy and anti-SARS-CoV-2 vaccination affect the clinical and pulmonary recovery needs to be investigated.

In roughly half of our study participants with abnormal lung CT findings, and especially in those with low-grade structural abnormalities, no overt lung function impairment at follow-up was discerned. Still, even sub-clinical lung alterations may bear the potential for clinically relevant progression of interstitial lung disease (51, 52) requiring systematic CT and lung function monitoring. Conversely, symptom persistence was weakly associated with incomplete functional or structural pulmonary recovery.

Since post-acute COVID-19 sequelae are found in as many as 10% of COVID-19 patients (10, 17, 19), robust, resource-saving tools assessing the individual risk of pulmonary complications are urgently needed (18, 20). Co-variables and characteristics of severe acute COVID-19 such as male sex, age and pre-existing comorbidities, hospitalization, ventilation and ICU stay were proposed as risk factors of persistent pulmonary impairment (9, 11, 13, 14, 20). However, their applicability in predicting complications of pulmonary recovery from mild or moderate COVID-19 is limited. Our results of univariate modeling, clustering and machine learning prediction point towards a distinct long-term pulmonary risk phenotype which manifests during acute COVID-19 and early recovery and whose central components are protracted systemic (IL-6, CRP, anemia of inflammation) and microvascular inflammation (D-dimer), strong humoral response (anti-S1/S2 IgG) demographic risk factors and comorbidities (48). Hence, consecutive monitoring of systemic inflammatory parameters analogous to concepts of interstitial lung disease in autoimmune disorders (53) and anti-

S1/S2 antibody levels may improve identification of the individuals at risk of chronic pulmonary damage irrespective of the acute COVID-19 severity.

Clustering and machine learning have been employed for deep phenotyping and predicting acute and post-acute COVID-19 outcomes in multi-variable data sets (10, 54–57). We demonstrate that subsets of COVID-19 patients which significantly differ in the risk for long-term CT abnormalities may be defined by an easily accessible clinical parameter set available at the early post-COVID-19 assessment (34, 35). This approach did not involve any CT or lung function variables. Furthermore, the cluster classification correlated with the risk of long-term pulmonary abnormalities independently of the acute COVID-19 severity. Thus, these characteristics provide a useful tool for broad screening of convalescent populations, including individuals who experienced mild or moderate COVID-19.

We show that technically unrelated machine learning classifiers and their ensemble trained without CT and lung function explanatory variables can predict lung CT findings independently of their grading at the six-month follow-up with good specificity and sensitivity in the training collective and cross-validation. By contrast, the more specific prediction of moderate-to-severe lung CT or risk estimation for lung function deficits demonstrated a limited sensitivity. For the moderate-to-severe CT abnormalities, this can be primarily traced back to their low frequency resulting in a sub-optimal classifier training, especially in cross-validation. A substantial fraction of the participants (20.7%, n = 30) suffered from a pre-existing respiratory condition (pulmonary disease, asthma or COPD) likely paralleled by lung function reduction, which possibly confounded the prediction of the post-COVID-19 lung function deficits both by clustering and ML. Accumulating evidence suggests that post-acute COVID-19 symptoms are highly heterogeneous conditions with multi-organ, neuro-cognitive and psychological manifestations (10, 58, 59), which may differ in risk factor constellations. This could explain, why univariate modeling, clustering and machine learning failed to estimate persistent symptom risk in our small study cohort. In general, the machine learning prediction quality may greatly benefit from a larger training data set and inclusion of additional explanatory variables such as cellular readouts of inflammation, in-depth medication and broader acute symptom data. Nevertheless, the herein described cluster- and ML classifiers represent resource-effective tools which may assist at the screening of medical record data and identification of COVID-19 patients requiring systematic CT and lung function monitoring. To facilitate the identification of patients at risk for protracted respiratory recovery, and to enable validation in an external collective, we implemented the clustering and prediction procedures in an open-source risk assessment application (<https://im2-ibk.shinyapps.io/CovILD/>).

370 Our study bears limitations primarily concerning the low sample size and the cross-sectional
371 character of the trial. Due to impaired availability of the patients because of prolonged inpatient
372 rehabilitation the 60- and 100-day follow-up visits in part showed a temporal overlap which may
373 have impacted on accuracy of the longitudinal data. Missingness of the consecutive outcome
374 variable record and the participant dropout, in particular of mild and moderate COVID-19 cases,
375 may have also potentially confounded the participant clustering results and machine learning risk
376 estimation for CT abnormalities and lung function impairment since prolonged hospitalization was
377 found a crucial cluster-defining and influential explanatory feature. Additionally, even though the
378 reproducibility of the risk assessment algorithms was partially addressed by cross-validation, cluster
379 and ML classifiers call for verification in a larger, independent multi-center collective of COVID-19
380 convalescents.

381 In summary, in our CovILD study cohort we found a high frequency of CT and lung function
382 abnormalities and persistent symptoms at the six-month follow-up, and a flattened recovery kinetics
383 after three months post-COVID-19. Systematic risk modeling reveled a set of clinical variables
384 linked to protracted pulmonary recovery apart from the severity of acute infection such as
385 inflammatory markers, anti-S1/S2 IgG levels, multi-morbidity, and male sex. We demonstrate that
386 clustering and machine learning classifiers may help to identify individuals at risk of persistent lung
387 lesions and to relocate medical resources to prevent long-term disability.

Acknowledgments

We acknowledge the commitment of the staff and providers of our institutions through the COVID-19 crisis and the suffering and loss of our patients as well as their families.

Conflict of interest

Piotr Tymoszek is (from May 2021 on) a free-lance data scientist working in his own enterprise ‘Data Analytics as a Service Tirol’. He received an honorary for the study data management, curation and analysis and minor manuscript work. The other authors declare no conflict of interest related to this study.

Funding information

The study was funded by the research fund of the state of Tyrol (Project GZ 71934, J.LR.) and an Investigator-Initiated Study grant by Boehringer Ingelheim (IIS 1199-0424, I.T.). The funding bodies did not influence the development of the research and manuscript.

Data availability

The anonymized participants’ data used for the analyzes presented in this report and the complete R analysis pipeline are available from GitHub (https://github.com/PiotrTymoszek/CovILD_6_Months). The code of the online risk assessment app is available at <https://github.com/PiotrTymoszek/COVILD-recovery-assessment-app>. The source data for Figures 2 – 10 are available as a Supplementary File.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 2020;20(5):533–534.
2. COVID-19 Map - Johns Hopkins Coronavirus Resource Center [Internet] <https://coronavirus.jhu.edu/map.html>. cited May 20, 2021
3. Perez-Saez J et al. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland [Internet]. *Lancet Infect. Dis.* 2021;21(4):e69–e70.
4. Huang C et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China [Internet]. *Lancet* 2020;395(10223):497–506.
5. Hui DS et al. The 1-Year Impact of Severe Acute Respiratory Syndrome on Pulmonary Function, Exercise Capacity, and Quality of Life in a Cohort of Survivors [Internet]. *Chest* 2005;128(4):2247–2261.
6. Ng CK et al. Six month radiological and physiological outcomes in severe acute respiratory syndrome (SARS) survivors [Internet]. *Thorax* 2004;59(10):889–891.
7. Ngai JC et al. The long-term impact of severe acute respiratory syndrome on pulmonary function, exercise capacity and health status [Internet]. *Respirology* 2010;15(3):543–550.
8. Lam MHB et al. Mental morbidities and chronic fatigue in severe acute respiratory syndrome survivors long-term follow-up [Internet]. *Arch. Intern. Med.* 2009;169(22):2142–2147.
9. Sonnweber T et al. Cardiopulmonary recovery after COVID-19 - an observational prospective multi-center trial. [Internet]. *Eur. Respir. J.* [published online ahead of print: December 10, 2020]; doi:10.1183/13993003.03481-2020
10. Sahanic S et al. Phenotyping of acute and persistent COVID-19 features in the outpatient setting: exploratory analysis of an international cross-sectional online survey [Internet]. *Clin. Infect. Dis.* [published online ahead of print: November 26, 2021]; doi:10.1093/CID/CIAB978
11. Caruso D et al. Postacute Sequelae of COVID-19 Pneumonia: 6-month Chest CT Follow-up [Internet]. *Radiology* 2021;210834.
12. Huang L et al. 1-year outcomes in hospital survivors with COVID-19: a longitudinal cohort study [Internet]. *Lancet* 2021;398(10302):747–758.
13. Huang C et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study [Internet]. *Lancet* 2021;397(10270):220–232.
14. Faverio P et al. Six-Month Pulmonary Impairment after Severe COVID-19: A Prospective, Multicentre Follow-Up Study [Internet]. *Respiration.* 2021;100(11):1078–1087.
15. Hellemons ME et al. Persistent Health Problems beyond Pulmonary Recovery up to 6 Months after Hospitalization for SARS-CoV-2; A Longitudinal Study of Respiratory, Physical and

442 Psychological Outcomes [Internet]. *Ann. Am. Thorac. Soc.* [published online ahead of print:
443 September 28, 2021]; doi:10.1513/ANNALSATS.202103-340OC

444 16. Zhou F et al. Assessment of Sequelae of COVID-19 Nearly 1 Year After Diagnosis. *Front. Med.*
445 2021;8:2260.

446 17. Venkatesan P. NICE guideline on long COVID [Internet]. *Lancet. Respir. Med.* 2021;9(2):129.

447 18. Shah W, Hillman T, Playford ED, Hishmeh L. Managing the long term effects of covid-19:
448 Summary of NICE, SIGN, and RCGP rapid guideline [Internet]. *BMJ* 2021;372.
449 doi:10.1136/bmj.n136

450 19. Sudre CH et al. Attributes and predictors of long COVID [Internet]. *Nat. Med.* 2021;27(4).
451 doi:10.1038/s41591-021-01292-y

452 20. Raghu G, Wilson KC. COVID-19 interstitial pneumonia: monitoring the clinical course in
453 survivors [Internet]. *Lancet Respir. Med.* 2020;8(9):839–842.

454 21. Coronavirus [Internet]https://www.who.int/health-topics/coronavirus#tab=tab_3. cited May 20,
455 2021

456 22. Ferrari D et al. Harmonization of six quantitative SARS-CoV-2 serological assays using sera of
457 vaccinated subjects [Internet]. *Clin. Chim. Acta.* 2021;522:144.

458 23. Wickham H et al. Welcome to the Tidyverse. *J. Open Source Softw.* 2019;4(43):1686.

459 24. Wickham H. *ggplot2: Elegant Graphics for Data Analysis [Internet]*. New York: Springer-
460 Verlag; 2016:

461 25. Sachs MC. PlotROC: A tool for plotting ROC curves [Internet]. *J. Stat. Softw.* 2017;79(1):1–19.

462 26. Wilke CO. *Fundamentals of Data Visualization: A Primer on Making Informative and*
463 *Compelling Figures*. Sebastopol: O'Reilly Media; 2019:

464 27. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful
465 Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 1995;57(1):289–300.

466 28. Pérez-Silva JG, Araujo-Voces M, Quesada V. nVenn: generalized, quasi-proportional Venn and
467 Euler diagrams [Internet]. *Bioinformatics* 2018;34(13):2322–2324.

468 29. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa.
469 *Psychol. Bull.* 1969;72(5):323–327.

470 30. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J.*
471 *Stat. Softw.* 2015;67(1):1–48.

472 31. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed
473 Effects Models . *J. Stat. Softw.* 2017;82(13):1–26.

474 32. Schubert E, Rousseeuw PJ. Faster k-Medoids Clustering: Improving the PAM, CLARA, and
475 CLARANS Algorithms [Internet]. In: *Lecture Notes in Computer Science (including subseries*

476 *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*). Springer; 2019:171–
 477 187

478 33. Boriah S, Chandola V, Kumar V. Similarity measures for categorical data: A comparative
 479 evaluation [Internet]. In: *Society for Industrial and Applied Mathematics - 8th SIAM International*
 480 *Conference on Data Mining 2008, Proceedings in Applied Mathematics 130*. 2008:243–254

481 34. Vesanto J, Alhoniemi E. Clustering of the self-organizing map. *IEEE Trans. Neural Networks*
 482 2000;11(3):586–600.

483 35. Kohonen T. *Self-Organizing Maps [Internet]*. Berlin, Heidelberg: Springer Berlin Heidelberg;
 484 1995:

485 36. Wehrens R, Kruisselbrink J. Flexible self-organizing maps in kohonen 3.0 [Internet]. *J. Stat.*
 486 *Softw.* 2018;87(7):1–18.

487 37. Lange T, Roth V, Braun ML, Buhmann JM. Stability-Based Validation of Clustering Solutions.
 488 *Neural Comput.* 2004;16(6):1299–1323.

489 38. Wang J. Consistent selection of the number of clusters via crossvalidation [Internet]. *Biometrika*
 490 2010;97(4):893–904.

491 39. Leng M, Wang J, Cheng J, Zhou H, Chen X. Adaptive semi-supervised clustering algorithm
 492 with label propagation. *J. Softw. Eng.* 2014;8(1):14–22.

493 40. Croux C, Filzmoser P, Oliveira MR. Algorithms for Projection-Pursuit robust principal
 494 component analysis. *Chemom. Intell. Lab. Syst.* 2007;87(2):218–225.

495 41. Breiman L. Random forests [Internet]. *Mach. Learn.* 2001;45(1):5–32.

496 42. Quinlan JR. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan
 497 Kaufmann Publishers Inc.; 1993:

498 43. Weston J, Watkins C. Multi-Class Support Vector Machines. In: 1998:

499 44. Ripley BD. *Pattern recognition and neural networks [Internet]*. Cambridge University Press;
 500 2014:

501 45. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via
 502 coordinate descent [Internet]. *J. Stat. Softw.* 2010;33(1):1–22.

503 46. Kuhn M. Building predictive models in R using the caret package. *J. Stat. Softw.* 2008;28(5):1–
 504 26.

505 47. Deane-Mayer ZA, Knowles JE. Ensembles of Caret Models [R package caretEnsemble version
 506 2.0.1] [Internet]2019;<https://cran.r-project.org/package=caretEnsemble>. cited December 13, 2021

507 48. Sonnweber T et al. Persisting alterations of iron homeostasis in COVID-19 are associated with
 508 non-resolving lung pathologies and poor patients’ performance: a prospective observational cohort
 509 study [Internet]. *Respir. Res.* 2020;21(1). doi:10.1186/s12931-020-01546-2

510 49. Wilcox ME et al. Radiologic outcomes at 5 years after severe ARDS [Internet]. *Chest*
511 2013;143(4):920–926.

512 50. Masclans JR et al. Quality of life, pulmonary function, and tomographic scan abnormalities after
513 ARDS [Internet]. *Chest* 2011;139(6):1340–1346.

514 51. Suliman YA et al. Pulmonary function tests: High rate of false-negative results in the early
515 detection and screening of scleroderma-related interstitial lung disease [Internet]. *Arthritis*
516 *Rheumatol.* 2015;67(12):3256–3261.

517 52. Hatabu H et al. Interstitial lung abnormalities detected incidentally on CT: a Position Paper from
518 the Fleischner Society [Internet]. *Lancet. Respir. Med.* 2020;8(7):726.

519 53. Khanna D et al. Etiology, risk factors, and biomarkers in systemic sclerosis with interstitial lung
520 disease [Internet]. *Am. J. Respir. Crit. Care Med.* 2020;201(6):650–660.

521 54. Sudre CH et al. Symptom clusters in COVID-19: A potential clinical prediction tool from the
522 COVID Symptom Study app [Internet]. *Sci. Adv.* 2021;7(12). doi:10.1126/sciadv.abd4177

523 55. Estiri H et al. Evolving phenotypes of non-hospitalized patients that indicate long COVID
524 [Internet]. *BMC Med.* 2021;19(1). doi:10.1186/S12916-021-02115-0

525 56. Demichev V et al. A time-resolved proteomic and prognostic map of COVID-19 [Internet]. *Cell*
526 *Syst.* 2021;12(8):780.

527 57. Benito-León J et al. Using Unsupervised Machine Learning to Identify Age- and Sex-
528 Independent Severity Subgroups Among Patients with COVID-19: Observational Longitudinal
529 Study [Internet]. *J. Med. Internet Res.* 2021;23(5). doi:10.2196/25988

530 58. Evans RA et al. Physical, cognitive, and mental health impacts of COVID-19 after
531 hospitalisation (PHOSP-COVID): a UK multicentre, prospective cohort study [Internet]. *Lancet*
532 *Respir. Med.* 2021;9(11):1275–1287.

533 59. Davis HE et al. Characterizing long COVID in an international cohort: 7 months of symptoms
534 and their impact [Internet]. *EClinicalMedicine* 2021;101019.

535 **Tables**

536 **Table 1. Characteristics of the study population.**

Characteristics (% cohort)	
Total participants – no.	145
Mean age, years	57.3 (SD = 14.3)
Female sex	42.4% (n = 63)
Obesity (body mass index > 30 kg/m ²)	19.3% (n = 28)
Ex-smoker	39.3% (n = 57)

Active smoker	2.8% (n = 4)
Acute COVID-19 severity (% cohort)	
Mild: outpatient	24.8% (n = 36)
Moderate: inpatient without oxygen therapy	25.5% (n = 37)
Severe: inpatient with oxygen therapy	27.6% (n = 40)
Critical: ICU	22.1% (n = 32)
Comorbidities (% cohort)	
None	22.8% (n = 33)
Cardiovascular disease	40% (n = 58)
Pulmonary disease	18.6% (n = 27)
Metabolic disease	43.4% (n = 63)
Chronic kidney disease	6.9% (n = 10)
Gastrointestinal tract diseases	13.8% (n = 20)
Malignancy	11.7% (n = 17)

537

538 **Table 2. Hospitalization and medication during acute COVID-19**

Parameter	Outpatient (n = 36)	Hospitalized (n = 37)	Hospitalized oxygen therapy (n = 40)	Hospitalized ICU (n = 32)
Mean hospitalization time, days	0 (SD = 0)	6.9 (SD = 3.6)	11.8 (SD = 6.3)	34.8 (SD = 15.7)
Hospitalized >7 days	0% (n = 0)	43.2% (n = 16)	80% (n = 32)	100% (n = 32)
Anti-infectives	11.1% (n = 4)	45.9% (n = 17)	72.5% (n = 29)	87.5% (n = 28)
Anti-platelet drugs	2.8% (n = 1)	10.8% (n = 4)	22.5% (n = 9)	25% (n = 8)
Anti-coagulatives	2.8% (n = 1)	2.7% (n = 1)	5% (n = 2)	15.6% (n = 5)
Corticosteroids*†‡	2.8% (n = 1)	5.4 % (n = 2)	22.5% (n = 9)	40.6% (n = 13)
Immunosuppression†‡	0% (n = 0)	2.7% (n = 1)	5% (n = 2)	9.4% (n = 3)

539

540 * from the week 4 post diagnosis on, at the discretion of the physician

541 † immunosuppressive medication prior to COVID-19

‡ subsumed under ‘Immunosuppression, acute COVID-19’ for data analysis

542

543 **Table 3. Radiological, functional and clinical study outcomes**

Outcome	60-day follow-up	100-day follow-up	180-day follow-up
Any lung CT abnormalities (complete: n = 103)	74.8% (n = 77)	60.2% (n = 62)	48.5% (n = 50)
Mild lung CT abnormalities (severity score ≤ 5) (complete: n = 103)	26.2% (n = 27)	36.9% (n = 38)	29.1% (n = 30)
Moderate-to-severe CT abnormalities (severity score > 5) (complete: n = 103)	48.5% (n = 50)	23.3% (n = 24)	19.4% (n = 20)
Functional lung impairment (complete: n = 116)	39.7% (n = 46)	37.1% (n = 43)	33.6% (n = 39)
Persistent symptoms (complete: n = 145)	79.3% (n = 115)	67.6% (n = 98)	49% (n = 71)

544

545

546

Figure Legends

Figure 1. Study inclusion flow diagram and analysis scheme.

Figure 2. Kinetic of recovery from COVID-19 symptoms.

Recovery from any COVID-19 symptoms was investigated by mixed-effect logistic modeling (random effect: individual, fixed effect: time). Significance was determined by the likelihood ratio test corrected for multiple testing with the Benjamini-Hochberg method, p values and the numbers of complete observations are indicated in the plots.

(A) Frequencies of individuals with any symptoms in the study cohort stratified by acute COVID-19 severity.

(B) Frequencies of participants with particular symptoms.

imp.: impaired.

Figure 3. Kinetic of pulmonary recovery.

Recovery from any lung CT abnormalities, moderate-to-severe lung CT abnormalities (severity score > 5) and recovery from functional lung impairment were investigated in the participants stratified by acute COVID-19 severity by mixed-effect logistic modeling (random effect: individual, fixed effect: time). Significance was determined by the likelihood ratio test corrected for multiple testing with the Benjamini-Hochberg method. Frequencies of the given abnormality at the indicated time points are presented, p values and the numbers of complete observations are indicated in the plots.

Figure 3 – figure supplement 1. Co-occurrence of lung CT abnormalities, functional lung impairment and any persistent symptoms.

Numbers and percentages of the study participants with any persistent symptoms, functional lung impairment or lung CT abnormalities at the consecutive follow-up visits presented in quasi-proportional Venn diagrams. The numbers of participants with CT abnormalities, LF impairment

and persistent symptoms are indicated in the diagrams, the numbers of complete observations are shown under the plots.

Figure 3 – figure supplement 2. Co-occurrence of moderate-to-severe lung CT abnormalities, functional lung impairment and any persistent symptoms.

Numbers and percentages of the study participants with any persistent symptoms, functional lung impairment or moderate-to-severe lung CT abnormalities (severity score > 5) at the consecutive follow-up visits presented in quasi-proportional Venn diagrams. The numbers of participants with CT abnormalities, LF impairment and persistent symptoms are indicated in the diagrams, the numbers of complete observations are shown under the plots.

Figure 3 – figure supplement 3. Frequency of mild and moderate-to-severe lung CT abnormalities. Prognostic value of functional lung impairment and persistent symptoms for prediction of radiological lung abnormalities.

(A) Relevance of functional lung impairment and persistent COVID-19 symptoms at predicting any lung CT abnormalities and moderate-to-severe lung CT abnormalities (severity score > 5) at the consecutive followup visits. The concordance of the outcome variables was determined by Cohen's κ coefficient. Statistical significance ($\kappa \neq 0$) was assessed by two-tailed T test corrected for multiple testing with the Benjamini-Hochberg method. Kappa with 95% confidence intervals and p values are presented as a heat map. The number of complete observations is indicated in the plot.

(B) Percentages of mild (severity score ≤ 5) and moderate-to-severe lung CT abnormalities at the consecutive follow-up visits in the study participants stratified by the severity of acute COVID-19. Statistical significance of frequency differences was determined by χ^2 test for trend corrected for multiple testing with the Benjamini-Hochberg method. The number of complete observations is indicated under the plots.

Figure 4. Risk factors of persistent radiological lung abnormalities.

Association of 52 binary explanatory variables (**Appendix 1 – table 1**) with the presence of any lung CT abnormalities (A) or moderate-to-severe lung CT abnormalities (severity score > 5, B) at the 180-day follow-up visit was investigated with a series of univariate logistic models (**Appendix 1 – table 2**). Odds ratio (OR) significance was determined by Wald Z test and corrected for multiple testing with the Benjamini-Hochberg method. OR with 95% confidence intervals for significant favorable and unfavorable factors are presented in Forest plots. Model baseline (ref) and numbers of complete observations are presented in the plot axis text. Q1, Q2, Q3, Q4: 1st, 2nd, 3rd and 4th quartile of anti-S1/S2 IgG titer, ICU: intensive care unit.

Figure 5. Risk factors of persistent functional lung impairment.

Association of 52 binary explanatory variables (**Appendix 1 – table 1**) with the presence of functional lung impairment at the 180-day follow-up visit was investigated with a series of univariate logistic models (**Appendix 1 – table 2**). Odds ratio (OR) significance was determined by Wald Z test and corrected for multiple testing with the Benjamini-Hochberg method. OR with 95% confidence intervals for the significant favorable and unfavorable factors are presented in a Forest plot. Model baseline (ref) and n numbers of complete observations are presented in the plot axis text. Q1, Q2, Q3, Q4: 1st, 2nd, 3rd and 4th quartile of anti-S1/S2 IgG titer, CKD: chronic kidney disease.

Figure 6. Association of incomplete symptom, lung function and radiological lung recovery with demographic and clinical parameters of acute COVID-19 and early recovery.

Clustering of 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 or at the early 60-day follow-up visit (**Appendix 1 – table 1**) was investigated by PAM algorithm with simple matching distance (SMD) dissimilarity measure (**Figure 6 – figure supplement 1, Appendix 1 – table 3**). The cluster assignment for the outcome variables at the 180-day follow-up visit (persistent symptoms, functional lung impairment, mild lung CT abnormalities [severity score ≤ 5] and moderate-to-severe lung CT abnormalities

[severity score > 5]) was predicted by k-nearest neighbor (k-NN) label propagation procedure. Numbers of complete observations and numbers of features in the clusters are indicated in (A). (A) Cluster assignment of the outcome variables (diamonds) presented in the plot of principal component (PC) scores. The first two major PC are displayed. The explanatory variables are visualized as points. Percentages of the data set variance associated with the PC are presented in the plot axes. (B) Five nearest neighbors (lowest SMD) of the outcome variables presented in radial plots. Font size, point radius and color code for SMD values. Q1, Q2, Q3, Q4: 1st, 2nd, 3rd and 4th quartile of anti-S1/S2 IgG titer, GITD: gastrointestinal disease, CKD: chronic kidney disease, ICU: intensive care unit, COPD: chronic obstructive pulmonary disease.

Figure 6 – figure supplement 1. Study feature clustering algorithm.

Clustering of 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 or at the early 60-day follow-up visit (**Appendix 1 – table 1**). (A, B) Comparison of ‘explained’ variances (between-cluster to total sum-of-squares ratio, A) and cluster stability (mean classification error in 20-fold cross-validation, B) in clustering of the data set with several algorithms with k = 3 centers/branches (algorithms: K-means, PAM: partitioning around medoids, HCl ward.D2: hierarchical clustering with Ward.D2 method, distances: SMD: simple matching distance, Jaccard, Dice and Cosine). (C, D) The optimal number of the feature clusters in clustering with the optimally performing PAM algorithm with SMD dissimilarity measure was determined by the bend of the total within-cluster sum-of-squares curve (C) and confirmed by good stability (low mean classification error) in 20-fold cross-validation (D).

Figure 6 – figure supplement 2. Semi-supervised clustering of mild and moderate-to-severe lung CT abnormalities, functional lung impairment and persistent symptoms at the 180-day follow-up with parameters of acute COVID-19 and early convalescence.

Clusters of 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 or at the 60-day follow-up visit (**Appendix 1 – table 1**) were defined by the optimally performing PAM algorithm and SMD (simple matching distance) dissimilarity measure (**Figure 6A, Figure 6 – figure supplement 1, Appendix 1 – table 3**). The cluster assignment for the outcome variables at the 180-day follow-up visit (persistent symptoms, functional lung impairment, mild lung CT abnormalities [severity score ≤ 5] and moderate-to-severe lung CT abnormalities [severity score > 5]) was predicted by k-nearest neighbor (k-NN) label propagation procedure. SMD between the features and their cluster assignments are shown in a heat map. The numbers of the features in the clusters and the total number of observations are indicated under the plot. CVD: cardiovascular disease, Q1, Q2, Q3, Q4: 1st, 2nd, 3rd and 4th quartile of anti-S1/S2 IgG titer, GI: gastrointestinal, PD: pulmonary disease, GITD: gastrointestinal disease, ICU: intensive care unit, COPD: chronic obstructive pulmonary disease, CKD: chronic kidney disease.

Figure 7. Clustering of the study participants by non-lung function and non-CT clinical features.

Study participants (n = 133 with the complete variable set) were clustered in respect to 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 or at the 60-day follow-up visit (**Appendix 1 – table 1**) using a combined self-organizing map (SOM, simple matching distance) and hierarchical clustering (Ward.D2 method, euclidean distance) procedure (**Figure 7 – figure supplement 1**). The numbers of participants assigned to low-risk (LR), intermediate-risk (IR) and high-risk (HR) clusters are indicated in (A).

(A) Cluster assignment of the study participants in the plot of principal component (PC) scores. The first two major PC are displayed. Percentages of the data set variance associated with the PC are presented in the plot axes.

(B) Presence of the most influential clustering features (**Figure 7 – figure supplement 2**) in the participant clusters presented as a heat map. Cluster #1, #2 and #3 refer to the feature clusters defined in **Figure 6**. Q1, Q2, Q3, Q4: 1st, 2nd, 3rd and 4th quartile of anti-S1/S2 IgG titer, GITD: gastrointestinal disease, CKD: chronic kidney disease, CVD: cardiovascular disease, GI: gastrointestinal, PD: pulmonary disease.

Figure 7 – figure supplement 1. Study participant clustering algorithm.

Clustering of the study participants (n = 133 with the complete variable set) in respect to 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 or at the 60-day follow-up visit (**Appendix 1 – table 1**). The procedure involved clustering of the observations with self-organizing maps (SOM, 4 × 4 hexagonal grid, distances: SMD: simple matching distance, Jaccard, Dice or cosine) followed by clustering of the SOM nodes (algorithms: HCl ward.D2: hierarchical clustering with Ward.D2 method, K-means, PAM: partitioning around medoids, distance: euclidean). Different combinations of observation dissimilarity measures and SOM node clustering algorithms were tested in the search for the optimal clustering algorithm.

(A, B) Comparison of ‘explained’ variances (between-cluster to total sum-of-squares ratio, A) and cluster stability (mean classification error in 20-fold cross-validation, B) in clustering of the data set with different observation distance measures and SOM node clustering algorithms.

(C – E) The optimal number of the SOM node clusters in clustering with the optimally performing SOM HCl algorithm with SMD observation dissimilarity measure. The optimal cluster number was determined by the bend of the total within-cluster sum-of-squares curve (C) and confirmed by visual inspection of the HCl dendrogram (E) and good stability (low mean classification error) in 20-fold cross-validation (D).

Figure 7 – figure supplement 2. Impact of specific variables on the quality of participant clustering.

The clusters of participants clusters were defined with the optimally performing SOM/HCl algorithm with SMD observation dissimilarity measure as presented in **Figure 7** and **Figure 7 – figure supplement 1**. The impact of a particular clustering variable was determined by

comparing the ‘explained’ clustering variance (between-cluster to total sum-of-squares ratio) between the initial cluster structure and the structure with random re-sampling of the variable. Differences in the clustering variances for the most influential clustering variables (Δ clustering variance > 0) are presented in the plot. Q1, Q3: 1st, 3rd quartile of anti-S1/S2 IgG titer, CKD: chronic kidney disease, GI: gastrointestinal, CVD: cardiovascular disease, PD: pulmonary disease, GITD: gastrointestinal disease.

Figure 8. Frequency of persistent radiological lung abnormalities, functional lung impairment and symptoms in the participant clusters.

The clusters of study participants were defined by non-lung function and non-CT features as presented in **Figure 7**. Frequencies of outcome variables at the 180-day follow-up visit (mild [severity score ≤ 5], moderate-to-severe lung CT abnormalities [severity score > 5]), functional lung impairment and persistent symptoms) were compared between the low-risk (LR), intermediate-risk (IR) and high-risk (HR) participant clusters by χ^2 test corrected for multiple testing with the Benjamini-Hochberg method. P values and numbers of participants assigned to the clusters are indicated in the plots.

(A) Frequencies of the outcome features in the participant clusters.

(B) Frequencies of specific symptoms in the participant clusters.

Figure 8 – figure supplement 1. Risk of radiological lung abnormalities at the 180-day follow-up in the participant clusters.

The clusters of participants were defined by non-lung function and non-CT clinical features of acute COVID-19 and early convalescence (60-day follow-up visit, **Appendix 1 – table 1**) with the optimally performing HCl algorithm with SMD observation dissimilarity measure as presented in **Figure 7** and **Figure 7 – figure supplement 1**.

(A) Distribution of mild, moderate, severe and critical acute COVID-19 cases in the participant clusters. Significance of the distribution differences was assessed with χ^2 test. The numbers of participants assigned to the clusters are indicated under the plot.

(B) Association of the participant cluster assignment (LR: low-risk, IR: intermediate-risk, HR: high-risk cluster) with the risk of any lung CT abnormalities and moderate-to-severe lung CT abnormalities (severity score > 5) at the 180-day follow-up visit was investigated by logistic modeling with and without inclusion of the acute COVID-19 severity effect (severity-adjusted). Odds ratio (OR) significance was determined by Wald Z test and corrected for multiple testing with the Benjamini-Hochberg method. OR with 95% confidence intervals are presented in Forest plots. Numbers of complete observations, outcome events, participants in the clusters and the acute COVID-19 severity subsets are indicated under the plot.

Figure 9. Prediction of persistent radiological lung abnormalities, functional lung impairment and symptoms by machine learning algorithms.

Single machine learning classifiers (C5.0, RF: random forests, SVM-R: support vector machines with radial kernel, NNet: neural network, glmNet: elastic net) and their ensemble (Ens) were trained in the cohort data set with 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 or at the 60-day follow-up visit (**Appendix 1 – table 1**) for predicting outcome variables at the 180-day follow-up visit (any lung CT abnormalities, moderate-to-severe lung CT abnormalities [severity score > 5], functional lung impairment and persistent symptoms) (**Appendix 1 – table 4**). The prediction accuracy was verified by repeated 20-fold cross-validation (5 repeats). Receiver operating characteristics (ROC) of the algorithms in the cross-validation are presented: area under the curve (AUC), sensitivity (Sens) and specificity (Spec) (**Appendix 1 – table 5**). The numbers of complete observations and outcome events are indicated under the plots.

Figure 9 – figure supplement 1. Correlation of the machine learning algorithm prediction accuracy.

Machine learning classifiers (C5.0, RF: random forests, SVM-R: support vector machines with radial kernel, NNet: neural network, glmNet: elastic net) were trained in the cohort data set with 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 or at the early 60-day follow-up visit (**Appendix 1 – table 1**) for predicting outcome variables at the

180-day follow-up visit (any lung CT abnormalities, moderate-to-severe lung CT abnormalities [severity score > 5], functional lung impairment and persistent symptoms) (**Figure 9, Appendix 1 – table 4**). The prediction accuracy was verified by repeated 20-fold cross-validation (5 repeats). Pearson’s correlation coefficients of the classifier prediction accuracy in the cross-validation are presented as heat maps. Numbers of complete observations and outcome events are indicated under the plots.

Figure 9 – figure supplement 2. Machine learning model ensembles.

Single machine learning classifiers (C5.0, RF: random forests, SVM-R: support vector machines with radial kernel, NNet: neural network, glmNet: elastic net) were trained as shown in **Figure 9**. The model ensembles based on the single classifiers were constructed with the elastic net (glmNet) procedure (**Appendix 1 – table 4**). Elastic net regression coefficients (β) are presented in the plots. Point and text color correspond to the β value. Numbers of complete observations and outcome events are indicated under the plots.

Figure 9 – figure supplement 3. Prediction of persistent radiological lung abnormalities, functional lung impairment and symptoms by machine learning algorithms in the training data sets.

Single machine learning classifiers (C5.0, RF: random forests, SVM-R: support vector machines with radial kernel, NNet: neural network, glmNet: elastic net) and their ensembles were trained as shown in **Figure 9**. Performance of the classifiers in the training data sets was investigated by receiver operating characteristic (ROC) of the algorithms (AUC: area under the curve, Sens: sensitivity, Spec: specificity, **Appendix 1 – table 5**). Numbers of complete observations and outcome events are indicated under the plots.

Figure 9 – figure supplement 4. Variable importance statistics for prediction of lung CT abnormalities at the 180-day follow-up by machine learning classifiers.

C5.0, random forests (RF) and elastic net (glmNet) classifiers were trained as presented in **Figure 9** for prediction of any lung CT abnormalities at the 180-day follow-up visit. Variable importance measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet: absolute value of the regression coefficient β) for the 10 most influential explanatory variables are presented. CKD: chronic kidney disease, Q1, Q4: 1st, 4th quartile of anti-S1/S2 IgG titer, PD: pulmonary disease, CKD: chronic kidney disease.

Figure 9 – figure supplement 5. Variable importance statistics for prediction of moderate-to-severe lung CT abnormalities at the 180-day follow-up by machine learning classifiers.

C5.0, random forests (RF) and elastic net (glmNet) classifiers were trained as presented in **Figure 9** for prediction of moderate-to-severe lung CT abnormalities (severity score > 5) at the 180-day follow-up visit. Variable importance measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet: absolute value of the regression coefficient β) for the 10 most influential explanatory variables are presented. PD: pulmonary disease, GITD: gastrointestinal disease, Q1, Q2, Q4: 1st, 2nd, 4th quartile of anti-S1/S2 IgG titer.

Figure 9 – figure supplement 6. Variable importance statistics for prediction of functional lung impairment at the 180-day follow-up by machine learning classifiers.

C5.0, random forests (RF) and elastic net (glmNet) classifiers were trained as presented in **Figure 9** for prediction of functional lung impairment at the 180-day follow-up visit. Variable importance measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet: absolute value of the regression coefficient β) for the 10 most influential explanatory variables are presented. CKD: chronic kidney disease, Q1, Q2: 1st, 2nd quartile of anti-S1/S2 IgG titer.

823

824 **Figure 9 – figure supplement 7. Variable importance statistics for**
825 **prediction of persistent symptoms at the 180-day follow-up by**
826 **machine learning classifiers.**

827 C5.0, random forests (RF) and elastic net (glmNet) classifiers were trained as presented in **Figure**
828 **9** for prediction of persistent symptoms at the 180-day follow-up visit. Variable importance
829 measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet:
830 absolute value of the regression coefficient β) for the 10 most influential explanatory variables are
831 presented. CVD: cardiovascular disease, GITD: gastrointestinal disease, COPD: chronic obstructive
832 lung disease.

833

834 **Figure 10. Performance of the machine learning ensemble classifier**
835 **in mild-to-moderate and severe-to-critical COVID-19 convalescents.**

836 The machine learning classifier ensemble (Ens) was developed as presented in **Figure 9**. Its
837 performance at predicting outcome variables at the 180-day follow-up visit (any CT lung
838 abnormalities, moderate-to-severe lung CT abnormalities [severity score > 5], functional lung
839 impairment and persistent symptoms in the entire cohort, mild-to-moderate (outpatient or
840 hospitalized without oxygen) and severe-to-critical COVID-19 convalescents (oxygen therapy or
841 ICU) in repeated 20-fold cross-validation (5 repeats) was assessed by receiver operating
842 characteristic (ROC) (**Appendix 1 – table 6**). ROC curves and statistics (AUC: area under
843 the curve, Se: sensitivity, Sp: specificity) in the cross-validation are shown. Numbers of complete
844 observations and outcome events are indicated in the plots.

845

Appendix 1

Appendix 1 – table 1. Study variables.

Variable: variable name in the analysis pipeline, Reference time point: study visit, the variable was recorded at, Label: variable label in figures and tables.

Variable	Reference time point	Label	Variable type	Stratification cutoff
sex_male_V0	acute COVID-19	Male sex	explanatory	
obesity_rec_V0	acute COVID-19	Obesity	explanatory	BMI > 30 kg/m2
current_smoker_V0	acute COVID-19	Current smoker	explanatory	
smoking_ex_V0	acute COVID-19	Ex-smoker	explanatory	
CVDis_rec_V0	acute COVID-19	CVD	explanatory	
hypertension_rec_V0	acute COVID-19	Hypertension	explanatory	
PDIs_rec_V0	acute COVID-19	PD	explanatory	
COPD_rec_V0	acute COVID-19	COPD	explanatory	
asthma_rec_V0	acute COVID-19	Asthma	explanatory	
endocrine_metabolic_rec_V0	acute COVID-19	Metabolic disorders	explanatory	
hypercholesterolemia_rec_V0	acute COVID-19	Hypercholesterolemia	explanatory	
diabetes_rec_V0	acute COVID-19	Diabetes	explanatory	
CKDis_rec_V0	acute COVID-19	CKD	explanatory	
GITDis_rec_V0	acute COVID-19	GITD	explanatory	
malignancy_rec_V0	acute COVID-19	Malignancy	explanatory	
immune_deficiency_rec_V0	acute COVID-19	Immune deficiency	explanatory	
weight_change_rec_V0	acute COVID-19	Weight loss, acute COVID-19	explanatory	≥ 1 kg
dyspnoe_rec_V0	acute COVID-19	Dyspnea, acute COVID-19	explanatory	
cough_rec_V0	acute COVID-19	Cough, acute COVID-19	explanatory	
fever_rec_V0	acute COVID-19	Fever, acute COVID-19	explanatory	
night_sweat_rec_V0	acute COVID-19	Night sweat, acute COVID-19	explanatory	
pain_rec_V0	acute COVID-19	Pain, acute COVID-19	explanatory	
GI_sympt_rec_V0	acute COVID-19	GI symptoms, acute COVID-19	explanatory	
anosmia_rec_V0	acute COVID-19	Anosmia, acute COVID-19	explanatory	
ECOG_imp_rec_V0	acute COVID-19	Impaired performance, acute COVID-19	explanatory	ECOG ≥ 1

Variable	Reference time point	Label	Variable type	Stratification cutoff
sleep_disorder_rec_V0	acute COVID-19	Sleep disorders, acute COVID-19	explanatory	
treat_antiinfec_rec_V0	acute COVID-19	Anti-infectives, acute COVID-19	explanatory	
treat_antiplat_rec_V0	acute COVID-19	Anti-platelet, acute COVID-19	explanatory	
treat_anticoag_rec_V0	acute COVID-19	Anti-coagulatives, acute COVID-19	explanatory	
treat_immunosuppr_rec_V0	acute COVID-19	Immunosuppression, acute COVID-19	explanatory	
anemia_rec_V1	60-day follow-up	Anemia, 60-day visit	explanatory	Male: Hb < 14 g/dL, Female: Hb < 12 g/dL
ferr_elv_rec_V1	60-day follow-up	Elevated ferritin, 60-day visit	explanatory	Male: > 300 ng/mL, Female: > 150 ng/ml
NTelv_rec_V1	60-day follow-up	Elevated NTproBNP, 60-day visit	explanatory	> 125 pg/mL
Ddimerelv_rec_V1	60-day follow-up	Elevated D-dimer, 60-day visit	explanatory	> 500 pg/mL FEU
CRP_elv_rec_V1	60-day follow-up	Elevated CRP, 60-day visit	explanatory	> 0.5 mg/dL
IL6_elv_rec_V1	60-day follow-up	Elevated IL-6, 60-day visit	explanatory	> 7 pg/mL
iron_deficiency_30_rec_V1	60-day follow-up	Iron deficiency, 60-day visit	explanatory	TF-Saturation < 15%
age_65_V0	acute COVID-19	Age over 65	explanatory	> 65 years
hosp_7d_V0	acute COVID-19	Hospitalized >7 days, acute COVID-19	explanatory	> 7 days
comorb_present_V0	acute COVID-19	Any comorbidity	explanatory	> 0 comorbidities
comorb_3_V0	acute COVID-19	>3 comorbidities	explanatory	> 3 comorbidities
overweight_V0	acute COVID-19	Overweight or obesity	explanatory	BMI > 25 kg/m2
sympt_6_V0	acute COVID-19	>6 symptoms, acute COVID-19	explanatory	> 6 symptoms
sympt_present_V1	60-day follow-up	Persistent symptoms, 60-day visit	explanatory	> 0 symptoms at 180-day visit
ab_0_V1	60-day follow-up	Anti-S1/S2 IgG Q1, 60-day visit	explanatory	(0, 312] BAU/ml
ab_25_V1	60-day follow-up	Anti-S1/S2 IgG Q2, 60-day visit	explanatory	(312, 644] BAU/ml
ab_50_V1	60-day follow-up	Anti-S1/S2 IgG Q3, 60-day visit	explanatory	(644, 975] BAU/ml

Variable	Reference time point	Label	Variable type	Stratification cutoff
ab_75_V1	60-day follow-up	Anti-S1/S2 IgG Q4, 60-day visit	explanatory	> 975 BAU/ml
pat_group_G1_V0	acute COVID-19	Ambulatory, acute COVID-19	explanatory	
pat_group_G2_V0	acute COVID-19	Hospitalized, acute COVID-19	explanatory	
pat_group_G3_V0	acute COVID-19	Oxygen therapy, acute COVID-19	explanatory	
pat_group_G4_V0	acute COVID-19	ICU, acute COVID-19	explanatory	
CT_findings_V3	180-day follow-up	CT abnormalities at 180-day visit	outcome	
CT_sev_low_V3	180-day follow-up	CT Severity Score 1-5 at 180-day visit	outcome	
CTsevabove5_V3	180-day follow-up	CT Severity Score >5 at 180-day visit	outcome	
sympt_present_V3	180-day follow-up	Symptoms at 180-day visit	outcome	
lung_function_impaired_V3	180-day follow-up	Lung function impairment at 180-day visit	outcome	

850

851

852

Appendix 1 – table 2. Results of univariate risk modeling.

Outcome: outcome variable at the 180-days follow-up visit. Co-variate: explanatory variable,
Baseline: reference level of the explanatory variable, OR: odds ratios with 95% confidence intervals,
pFDR: significance p-value corrected for multiple testing with the Benjamini-Hochberg method
(FDR: false discovery rate). The table is available as a supplementary Excel file.

Appendix 1 – table 3. Feature cluster assignment scheme.

Cluster	Variable
Cluster #1	Male sex, CVD, Hypertension, Metabolic disorders, Anti-infectives, acute COVID-19, Elevated NTproBNP, 60-day visit, Elevated D-dimer, 60-day visit, Hospitalized >7 days, acute COVID-19, >3 comorbidities, Overweight.
Cluster #2	Obesity, Current smoker, Ex-smoker, PD, COPD, Asthma, Hypercholesterolemia, Diabetes, CKD, GITD, Malignancy, Immune deficiency, GI symptoms, acute COVID-19, Anosmia, acute COVID-19, Sleep disorders, acute COVID-19, Anti-platelet, acute COVID-19, Anti-coagulatives, acute COVID-19, Immunosuppression, acute COVID-19, Anemia, 60-day visit, Elevated ferritin, 60-day visit, Elevated CRP, 60-day visit, Elevated IL-6, 60-day visit, Iron deficiency, 60-day visit, Age over 65, >6 symptoms, acute COVID-19, Anti-S1/S2 IgG Q1, 60-day visit, Anti-S1/S2 IgG Q2, 60-day visit, Anti-S1/S2 IgG Q3, 60-day visit, Anti-S1/S2 IgG Q4, 60-day visit, Ambulatory, acute COVID-19, Hospitalized, acute COVID-19, Oxygen therapy, acute COVID-19, ICU, acute COVID-19, CT Severity Score 1-5 at 180-day visit, CT Severity Score >5 at 180-day visit, Lung function impairment at 180-day visit
Cluster #3	Weight loss, acute COVID-19, Dyspnea, acute COVID-19, Cough, acute COVID-19, Fever, acute COVID-19, Night sweat, acute COVID-19, Pain, acute COVID-19, Impaired performance, acute COVID-19, Any comorbidity, Persistent symptoms, 60-day visit, Symptoms at 180-day visit

862 **Appendix 1 – table 4. Development of machine learning models.**

863 Outcome: outcome variable at the 180-days follow-up visit.

Outcome	Classifier type	Caret method	Description	Package	Optimal arguments
CT abnormalities at 180-day visit	model	C5.0	C5.0	C50	trials = 10, model = tree, winnow = FALSE
		rf	Random Forest	randomForest	mtry = 27
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.0105, C = 0.5
		nnet	Neural Network	nnet	size = 1, decay = 0
		glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.000431
	ensemble	glmnet	glmnet	glmnet	alpha = 1, lambda = 0.0523
CT Severity Score >5 at 180-day visit	model	C5.0	C5.0	C50	trials = 1, model = rules, winnow = TRUE
		rf	Random Forest	randomForest	mtry = 52
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.00979, C = 0.5
		nnet	Neural Network	nnet	size = 1, decay = 0.1
		glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.0419
	ensemble	glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.00379
Symptoms at 180-day visit	model	C5.0	C5.0	C50	trials = 1, model = tree, winnow = FALSE
		rf	Random Forest	randomForest	mtry = 27
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.0109, C = 1
		nnet	Neural Network	nnet	size = 3, decay = 0.1
		glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.000247
	ensemble	glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.0167
Lung function impairment at 180-day visit	model	C5.0	C5.0	C50	trials = 1, model = rules, winnow = FALSE

Outcome	Classifier type	Caret method	Description	Package	Optimal arguments
		rf	Random Forest	randomForest	mtry = 52
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.0108, C = 0.5
		nnet	Neural Network	nnet	size = 1, decay = 0.1
		glmnet	glmnet	glmnet	alpha = 0.55, lambda = 0.0341
	ensemble	glmnet	glmnet	glmnet	alpha = 0.55, lambda = 0.0387

864

865

866 **Appendix 1 – table 5. Performance of machine learning classifiers.**

867 Outcome: outcome variable at the 180-days follow-up visit. The table is available as a
868 supplementary Excel file.

869

870 **Appendix 1 – table 6. Performance of machine learning classifiers in**
871 **the acute COVID-19 severity strata.**

872 Outcome: outcome variable at the 180-days follow-up visit, Cohort subset: cohort acute COVID-19
873 severity strata (mild-moderate: outpatient or hospitalized without oxygen, severe-critical: oxygen
874 therapy or ICU). The table is available as a supplementary Excel file.

A**Persistent symptoms** **B****Symptom recovery**

Pulmonary recovery

60 days

n = 126

100 days

n = 133

180 days

n = 107

60 days

n = 126

100 days

n = 133

180 days

n = 107

A **B**

A **B**

Lung function impairment at 180-day visit

A **B****CT Severity Score 1-5 at 180-day visit**

Cluster #2

CT Severity Score >5 at 180-day visit

Cluster #2

Lung function impairment at 180-day visit

Cluster #2

Symptoms at 180-day visit

Cluster #3

A

Clustering variance
Between/total sum of squares

B

Cluster stability
20-fold cross-validation

C

Optimal cluster number
PAM clustering, SMD distance

D

Cluster stability vs cluster number
20-fold cross-validation

Distances between observations

A

Participant clustering

Cluster ID

LR

IR

HR

LR: n = 60

IR: n = 31

HR: n = 42

B

The most relevant clustering features

A**Clustering variance**

Between/total sum of squares

B**Cluster stability**

20-fold cross-validation

C**SOM clustering: training**

SMD

D**Opimal number of SOM node clusters**

HCI, Euncclidean distance, Ward.D2 method

E**Opimal number of SOM node clusters** **F****Cluster stability vs cluster number**

20-fold cross-validation

Importance of clustering variables

A**CT abnormalities at 180-day visit**

p = 2e-05

Lung function imp. at 180-day visit

ns (p = 0.057)

Symptoms at 180-day visit

ns (p = 0.48)

B**Symptoms present at 180-day visit**

A**Acute COVID-19 severity** $p = 7.1e-21$

LR : n = 60, IR : n = 31, HR : n = 42

B**Participant clustering and CT abnormality risk**

Model type — severity-adjusted — unadjusted

complete: CT: n = 109, lung function: n = 111

CT abnormalities: n = 49, CT Severity Score > 5: n = 21, Lung function impaired: n = 21

LR : n = 60, IR : n = 31, HR : n = 42

Outpatient: n = 27, Hospitalized: n = 31, Hospitalized oxygen therapy: n = 33, Hospitalized ICU: n = 18

CT abnormalities at 180-day visit

Total: n = 109, Events: n = 49

CT Severity Score >5 at 180-day visit

Total: n = 109, Events: n = 21

Lung function impairment at 180-day visit

Total: n = 111, Events: n = 38

Symptoms at 180-day visit

Total: n = 133, Events: n = 65

—●— AUC
—●— Sens
—●— Spec

CT abnormalities at 180-day visit

SVM-R	0.44	0.36	0.51	0.76	1
RF	0.5	0.37	0.49	1	0.76
NNet	0.29	0.42	1	0.49	0.51
glmNet	0.4	1	0.42	0.37	0.36
C5.0	1	0.4	0.29	0.5	0.44
	C5.0	glmNet	NNet	RF	SVM-R

Total: n = 109, Events: n = 49

CT Severity Score >5 at 180-day visit

SVM-R	0.71	0.75	0.48	0.74	1
RF	0.85	0.6	0.53	1	0.74
NNet	0.48	0.73	1	0.53	0.48
glmNet	0.6	1	0.73	0.6	0.75
C5.0	1	0.6	0.48	0.85	0.71
	C5.0	glmNet	NNet	RF	SVM-R

Total: n = 109, Events: n = 21

Symptoms at 180-day visit

SVM-R	0.28	0.39	0.33	0.46	1
RF	0.52	0.33	0.37	1	0.46
NNet	0.3	0.4	1	0.37	0.33
glmNet	0.17	1	0.4	0.33	0.39
C5.0	1	0.17	0.3	0.52	0.28
	C5.0	glmNet	NNet	RF	SVM-R

Total: n = 133, Events: n = 65

Lung function impairment at 180-day visit

SVM-R	0.57	0.73	0.5	0.71	1
RF	0.62	0.78	0.65	1	0.71
NNet	0.37	0.66	1	0.65	0.5
glmNet	0.55	1	0.66	0.78	0.73
C5.0	1	0.55	0.37	0.62	0.57
	C5.0	glmNet	NNet	RF	SVM-R

Total: n = 111, Events: n = 38

CT abnormalities at 180-day visit

Total: n = 109, Events: n = 49

CT Severity Score >5 at 180-day visit

Total: n = 109, Events: n = 21

Symptoms at 180-day visit

Total: n = 133, Events: n = 65

Lung function impairment at 180-day visit

Total: n = 111, Events: n = 38

CT abnormalities at 180-day visit

Total: n = 109, Events: n = 49

CT Severity Score >5 at 180-day visit

Total: n = 109, Events: n = 21

Lung function impairment at 180-day visit

Total: n = 111, Events: n = 38

Symptoms at 180-day visit

Total: n = 133, Events: n = 65

—●— AUC
—●— Sens
—●— Spec

A**C5.0: CT abnormalities at 180-day visit** **B****RF: CT abnormalities at 180-day visit** **C****glmNet: CT abnormalities at 180-day visit**

A**C5.0: CT Severity Score >5 at 180-day visit** **B****RF: CT Severity Score >5 at 180-day visit** **C****glmNet: CT Severity Score >5 at 180-day visit**

A**C5.0: Lung function impairment at 180-day visit** **B****RF: Lung function impairment at 180-day visit** **C****glmNet: Lung function impairment at 180-day visit**

A C5.0: Symptoms at 180-day visit

B RF: Symptoms at 180-day visit

C glmNet: Symptoms at 180-day visit

CT abnormalities at 180-day visit: Ens

CT Severity Score >5 at 180-day visit: Ens

Lung function impairment at 180-day visit: Ens

Symptoms at 180-day visit: Ens

—●— cohort —●— mild – moderate —●— severe – critical