Brain, Behavior, & Immunity - Health 54 (2026) 101267

Contents lists available at ScienceDirect

Brain, Behavior, & Immunity - Health

journal homepage: www.editorialmanager.com/bbih/default.aspx

An externally validated machine learning algorithm for predicting mental and physical health outcomes three months post-hospitalization for severe viral infection with SARS-CoV-2

Katharina Schultebraucks [a,b,*], Sapir Gershov [a], Felix Fischer [c,z], Katja Wingenfeld [d], Sein Schmidt [e], Sarah Steinbrecher [f], Thomas Zoller [f], Fridolin Steinbeis [f], Sina M. Pütz [g], Jürgen Deckert [h], Margarete Scherer [i,w], Isabel Bröhl [i], Patricia Wagner [g], Katharina S. Appel [i], Mirjam Kohls [j], Steffi Jiru-Hillmann [j], Matthias Nauck [k], Bettina Lorenz-Depiereux [l], Sabine Blaschke [m], Anna Muzalyova [n], Christoph Stellbrink [o], Anke Steinmetz [p], Marylyn Martina Addo [q], Edgar Dahl [r], Markus Zettler [s], Stefan Hansch [t], Andreas Dinkel [u], Verena Keitel [v], Maria J.G.T. Vehreschild [w], Jörg J. Vehreschild [g,l,w], Friedemann Paul [x], Martin Witzenrath [f,y], Matthias Rose [c,z], Christian Otte [d,z]

[a] Department of Psychiatry, NYU Grossman School of Medicine, New York, USA
[b] Division of Healthcare Delivery Science, Department of Population Health, NYU Grossman School of Medicine, New York, USA
[c] Charité – Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt Universität zu Berlin, Medizinische Klinik Mit Schwerpunkt für Psychosomatik, Center for Patient-Centered Outcomes Research, Germany
[d] Department of Psychiatry and Psychotherapy, Campus Benjamin Franklin, Charité–Universitätsmedizin, Berlin, Germany
[e] Clinical Trial Unit of the Clinical Study Center (CSC), Berlin Institute of Health (BIH), Germany
[f] Department of Infectious Diseases, Respiratory Medicine and Critical Care, Charité – Universitätsmedizin Berlin, Germany
[g] University of Cologne, Faculty of Medicine and University Hospital Cologne, Department I of Internal Medicine, Center for Infectious Diseases, Cologne, Germany
[h] Department of Psychiatry, University Hospital Würzburg, Germany
[i] Institute for Digital Medicine and Clinical Data Science, Faculty of Medicine, Goethe University Frankfurt, Germany
[j] Institute for Clinical Epidemiology and Biometry, University Hospital Würzburg, Germany
[k] Klinische Chemie und Laboratoriumsmedizin, University Medicine Greifswald, Germany
[l] Institute of Epidemiology, Research Unit Molecular Epidemiology, Helmholtz Munich, Germany
[m] Emergency Department, University Medical Center Göttingen, Germany
[n] Institute for Digital Medicine, University of Augsburg, Germany
[o] Department of Cardiology, Universitäty Medical Center Bielefeld, Germany
[p] Department of Orthopaedics, Trauma and Rehabilitation Medicine, Universitätsmedizin Greifswald, Germany
[q] Infection Research and Vaccine Development, University Medical Center Hamburg-Eppendorf, Germany
[r] RWTH Centralized Biomaterial Bank (RWTH CBMB) at the Institute of Pathology, RWTH Aachen University, Germany
[s] Institute of Infectiology, University Medical Center Essen, Germany
[t] Department for Infectious Diseases and Infection Control, University Hospital Regensburg, Germany
[u] Department of Psychosomatic Medicine and Psychotherapy, Klinikum rechts der Isar, Technichal University of Munich, Germany
[v] Clinic for Gastroenterology, Hepatology and Infectiology, Medical Faculty, Otto von Guericke University Magdeburg, Germany
[w] Goethe University Frankfurt, University Hospital Frankfurt, Department II of Internal Medicine, Infectious Diseases, Frankfurt am Main, Germany
[x] Experimental and Clinical Research Center (ECRC), Charité – Universitätsmedizin Berlin, Germany
[y] German Center for Lung Research (DZL), Partner Site Berlin, Germany
[z] German Center for Mental Health, Partner Site Berlin/Potsdam, Germany

ARTICLE INFO

Keywords:
SARS-CoV-2
PROMIS

ABSTRACT

Many individuals hospitalized due to severe viral infections develop mental and physical sequelae, which could potentially be prevented by targeted interventions for those at risk. Our goal was to develop and externally validate an algorithm for predicting mental and physical symptoms after SARS-CoV-2 hospitalization utilizing

This article is part of a special issue entitled: Post-COVID19 condition published in Brain, Behavior, & Immunity - Health.
* Corresponding author. Department of Psychiatry, New York University Grossman School of Medicine, One Park Avenue, New York, NY, 10016, USA.
E-mail address: katharina.schultebraucks@nyulangone.org (K. Schultebraucks).

https://doi.org/10.1016/j.bbih.2026.101267
Received 23 July 2025; Received in revised form 6 May 2026; Accepted 19 May 2026
Available online 20 May 2026

Early risk prediction
Machine learning
Clinical decision-making

routinely collected clinical data. Participants were included from two independent samples of the German National Pandemic Cohort Network (NAPKON): a model development sample (SUEP; N = 451; mean age: 55.6 ± 15.3; 36.2% female) and an external validation sample (HAP: N = 158; mean age: 55.1 ± 12.1; 39.9% female). Machine learning models leveraging demographic, clinical and biological variables collected at the time of admission were employed to predict Patient-Reported Outcomes Measurement Information System scores (PROMIS) across 7 domains (physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles and activities, and pain) three months after SARS-CoV-2 hospitalization. Shapley Additive exPlanation values were used to provide interpretable information about key predictive factors. Approximately 15-20% of participants reported moderate to severe impairment in at least one PROMIS domain three months after hospitalization. For the mental health composite score, the best-performing model achieved RMSE = 1.833 ± 0.341 and R^2 = 0.927 ± 0.031 in SUEP and RMSE = 3.131 and R^2 = 0.893 in HAP. For the physical health composite, the best-performing model achieved RMSE = 2.908 ± 0.703 and R^2 = 0.824 ± 0.052 in SUEP and RMSE = 3.019 and R^2 = 0.850 in HAP. Furthermore, the models achieved high predictive performance across all individual PROMIS domain scores in both samples. We provide an externally validated methodology for accurately predicting mental and physical symptomatology following hospitalization due to a severe viral infection. This approach may facilitate the development of a brief risk stratification tool at the point of hospitalization, enabling early identification of at-risk patients, improving the prediction accuracy of subsequent psychological and physical sequelae, and supporting timely preventive interventions.

1. Introduction

Severe viral infections requiring hospitalization can lead to significant downstream health issues that may become chronic if not adequately treated. A substantial proportion of intensive care unit (ICU) survivors experience persistent physical, cognitive, and mental health impairments, including anxiety, depression, and post-traumatic stress disorder (Davydow et al., 2008; Geense et al., 2021; Needham et al., 2012; Pandharipande et al., 2013). This pattern is particularly evident among patients hospitalized with SARS-CoV-2, where persistent post-acute symptoms, commonly referred to as Long COVID or post-acute sequelae of SARS-CoV-2 (PASC), affect an estimated 10–30% of non-hospitalized and up to 50–70% of hospitalized patients (Ballering et al., 2022; Carfì et al., 2020; Huang et al., 2023; Weich, 2022; Zhao et al., 2023). Long COVID encompasses a broad spectrum of physical, cognitive, and psychological symptoms, with fatigue, muscle weakness, cognitive impairment, dyspnea, anxiety, and depression among the most frequently reported (Davis et al., 2023; Huang et al., 2022; Nalbandian et al., 2021; Taquet et al., 2021). .These symptoms frequently co-occur, suggesting shared pathophysiological mechanisms involving persistent inflammation, immune dysregulation, and endothelial dysfunction (Phetsouphanh et al., 2022; Su et al., 2022).

Early prediction of post-acute mental and physical symptoms is critical for several reasons. Post-acute symptoms significantly impair patients' quality of life (Davydow et al., 2008), and early identification of at-risk individuals enables timely interventions that may mitigate symptom severity and duration (Needham et al., 2012). Furthermore, accurate risk prediction facilitates more effective healthcare resource allocation (Herridge et al., 2011) and supports the development of tailored rehabilitation programs targeting specific physical and cognitive impairments, thereby improving recovery outcomes (Pandharipande et al., 2013). This is particularly relevant in the context of Long COVID, where the heterogeneity of symptom profiles and trajectories underscores the need for individualized, data-driven risk stratification approaches (Sudre et al., 2021; Whitaker et al., 2022).

Identifying at-risk individuals is a multifaceted challenge requiring consideration of diverse risk factors. Demographic, clinical, and biological variables play significant roles in determining post-viral health outcomes for patients recovering from severe viral infections (Cao, 2020; Carfì et al., 2020; Huang et al., 2023; Mazza et al., 2020; Peluso et al., 2021; Phetsouphanh et al., 2022). Emerging Long COVID evidence further highlights that female sex, higher acute-phase symptom burden, pre-existing comorbidities, and elevated inflammatory markers are associated with an increased risk of developing persistent post-acute sequelae (Subramanian et al., 2022; Thompson et al., 2022; Xie et al., 2022). While several prediction models for post-COVID outcomes have been proposed, including externally validated algorithms for hospitalization, mortality, and persistent symptomatology (Hippisley-Cox et al., 2021; Pfaff et al., 2022; Sudre et al., 2021), a critical gap remains in models that simultaneously predict multiple mental and physical health domains using routinely available clinical data, thereby providing a comprehensive risk profile to inform targeted, domain-specific preventive interventions.

The primary aim of this study was to develop and externally validate a predictive model for early identification of individuals at risk of developing mental and physical health sequelae after hospitalization due to a severe viral infection. As a first application, the model was developed and validated in patients hospitalized with SARS-CoV-2. By leveraging demographic, clinical, and biological variables collected during the initial hospitalization, we aimed to predict which patients would develop psychological distress and physical sequelae three months after infection. This three-month window provides a clinically actionable opportunity for implementing preventive interventions and mitigating the burden of post-acute sequelae.

2. Methods

2.1. Study design

Participants were included from two independent samples recruited from NAPKON, the German National Pandemic Cohort Network (Schons et al., 2022). NAPKON is Germany's most comprehensive SARS-CoV-2 cohort, initiated in July 2020 as part of the Network University Medicine. Its primary goal is to establish standardized, high-quality data and biosample collection from patients, citizens, and controls with comparator respiratory infections to support the SARS-CoV-2 pandemic and future pandemics.

NAPKON comprises three parallel and complementary cohort platforms: the Cross-Sectoral Platform (SUEP), the High-Resolution Platform (HAP), and the Population-Based Platform (POP). In the current study, we are including only adult participants (≥18 years of age) from SUEP and HAP who were hospitalized due to severe SARS-CoV-2 illness and for whom 3-month Patient-Reported Outcomes were available. The SUEP cohort recruited SARS-CoV-2-infected in- and outpatients of all ages across all departments, collecting primary health record data, basic clinical phenotyping, biosamples, and patient interviews. The HAP cohort particularly focused on adult SARS-CoV-2-positive inpatients with severe SARS-CoV-2 illness. It extends data and biosample collection by adding additional clinical examinations, cytokine profiling, and standardized imaging, conducted at 10 German university hospitals.

Both Platforms (SUEP and HAP) are described in greater detail in Hopff et al. (2024) and Steinbeis et al. (2024). For NAPKON-SUEP, a primary ethics vote was obtained at the Ethics Committee of the Department of Medicine at Goethe University Frankfurt (local ethics ID approval 20-924). All further study sites received their local ethics votes at the respective ethics commissions. The NAPKON-SUEP is registered at ClinicalTrials.gov (Identifier: NCT04768998). For NAPKON-HAP, the study protocol and its amendments were reviewed and approved by the Charité Ethics Committee (EA2/066/20, EA2/226/21) as well as local ethics committees at each participating study center.

2.2. Primary outcomes

We assessed patients' self-reported psychological distress and physical symptoms three months after SARS-CoV-2 diagnosis using PROMIS-57 Profile questionnaire (Cella et al., 2019) from the Patient-Reported Outcomes Measurement Information System (PROMIS).

The PROMIS-57 Profile comprises 57 items distributed across seven domains: physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles and activities, and pain interference, each assessed by 8 items, supplemented by a single item measuring pain intensity (Cella et al., 2010; Rothrock et al., 2010). For each domain, item responses were converted to standardized PROMIS T-scores using the established scoring framework, in which scores are calibrated to a reference metric based on the US general population (mean = 50, SD = 10). In addition, we derived two continuous composite scores, a physical health and a mental health summary score, following the methodology described by Hays et al. (2018). These summary scores synthesize information across domain-level T-scores and provide complementary endpoints that capture broader patterns of physical and psychological health impact while maintaining interpretability on the same standardized metric.

2.3. Predictor variables

Predictor variables comprised demographic, clinical and biological variables routinely collected at the time of admission of a SARS-CoV-2-related encounter. These variable categories were selected based on converging evidence that demographic, clinical, and biological factors play significant roles in determining health outcomes following severe viral infections (Cao, 2020; Carfì et al., 2020; Huang et al., 2023; Mazza et al., 2020; Peluso et al., 2021; Phetsouphanh et al., 2022) and have been identified as relevant predictors of post-acute sequelae. Critically, the selection was further guided by the pragmatic requirement that all predictors be readily available within EMR systems during routine clinical workflows, ensuring that the resulting model is directly implementable in real-world acute care settings without imposing additional data collection burden. A comprehensive list of all included predictor variables, with descriptive statistics for both the SUEP and HAP samples, is provided in Supplementary Table 1. Additionally, Supplementary Table 2 presents the level of missingness of each variable in the dataset. To ensure cross-cohort harmonization, predictor definitions, units, and category mappings were systematically aligned across SUEP and HAP prior to modeling.

2.4. Statistical analyses

Data pre-processing. To ensure methodological rigor and prevent information leakage, all preprocessing steps were derived exclusively from the SUEP training data within each cross-validation iteration and subsequently applied without modification to the corresponding held-out SUEP test fold and to the HAP external validation sample. Within each training fold, features were excluded if they exhibited greater than 40% missingness or near-zero variance (<0.001). This threshold was chosen a priori as a pragmatic compromise between preserving clinically relevant routinely collected variables and reducing dependence on heavily imputed predictors, for which imputation estimates may become less robust as the proportion of missing observations increases. This approach is consistent with recommendations in the clinical prediction modeling literature (Madley-Dowd et al., 2019; Moons et al., 2015), which emphasize that rigid missingness thresholds should not be applied mechanistically but should balance data availability against imputation reliability. For the remaining features, missing values were imputed using a multivariate chained equations approach implemented in a round-robin fashion (IterativeImputer, scikit-learn v1.5.2). The imputation model was fitted solely on the training fold data and applied to both the corresponding test fold and the HAP sample, ensuring that no information from the evaluation data influenced the imputation process. Across all included variables, the median proportion of missing data was 4.43% (range: 0.22–30%) in the SUEP sample and 10.12% (range: 1.27–34.65%) in the HAP sample.

Model development. The predictive models were developed and tuned using only the SUEP sample and trained separately for each outcome. Due to their robustness and capability to handle various data types, we utilized the following regression models: (1) Gradient Boosting Regressor (GBR) (Friedman, 2001); (2) Histogram-based Gradient Boosting Regression Tree (HistGB) (Ke et al., 2017); (3) Random Forest Regressor (Breiman, 2001); (4) Extra-trees Regressor (Geurts et al., 2006). All models were incorporated in Python (v3.9, 'scikit-learn' library version 1.5.2). To guard against overfitting, model selection and performance were evaluated using a 5-fold nested cross-validation.

Model's hyperparameter optimization was performed via grid search on the inner folds, optimizing Negative Mean Squared Error (NMSE), and each model's performance was evaluated on the outer folds.

Model evaluation. The model's predictive performance was evaluated for both samples using Root Mean Squared Error (RMSE) and R^2. To examine generalizability, the final SUEP-trained models were applied once to the external validation sample HAP to ensure our findings are robust within the initial study cohort and applicable to a broader population. Additionally, we evaluated the calibration of the best-performing model for each PROMIS domain using out-of-fold predictions from nested cross-validation. Agreement between predicted and observed PROMIS scores was assessed visually using calibration plots and quantified by estimating calibration intercepts and slopes from linear regression of observed outcomes on predicted values.

Variable importance. To interpret the model's predictions and understand the importance of features, we utilized SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). This approach allowed us to rank the contributions of individual features to the model's output. We used TreeExplainer with a background sample size of 200 and summarized global importance as the mean absolute SHAP value.

Subgroup performance and bias assessment. To assess whether the model performs equitably across clinically and demographically relevant subpopulations, we conducted a prespecified subgroup performance analysis. Predictive performance was evaluated separately within strata defined by sex (female vs. male), vaccination status (vaccinated vs. unvaccinated), smoking status (current smoker vs. non-smoker), acute infection severity (without complications, with complications, critical complications), and country of birth (Germany vs. other). For each stratum, we computed R^2, RMSE, and mean signed error (bias; $\text{mean}(\hat{y} - y)$), where negative values indicate systematic under-prediction and positive values indicate systematic over-prediction.

All analyses were conducted and reported in accordance with the TRIPOD-AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis - Artificial Intelligence) guidelines for clinical prediction models that use machine learning methods (Collins et al., 2015). A completed TRIPOD-AI checklist detailing how each reporting item is addressed in the manuscript is provided in the Supplementary Material.

3. Results

Descriptive statistics of sample characteristics of the model development sample (SUEP; N = 451) and the external validation sample

(HAP; N = 158) are presented in Table 1. Both samples differed significantly in several clinical and demographic characteristics. Regarding the severity of SARS-CoV-2, SUEP had more cases with no complications (68.03%), while HAP had more cases with complications (52.11%) and severe complications (8.45%) ($p \leq 0.001$). Additionally, the SARS-CoV-2 vaccination rate was significantly higher in SUEP (72.28%) compared to HAP (27.85%) ($p \leq 0.001$). Smoking status showed significant differences, with more non-smokers in SUEP (53.20% non-smokers and 40.64% former smokers) than in HAP (57.38%) ($p \leq 0.001$). Dyspnea was more frequent in SUEP (65.63%) compared to HAP (55.70%) ($p \leq 0.05$). Moreover, heart rate was higher in SUEP (81.22 ± 15.82) compared to HAP (76.04 ± 15.88) ($p \leq 0.001$).

Overall, among patients in both samples, a significant portion (around 15-20%) reported moderate or severe mental and physical sequelae three months after the viral infection in at least one PROMIS domain (Tables 2 and 3) (Cella et al., 2010; Rothrock et al., 2010).

Table 1
Demographic and Clinical Characteristics at the time of admission due to a SARS-CoV-2 infection.

Features		SUEP (n = 451)	HAP (n = 158)	P-value
Age (mean, range)		55.57 (18 - 91)	55.05 (20 - 75)	0.66
Gender (n, %)		166 (36.81%) Female	52 (32.91%) Female	1.00
Body Mass Index (mean, SD)		28.44 (6.54)	28.99 (5.60)	0.51
Severity of SARS-CoV-2 infection (n, %)	1 – Without complications	307 (68.03%)	62 (39.44%)	≤0.001
	2 – With complications	127 (28.12%)	82 (52.11%)	
	3 – Critical complications	17 (3.85%)	13 (8.45%)	
Smoking Status (n, %)	1 - Current Smoker	28 (6.16%)	No – 91 (57.38%)	≤0.001[a]
	2 - Former Smoker	183 (40.64%)		
	3 - Non-Smoker	240 (53.20%)		
Comorbid diseases (n, %)	Cardiovascular	230 (51.00%)	80 (50.63%)	0.82
	Chronic lung	93 (20.62%)	34 (21.52%)	
	Chronic kidney	59 (13.08%)	28 (17.72%)	
	Chronic liver	34 (7.54%)	10 (6.33%)	
	Rheumatological/ immunological	37 (8.20%)	10 (6.33%)	
	Diabetes mellitus	74 (16.41%)	38 (24.05%)	
	Solid tumor	62 (13.75%)	16 (10.13%)	
Fever (n, %)		284 (62.97%)	108 (68.35%)	0.35
Shortness of breath (dyspnea) (n, %)		296 (65.63%)	88 (55.70%)	0.047
SARS-CoV-2 vaccination (n, %)		326 (72.28%)	44 (27.85%)	≤0.001
Systolic blood pressure (mean, SD)		127.18 (18.80)	128.72 (17.83)	0.37
Diastolic blood pressure (mean, SD)		75.40 (11.46)	74.20 (10.47)	0.24
Heart rate (mean, SD)		81.22 (15.82)	76.04 (15.88)	≤0.001
Oxygen saturation (mean, SD)		95.18 (3.19)	94.64 (2.85)	0.056
Oxygen Support (n, %)		286 (63.41%)	110 (69.62%)	0.21

[a] Smoking status "Non-Smoker" and "Former Smoker" were combined to enable comparison, as the SUEP dataset had three categories while the HAP dataset had only two. This ensures a valid Chi-Square Test of Independence p-value.

Table 2
Model development sample (SUEP) Symptom Severity 3 months after infection. To standardize the interpretation of PROMIS scores across profile domains, we used the mean of the US general population (50) and a standard deviation of 10. Then, thresholds of 0.5 (mild), 1.0 (moderate), and 2.0 (severe) standard deviations were incorporated across domains (Cella et al., 2010; Rothrock et al., 2010). Physical and mental health summary scores were computed using PROMIS summary-scoring factor coefficients (z-score method) (Hays et al., 2018).

Symptom	No Significant Impairment N (%)	Mild N (%)	Moderate N (%)	Severe N (%)
Anxiety/Fear	313 (69.40%)	76 (16.85%)	57 (12.64%)	5 (1.11%)
Depression/ Sadness	329 (72.95%)	65 (14.41%)	53 (11.75%)	4 (0.89%)
Fatigue	310 (68.74%)	56 (12.42%)	72 (15.96%)	13 (2.88%)
Pain	257 (56.98%)	80 (17.74%)	98 (21.73%)	16 (3.55%)
Sleep	319 (70.73%)	62 (13.75%)	62 (13.75%)	8 (1.77%)
Social Roles	366 (81.15%)	73 (16.19%)	12 (2.66%)	0 (0%)
Physical Function	326 (72.28%)	107 (23.72%)	18 (3.99%)	0 (0%)
Physical Health Summary Score	325 (72.06%)	107 (23.73%)	19 (4.21%)	0 (0%)
Mental Health Summary Score	376 (83.37%)	69 (15.30%)	6 (1.33%)	0 (0%)

Table 3
Model external validation sample (HAP) Symptom Severity 3 months after infection. To standardize the interpretation of PROMIS scores across profile domains, we used the mean of the US general population (50) and a standard deviation of 10. Then, thresholds of 0.5 (mild), 1.0 (moderate), and 2.0 (severe) standard deviations were incorporated across domains (Cella et al., 2010; Rothrock et al., 2010). Physical and mental health summary scores were computed using PROMIS summary-scoring factor coefficients (z-score method) (Hays et al., 2018).

Symptom	No Significant Impairment N (%)	Mild N (%)	Moderate N (%)	Severe N (%)
Anxiety/Fear	110 (69.62%)	22 (13.92%)	16 (10.13%)	10 (6.33%)
Depression/ Sadness	112 (70.89%)	22 (13.92%)	22 (13.92%)	2 (1.27%)
Fatigue	104 (65.82%)	18 (11.39%)	32 (20.25%)	4 (2.53%)
Pain	86 (54.43%)	32 (20.25%)	38 (24.05%)	2 (1.27%)
Sleep	110 (69.62%)	22 (13.92%)	22 (13.92%)	4 (2.53%)
Social Roles	102 (64.56%)	24 (15.19%)	32 (20.25%)	0 (0%)
Physical Function	72 (45.57%)	30 (18.99%)	54 (34.18%)	2 (1.27%)
Physical Health Summary Score	76 (48.10%)	30 (18.99%)	48 (30.38%)	4 (2.53%)
Mental Health Summary Score	106 (67.09%)	24 (15.19%)	24 (15.19%)	4 (2.53%)

The predictive model based on demographic, clinical, and biological variables collected during SARS-CoV-2-related encounters showed high predictive performance for mental and physical sequelae in both the model development and external validation samples. For the mental composite score, the best-performing model was the ExtraTrees model, which achieved RMSE = 1.833 ± 0.341 and R^2 = 0.927 ± 0.031 in the model development sample (SUEP) and RMSE = 3.131 and R^2 = 0.893 in the external validation sample (HAP). For the physical composite score, the best-performing model was HistGB, which achieved

RMSE = 2.908 ± 0.703 and R^2 = 0.824 ± 0.052 in SUEP and RMSE = 3.019 and R^2 = 0.850 in HAP. Furthermore, the machine learning models achieved high predictive performance for each PROMIS domain in both the model development and external validation samples (Table 4). The models' performances and the final hyperparameters of the best-performing models are included in the supplementary material (Supplementary Tables 3 and 4). The calibration analyses demonstrated good agreement between predicted and observed PROMIS scores, with calibration slopes close to 1 and minimal bias (see Supplementary Fig. 1–7).

The variable importance for each PROMIS domain, estimated using SHAP values derived from the SUEP sample, is depicted in Figs. 1–7. Across all outcomes, we found that the presence of chronic diseases, shortness of breath, higher creatinine and leukocyte levels, older age, male gender, higher BMI, and lower blood pressure at the time of admission were among the most important predictors for more severe mental and physical sequelae three months after a SARS-CoV-2 infection.

The subgroup performance and bias assessment revealed broadly consistent predictive performance across all evaluated demographic and clinical strata in both the SUEP and HAP samples (Supplementary Table 5–6). Across most subgroups, the mean signed error was close to zero, indicating minimal systematic directional bias. Model accuracy, as measured by R^2 and RMSE, remained stable across sex, vaccination status, country of birth, and severity categories. The largest apparent deviations in performance were confined to smaller strata, notably current smokers and patients with critical acute infection severity, where reduced sample sizes are expected to produce less stable estimates. These findings suggest that the model performs equitably across the evaluated subpopulations, with no evidence of clinically meaningful differential bias.

4. Discussion

The present study aimed to develop and externally validate a predictive model for identifying individuals at risk of developing mental and physical sequelae three months after hospitalization due to severe infection with SARS CoV-2 (COVID-19). We found that 15-20% of participants reported moderate or severe sequelae in at least one domain of the PROMIS questionnaire. This finding is consistent with the broader Long COVID literature, which estimates that a substantial proportion of hospitalized patients develop persistent PASC, encompassing a wide spectrum of physical, cognitive, and psychological symptoms (Ballering et al., 2022; Davis et al., 2023; Nalbandian et al., 2021). Utilizing routinely collectable data integrated within EMR, the algorithm demonstrated high predictive performance for both mental and physical health outcomes. These findings underscore the potential of leveraging EMR data to facilitate early identification and intervention for patients at risk of subsequent sequelae, thereby improving patient outcomes and optimizing healthcare resource allocation.

Several prediction models for post-COVID outcomes have been developed in recent years, reflecting the urgent need for early risk stratification tools. For example, Sudre et al. (2021) identified early symptom predictors of Long COVID using data from a mobile application, while Su et al. (2022) leveraged multi-omic data to identify biological factors that anticipate post-acute sequelae. More recently, externally validated models have been developed using EMR data to predict specific post-COVID outcomes such as hospitalization, mortality, and persistent symptomatology (Hippisley-Cox et al., 2021; Pfaff et al., 2022). Our study extends this literature in several important ways. First, whereas most existing models focus on a single outcome domain (e.g., hospitalization or mortality), our approach simultaneously predicts multiple mental and physical health domains using the validated PROMIS framework, providing a comprehensive and clinically interpretable risk profile. Second, unlike models that rely on specialized research data or multi-omic profiling, our algorithm utilizes exclusively routinely collected EMR data available at the point of admission, enhancing its scalability and real-world implementability. Third, our model was externally validated in an independent cohort with substantially different clinical and demographic characteristics, providing a rigorous test of generalizability that many existing models lack (Siontis et al., 2015).

The predictive model exhibited robust performance in both the model development sample (SUEP) and the external validation sample (HAP). The high R^2 values for the mental health composite score (0.927 in SUEP and 0.893 in HAP) and physical health composite score (0.824 in SUEP and 0.850 in HAP) indicate that the model can accurately predict subsequent mental and physical health outcomes based on data collected during the initial SARS-CoV-2-related encounter. These results are particularly significant given the complexity and multifactorial nature of post-viral sequelae. Notably, the stable performance across two independent cohorts with substantially different clinical profiles, including disparate vaccination rates (72% vs. 28%), disease severity distributions, and smoking prevalence, provides strong evidence that the model captures generalizable predictive relationships rather than cohort-specific patterns. This is especially relevant in the context of Long COVID, where the heterogeneity of symptom presentations and recovery trajectories has posed a significant challenge for risk prediction efforts (Sudre et al., 2021; Whitaker et al., 2022).

While the observed R^2 values are high relative to those typically reported in psychological outcome prediction, several features of our analytic design mitigate concerns about overfitting. All model development and evaluation were conducted within a 5-fold nested cross-validation framework, ensuring strict separation between hyperparameter tuning and performance estimation. The low standard deviations across cross-validation folds indicate stable performance across data partitions. Most importantly, the model demonstrated consistently high predictive accuracy in the HAP external validation sample, which differed substantially from the development sample across multiple clinical and demographic characteristics. This robust generalization to an independent and heterogeneous cohort provides strong evidence that the reported performance reflects a genuine predictive signal rather than overfitting to the training data.

The model's ability to predict specific domains of the PROMIS, such

Table 4
Model results (mean and SD) per PROMIS domain for the model development sample (SUEP) and external validation sample (HAP).

		SUEP		HAP	
Symptom	Model	RMSE	R^2	RMSE	R^2
Anxiety/Fear	HistGB	1.552 ± 0.899	0.903 ± 0.046	2.321	0.884
Depression/Sadness	GBR	5.234 ± 1.886	0.757 ± 0.120	4.947	0.734
Fatigue	HistGB	1.560 ± 0.144	0.924 ± 0.021	1.646	0.907
Pain	Random Forest	3.737 ± 0.388	0.859 ± 0.020	3.899	0.846
Sleep	HistGB	1.981 ± 0.246	0.925 ± 0.0105	3.188	0.884
Social Roles	Random Forest	5.167 ± 0.213	0.734 ± 0.009	5.341	0.706
Physical Function	GBR	2.053 ± 1.366	0.915 ± 0.079	2.650	0.905
Meantal Health Summary Score	ExtraTrees	1.833 ± 0.341	0.927 ± 0.031	3.131	0.893
Physical Health Summary Score	HistGB	2.908 ± 0.703	0.824 ± 0.052	3.019	0.850

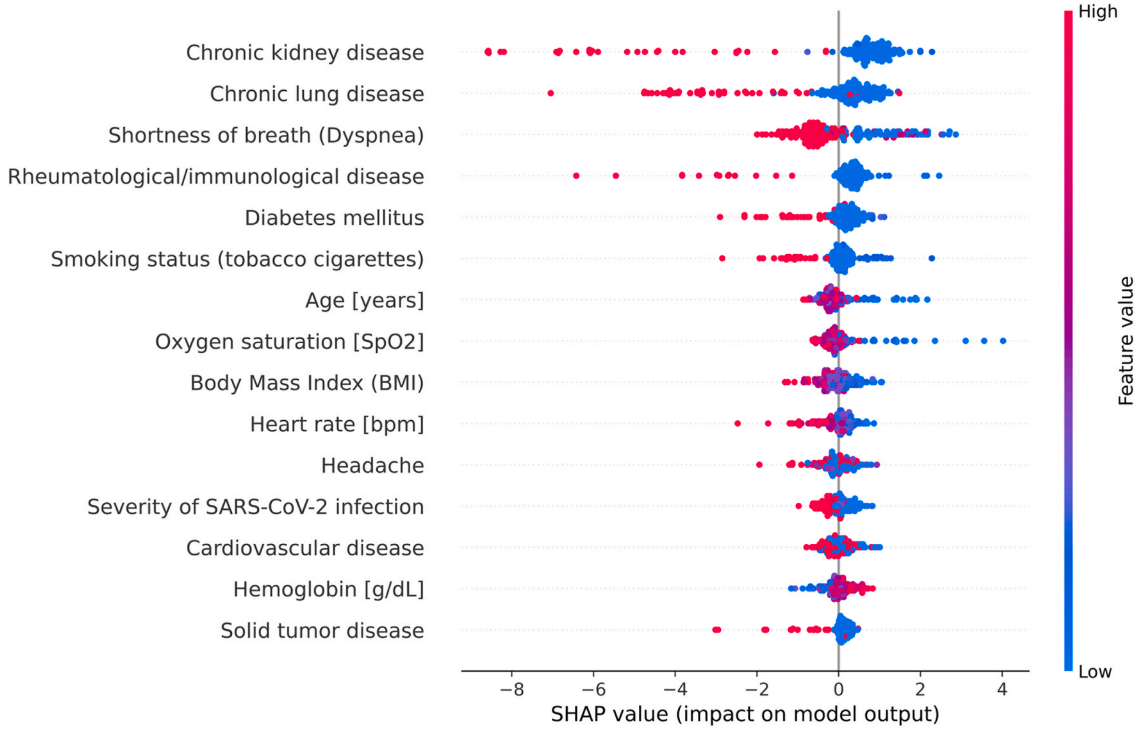

Fig. 1. SHAP values for Ability to Participate in Social Roles and Activities. The SHAP beeswarm plot above shows the impact of different features on the model's output for the PROMIS domain "Social Roles". Each dot represents an observation, and the position of the dot along the x-axis shows how much that feature's value influences the model's prediction (SHAP value). Features with high SHAP values (right side) increase the model's output, while features with low SHAP values (left side) decrease it. The color of each dot represents the feature's value, with pink indicating high values and blue indicating low values. For example, having a chronic kidney disease is associated with a decreased likelihood of having a high PROMIS score for "Social Roles". This is indicated by the negative SHAP values for 'Chronic kidney disease', which shows that the presence of this condition (i.e., pink dots) contributes to lower predicted PROMIS scores (i.e., negative SHAP values). SHAP values were computed using the SUEP sample.

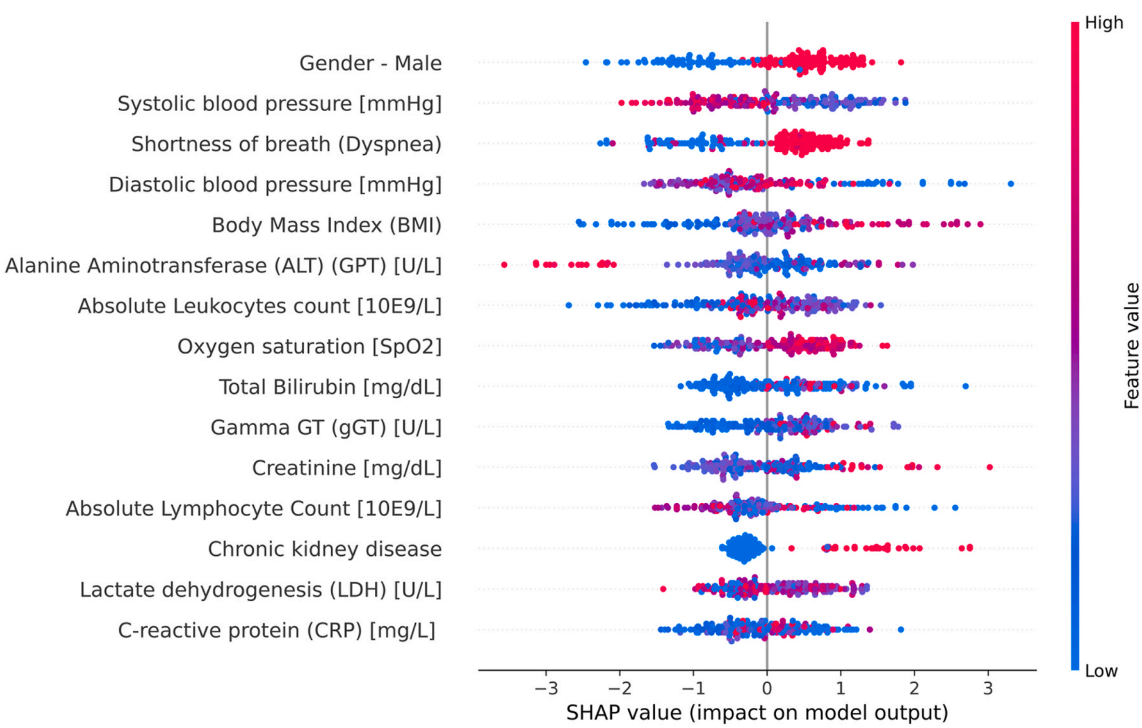

Fig. 2. SHAP beeswarm plot for Depression. SHAP analysis indicates that male sex, lower blood pressure, the presence of dyspnea, and higher BMI were key variables associated with higher predicted PROMIS depression scores. SHAP values were computed using the SUEP sample.

as anxiety, depression, fatigue, pain, sleep disturbance, ability to participate in social roles and activities, and physical function, further

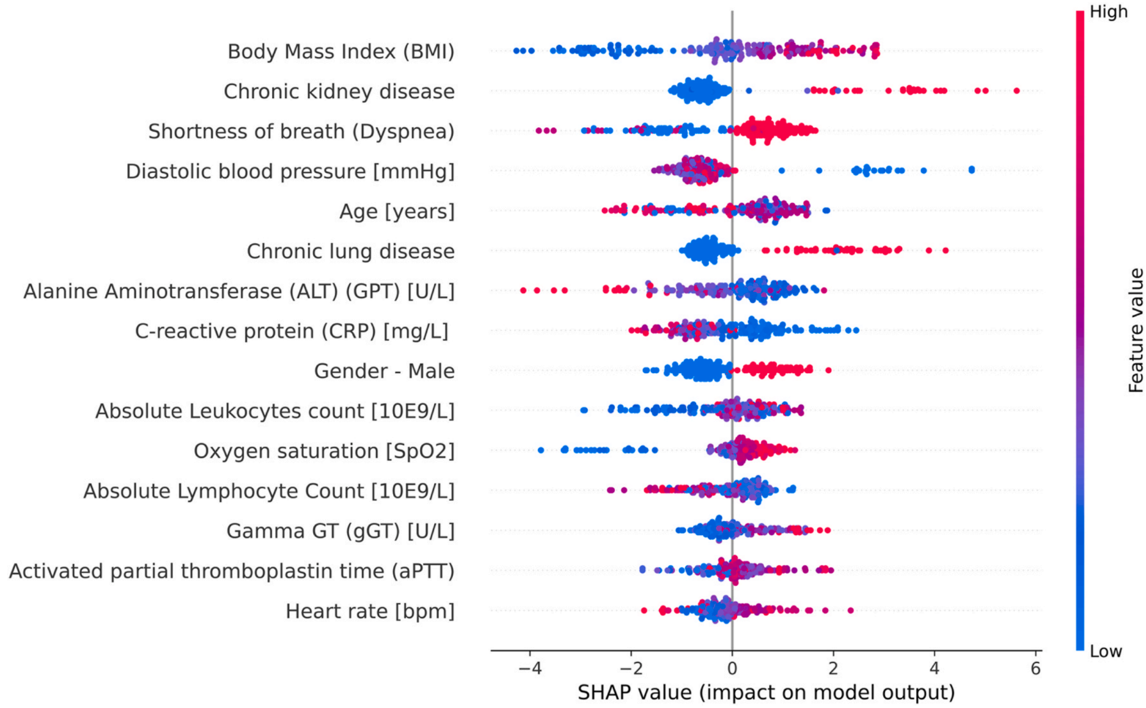

Fig. 3. SHAP beeswarm plot for Fatigue. SHAP analysis indicates that higher BMI, the presence of chronic kidney and lung disease, and dyspnea were key variables associated with higher predicted PROMIS fatigue scores. SHAP values were computed using the SUEP sample.

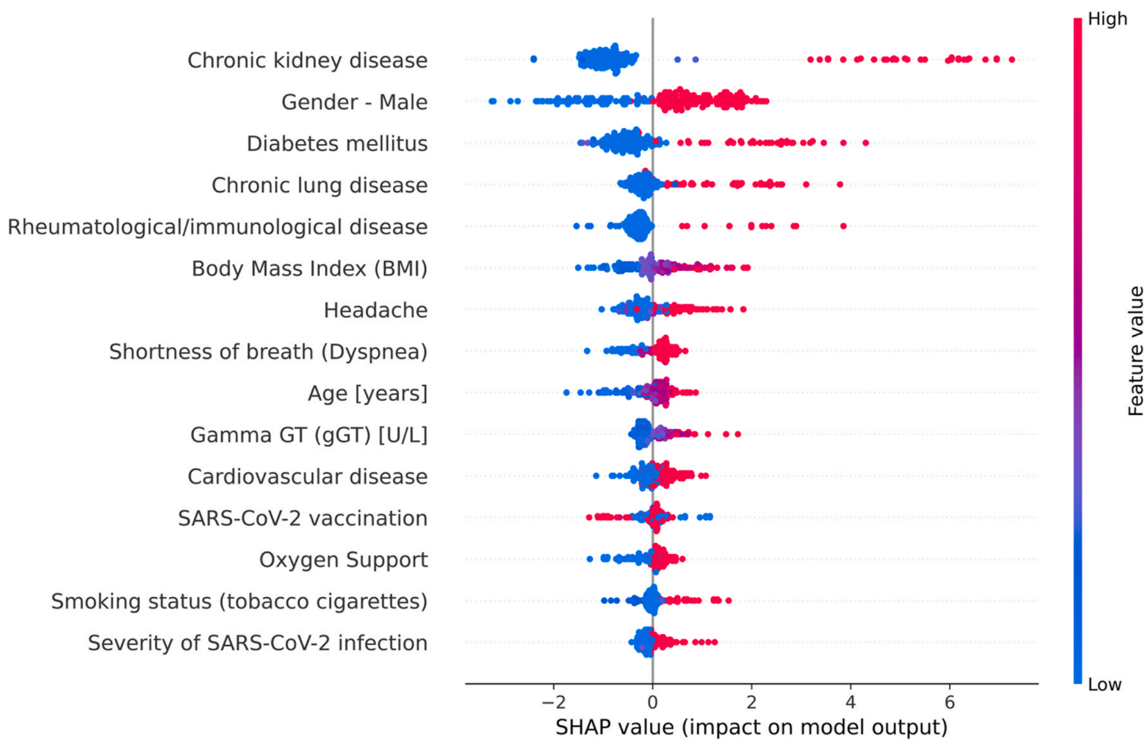

Fig. 4. SHAP beeswarm plot for Pain. SHAP analysis indicates that the presence of chronic kidney disease, diabetes mellitus, chronic lung disease, and rheumatological/immunological disease, as well as male sex and higher BMI, were key variables associated with higher predicted PROMIS pain scores. SHAP values were computed using the SUEP sample.

highlights its utility. These domains closely mirror the core symptom clusters reported in the Long COVID literature, where fatigue, cognitive impairment, psychological distress, pain, and functional limitations are among the most prevalent and debilitating complaints (Davis et al., 2023; Subramanian et al., 2022; Taquet et al., 2021). The high predictive performance across these domains suggests that the model can provide a comprehensive assessment of a patient's risk profile, enabling targeted interventions tailored to individual needs. This domain-specific predictive capability is particularly valuable given emerging evidence that Long COVID symptom trajectories are not

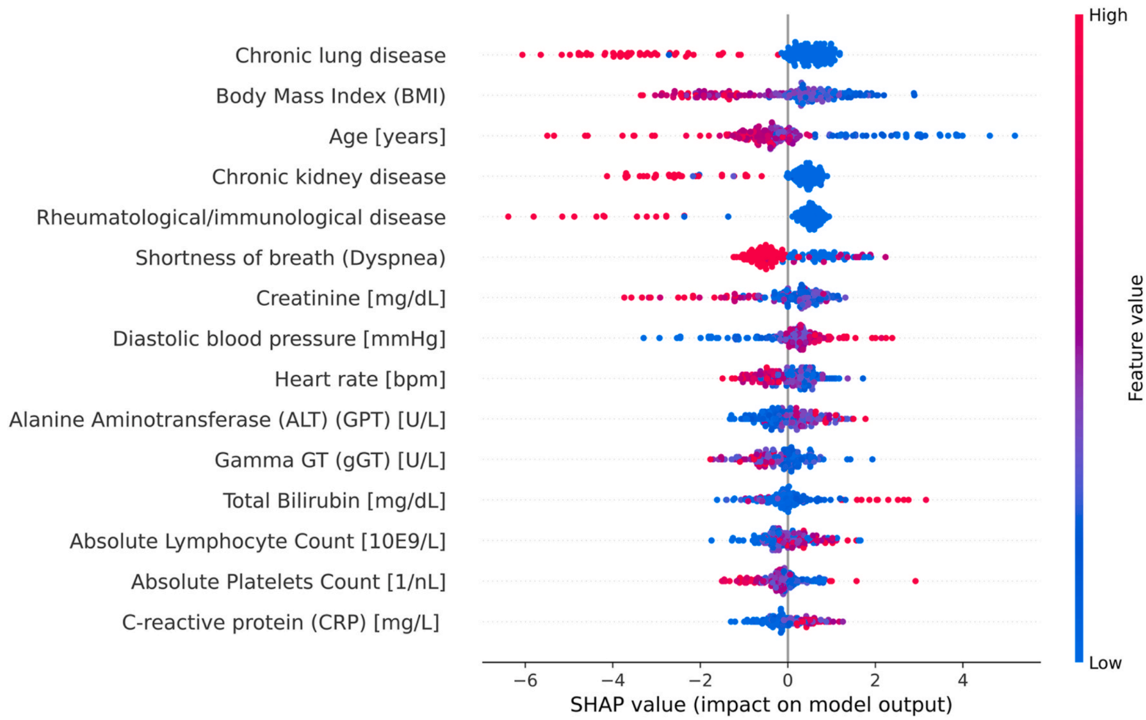

Fig. 5. SHAP beeswarm plot for Physical Function. SHAP analysis indicates that the presence of chronic lung disease, chronic kidney disease, and rheumatological/immunological disease, as well as higher BMI, older age, and dyspnea, were key variables associated with lower predicted PROMIS physical function scores. SHAP values were computed using the SUEP sample.

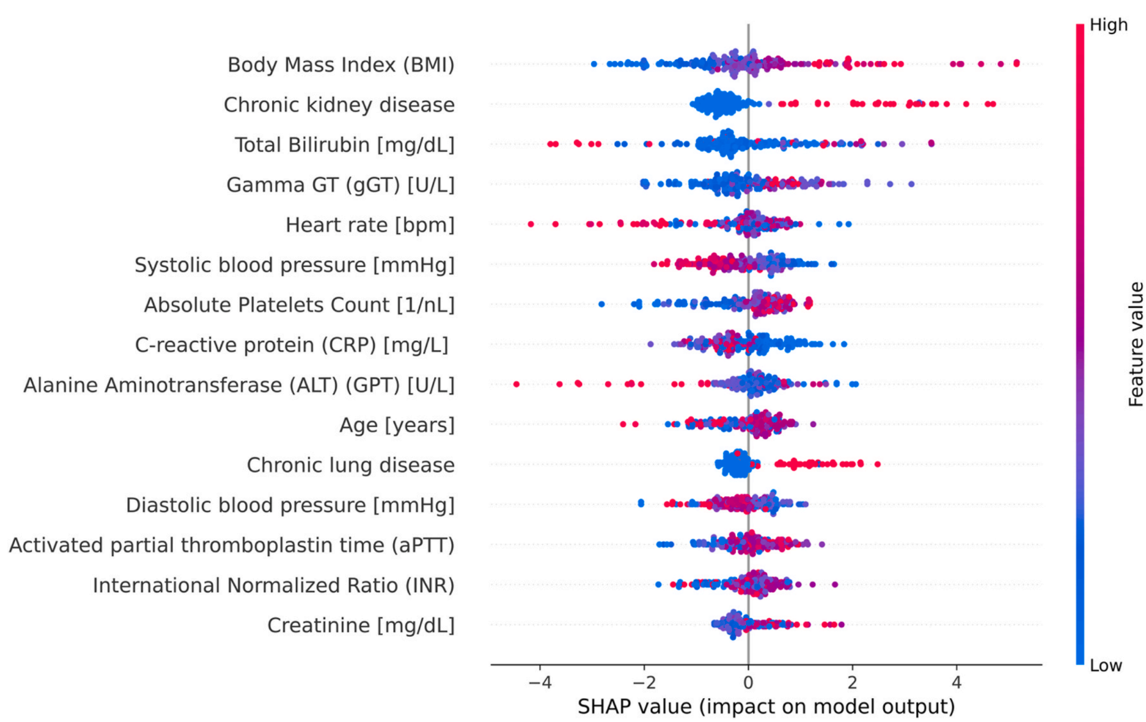

Fig. 6. SHAP beeswarm plot for Sleep. SHAP analysis indicates that higher BMI and the presence of chronic kidney disease were key variables associated with higher predicted PROMIS sleep scores. Additionally, a higher heart rate was associated with lower sleep scores. SHAP values were computed using the SUEP sample.

uniform but instead cluster into distinct phenotypic profiles with different underlying risk factors and prognostic implications (Su et al., 2022; Sudre et al., 2021). By providing granular, domain-level predictions, the model offers the potential to inform not only whether a patient is at risk, but which specific symptom domains are most likely to be affected, thereby enabling more precise and personalized intervention planning. Moreover, the use of SHAP values to interpret the model's predictions provides valuable insights into the key predictive factors. In line with previous findings, we found that demographic characteristics such as age and gender are associated with an increased risk of

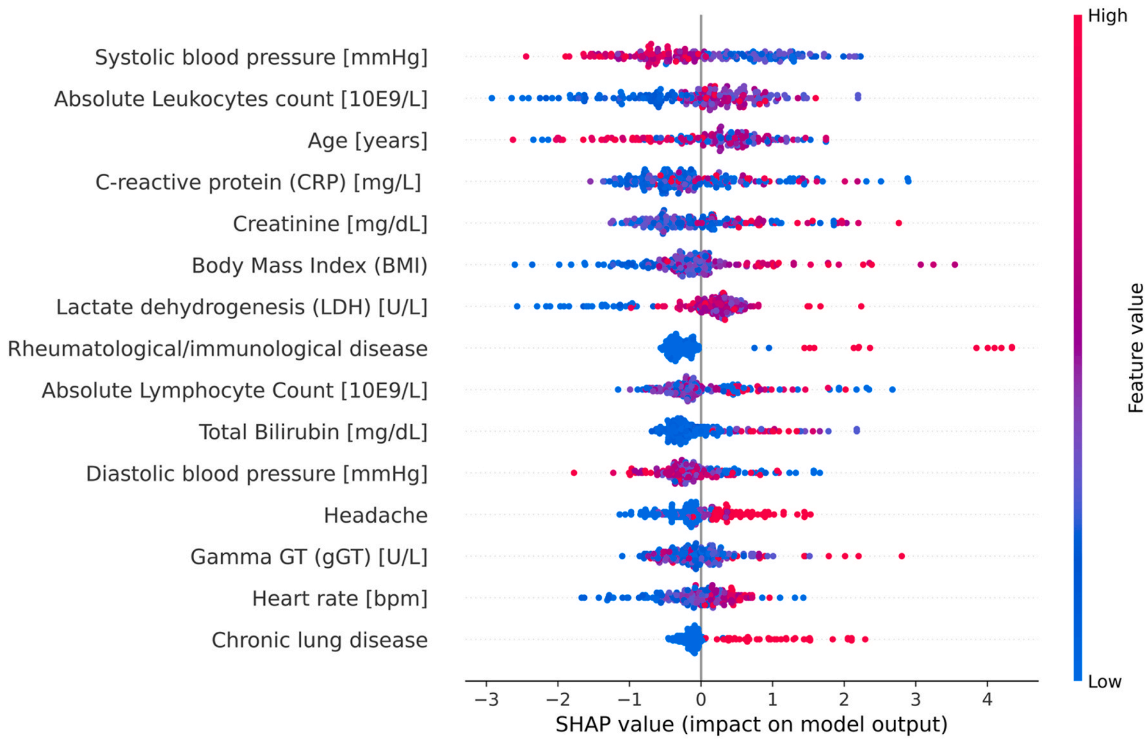

Fig. 7. SHAP beeswarm plot for Anxiety. SHAP analysis indicates that lower systolic blood pressure, higher leukocyte count, and higher creatinine were key variables associated with higher predicted PROMIS anxiety scores. SHAP values were computed using the SUEP sample.

developing post-infection sequelae (Cao, 2020; Mazza et al., 2020). Our SHAP analysis shows that older age is associated with increased risk in terms of its impact on the deterioration of mental and physical health. Sex differences also show clear patterns in risk profiles, with men often showing higher susceptibility to adverse outcomes compared to women. Interestingly, these findings diverge from the broader Long COVID literature, which generally reports a higher prevalence of persistent symptoms among women (Subramanian et al., 2022; Thompson et al., 2022). This discrepancy may reflect differences between community-based Long COVID cohorts, which include predominantly non-hospitalized individuals, and our hospitalized sample, where the severity of acute illness and its associated physiological burden may differentially affect men and shift the risk profile accordingly. Furthermore, consistent with previous studies, we found that indicators of the severity of the initial infection and the presence of chronic diseases increase the risk of developing physical and mental health issues post-recovery (Carfì et al., 2020; Huang et al., 2023). The prominent role of chronic comorbidities, including cardiovascular disease, chronic kidney disease, and diabetes mellitus, aligns with evidence that pre-existing organ vulnerability amplifies the risk of post-acute sequelae, potentially through mechanisms involving endothelial dysfunction, persistent inflammation, and impaired immune resolution (Nalbandian et al., 2021; Xie et al., 2022). In addition, we confirmed that inflammatory markers and immune response indicators, e.g., C-reactive protein or leucocytes, are associated with an increased risk of developing mental and physical sequelae after severe viral infection (Mazza et al., 2020; Peluso et al., 2021; Phetsouphanh et al., 2022). This is consistent with the emerging understanding that immune dysregulation and persistent low-grade inflammation are central pathophysiological drivers of Long COVID, contributing to both somatic symptoms such as fatigue and pain, and neuropsychiatric manifestations including anxiety and depression (Phetsouphanh et al., 2022; Su et al., 2022).

Although several predictors, including BMI, age, chronic comorbidities, and inflammatory markers, were consistently important across PROMIS domains, the SHAP beeswarm plots (Figs. 1–7) reveal notable differences in the strength and directionality of these associations. For instance, higher BMI demonstrated a strong, unidirectional association with increased pain interference, consistent with the well-established relationship between elevated BMI and chronic pain mediated through mechanical loading and systemic inflammation (Okifuji and Hare, 2015; Walsh et al., 2018). In contrast, the effect of BMI on social participation was more diffuse, suggesting moderation by additional factors such as mobility limitations and psychological well-being (Ul-Haq et al., 2013). Inflammatory markers such as leukocyte count and creatinine showed the strongest associations with fatigue and physical function, aligning with evidence that persistent inflammation drives post-acute somatic symptoms (Phetsouphanh et al., 2022; Su et al., 2022), whereas depression and anxiety were more strongly influenced by chronic comorbidities and indicators of acute illness severity such as dyspnea, suggesting that the psychological burden of disease severity may be more relevant to mental health outcomes than acute inflammatory markers alone (Mazza et al., 2020). These domain-specific patterns underscore the value of granular, domain-level prediction and suggest that intervention strategies should be tailored not only to a patient's overall risk level but to the specific symptom domains for which they are most vulnerable.

5. Clinical implications

The integration of this predictive model into EMR systems could significantly improve the management of patients recovering from severe viral infections. By identifying individuals at high risk for subsequent mental and physical sequelae at the point of admission, healthcare providers can implement preventive interventions during hospitalization or shortly thereafter for those who need them. This is particularly pertinent in the context of Long COVID, where the absence of reliable early risk stratification tools has been identified as a major barrier to the timely delivery of preventive and rehabilitative care (Davis et al., 2023; Greenhalgh et al., 2020). Early interventions, such as psychological support, physical rehabilitation, and tailored follow-up care, could mitigate the severity and duration of post-acute symptoms, ultimately enhancing patient quality of life and reducing healthcare costs (Davydow et al., 2008; Herridge et al., 2011; Jackson et al., 2014; Needham et al., 2012; Pandharipande et al., 2013).

Understanding the relative importance of demographic, clinical, and biological variables can inform the development of more effective intervention strategies. For instance, given that certain clinical markers were identified as strong predictors of mental and physical health outcomes, targeted monitoring and early therapeutic interventions could be prioritized for patients exhibiting these markers. Moreover, the model's reliance on routinely collected EMR data ensures that risk stratification can be performed without imposing additional data collection burdens, facilitating seamless integration into existing clinical workflows and enabling scalable deployment across diverse healthcare settings.

6. Strengths, limitations and future directions

A key strength of the present study is the use of an independent external validation cohort (HAP) that differed substantially from the model development sample (SUEP) across several clinical and demographic characteristics, including vaccination rates, disease severity distribution, and smoking status. External validation in a cohort with such heterogeneous characteristics provides a far more rigorous test of model generalizability than validation in a closely matched sample (Collins et al., 2015; Steyerberg and Harrell Jr, 2015). The fact that predictive performance remained consistently high across all PROMIS domains in the HAP cohort, despite these differences, provides strong evidence that the model captures robust predictive relationships that are transportable across diverse clinical contexts. Moreover, external validation itself remains uncommon in the prediction modeling literature, with many published models relying solely on internal validation (Siontis et al., 2015). The inclusion of an independent external cohort therefore represents a methodological strength that enhances confidence in the clinical applicability of the reported findings.

Both the SUEP and HAP cohorts were recruited across multiple sites within the NAPKON network, with HAP encompassing 10 German university hospitals and SUEP spanning a broader network of participating centers nationwide. No site-level corrections were applied, as the model was designed to generalize using patient-level variables. Site-level identifiers were not available in the analytic dataset. However, the robust predictive performance observed across two independently recruited, multi-site cohorts with substantially different demographic and clinical profiles provides strong evidence that the model generalizes across diverse clinical settings without requiring explicit site-level adjustment.

Despite these strengths, several limitations should be acknowledged. First, the study focused on a specific cohort of SARS-CoV-2 patients with patients treated in Germany, which may limit the generalizability of the findings to other viral infections or different patient populations. Although the predictive factors identified in our model, such as chronic comorbidities, inflammatory markers, and disease severity, are broadly relevant across post-viral syndromes, and our findings are in line with previous studies examining risk in patients after severe viral infection, future research should aim to validate the model in diverse cohorts and across various viral infections to enhance its applicability. Second, our outcome assessment was conducted at three months post-hospitalization. However, the three-month time point was deliberately chosen for its clinical relevance, as this period represents a critical intervention window during which emerging symptoms are still modifiable through lower-threshold interventions before trajectories consolidate into chronic conditions (Davydow et al., 2008; Herridge et al., 2011; Jackson et al., 2014; Needham et al., 2012; Pandharipande et al., 2013). Untreated early symptoms are important predictors of persistent impairment at later time points, and early psychological distress independently worsens physical recovery, including fatigue, functional disability, and pain. Accurate prediction at this juncture therefore offers a clinically meaningful opportunity to intervene before the reinforcing cycle between mental and physical sequelae becomes entrenched. Nonetheless, future research should validate these models against longer-term outcome data to assess the durability of predictive accuracy. Third, the study relied on self-reported outcomes measured by the PROMIS-57 Profile. While this is a validated and widely used tool, incorporating objective measures of physical and mental health could provide a more comprehensive assessment of patient outcomes.

An important consideration is that pre-infection mental health and physical symptom assessments were not incorporated into the predictive model. This was a deliberate design decision rather than an oversight, driven by the pragmatic goal of developing a risk stratification tool that relies exclusively on data routinely available within EMR systems during an acute hospitalization encounter. While pre-existing symptomatology would undoubtedly provide additional predictive value, the systematic collection of mental health history in acute care settings is hampered by well-documented barriers, including patient reluctance to disclose due to stigma (Clement et al., 2015; Henderson et al., 2013), provider discomfort and time constraints in eliciting psychiatric history (Mitchell et al., 2009), and inconsistent documentation practices across institutions. Importantly, our model partially accounts for pre-existing vulnerability through the inclusion of chronic comorbid conditions such as cardiovascular disease, chronic lung disease, and diabetes mellitus, which emerged as important predictors in the SHAP analyses and are well-established correlates of pre-existing mental health conditions (J. Katon, 2011; Scott et al., 2016). Nevertheless, our findings should be interpreted with the caveat that post-hospitalization symptoms identified at three months may reflect a combination of infection-related sequelae and pre-existing conditions. Moreover, because both cohorts were enrolled prospectively after SARS-CoV-2 infection, no pre-infection patient-reported outcome data were available to assess within-person symptom change, precluding a direct sensitivity analysis comparing pre- and post-infection levels. However, the substantial inter-individual variability in post-hospitalization outcomes (Tables 2 and 3), the prominence of acute clinical predictors in the SHAP analyses and the high proportion of participants scoring within normal ranges all suggest that the model captures meaningful post-infection health trajectories rather than stable pre-existing characteristics alone. Future research incorporating validated pre-infection mental health measures, where feasible, would help disentangle these contributions and further refine predictive accuracy. Pragmatically, however, the current approach prioritizes clinical scalability and real-world implementability, ensuring that the model can be deployed within existing EMR infrastructures without imposing additional data collection burdens on patients or providers.

7. Conclusion

This study presents an externally validated predictive model for identifying individuals at risk of developing mental and physical sequelae after hospitalization due to a severe viral infection. The findings contribute to the growing body of Long COVID research by demonstrating that routinely collected clinical data available at the time of hospital admission can accurately predict which patients will develop significant post-acute mental and physical symptomatology. By leveraging routinely collectable EMR data, the model offers a practical and effective tool for acute risk stratification and early preventive intervention, with the potential to significantly improve patient outcomes and optimize healthcare resource allocation. Future research should focus on validating and refining the model in diverse populations, across different viral infections, and on exploring the integration of objective health measures to further enhance its clinical utility.

CRediT authorship contribution statement

Katharina Schultebraucks: Conceptualization, Formal analysis, Methodology, Writing – original draft. **Sapir Gershov:** Formal analysis, Writing – review & editing. **Felix Fischer:** Data curation, Writing – review & editing. **Katja Wingenfeld:** Writing – review & editing. **Sein Schmidt:** Conceptualization, Writing – review & editing. **Sarah Steinbrecher:** Writing – review & editing. **Thomas Zoller:** Writing – review & editing. **Fridolin Steinbeis:** Writing – review & editing. **Sina M. Pütz:** Writing – review & editing. **Jürgen Deckert:** Writing – review & editing. **Margarete Scherer:** Writing – review & editing. **Isabel Bröhl:** Writing – review & editing. **Patricia Wagner:** Writing – review & editing. **Katharina S. Appel:** Writing – review & editing. **Mirjam Kohls:** Writing – review & editing. **Steffi Jiru-Hillmann:** Writing – review & editing. **Matthias Nauck:** Writing – review & editing. **Bettina Lorenz-Depiereux:** Writing – review & editing. **Sabine Blaschke:** Writing – review & editing. **Anna Muzalyova:** Writing – review & editing. **Christoph Stellbrink:** Writing – review & editing. **Anke Steinmetz:** Writing – review & editing. **Marylyn Martina Addo:** Writing – review & editing. **Edgar Dahl:** Writing – review & editing. **Markus Zettler:** Writing – review & editing. **Stefan Hansch:** Writing – review & editing. **Andreas Dinkel:** Writing – review & editing. **Verena Keitel:** Writing – review & editing. **Maria J.G.T. Vehreschild:** Writing – review & editing. **Jörg J. Vehreschild:** Writing – review & editing. **Friedemann Paul:** Writing – review & editing. **Martin Witzenrath:** Writing – review & editing. **Matthias Rose:** Writing – review & editing. **Christian Otte:** Conceptualization, Funding acquisition, Resources, Writing – review & editing.

Declaration of competing interest

KS: received support from the National Institute of Mental Health (R01MH129856) and the National Heart, Lung, and Blood Institute (R01HL156134).

AD: Honoraria for lectures and educational workshops from Gilead, Novartis, Coliquio.

CO: Honoraria for lectures and/or scientific advice from Boehringer-Ingelheim, Janssen, Limes Klinikgruppe, Neuraxpharm, Oberberg Kliniken and Peak Profiling; Research Funding from the German Research Foundation (OT 209/7-3; 14-1, 19-1, 21-1, EXC, 2049), the European Commission (IMI2 859366), the German Federal Ministry of Education and Research (KS2017-067), the Berlin Institute of Health (B3010350), and the Wellcome Trust.

SB: Funding from the German Federal Ministry of Education and Research (Grant No. 16SV8977) an Federal Ministry of Health (Grant No. 2520DAT803).

SMP: SMH received travel grants from Tillotts.

MJGTV: Research grants: 3M, Astellas Pharma, Biontech, DaVolterra, Gilead Sciences, Glycom, Immunic, Merck/MSD, Takeda Pharmaceutical, Tillots pharma; speaker fees and consulting: Astellas Pharma, Basilea, Bio-Mérieux, DaVolterra, Falk, Farmak International Holding GmbH, Ferring, Gilead Sciences, Immunic AG, Merck/MSD, Pfizer, Roche, Organobalance, SocraTec R&D GmbH, Tillots Pharma.

SH: Travel grants from Gilead Sciences.

MW: Research grants: Biotest, Pantherna, Aptarion; consulting fees: Biotest, Pantherna, Aptarion; honoraria for lectures: Astra Zeneca, Chiesi, Insmed, Gilead, Pfizer, Boehringer Ingelheim.

All other authors declare no conflict of interest.

Acknowledgements

NAPKON is part of the NUM, funded by the German Federal Ministry of Education and Research (BMBF) (FKZ: 01KX2121).Katharina Schultebraucks received support from the National Institute of Mental Health (R01MH129856) and the National Heart, Lung, and Blood Institute (R01HL156134).We gratefully thank all NAPKON sites that contributed patient data and/or biosamples for this analysis. The representatives of NAPKON sites contributing at least 5 per mille to this analysis are (alphabetical order): Bielefeld University, Medical School and University Medical Center OWL, Bielefeld (Alsaad K, Hamelmann E, Heidenreich H, Hornberg C, Kulamadayil-Heidenreich NSA, Maasjosthusmann P, Muna A, Ruwe M, Stellbrink C, Tebbe J), Charité - Universitaetsmedizin Berlin, Berlin (Finke C, Mittermaier M, Mueller-Plathe M, Pley C, Stege A, Steinbeis F, Steinbrecher S, Wildberg C, Zessin E), Hannover Medical School, Hannover (Bernemann I, Drick N, Hoeper M, Illig T, Kersting M, Klopp N, Kopfnagel V, Pink I, Zetzsche F), Thoraxklinik at University Hospital Heidelberg, Heidelberg (Altner D, Fallenbuechel M, Hoppstock C, Kress A, Schaufler B, Suengue J), University Hospital Augsburg, Augsburg (Bader S, Engelmann M, Fuchs A, Langer A, Maerkl B, Messmann H, Muzalyova A, Roemmele C), University Hospital Carl Gustav Carus, Dresden (Altmann H, Berner R, Dressen S, Koch T, Lindemann D, Seele K, Spieth P, Tausche K, Toepfner N, v. Bonin S), University Hospital Duesseldorf, Duesseldorf (Feldt T, Feuerhake ML, Jensen BEO, Killer A, Luedde T, Lutterbeck M, Paluschinski M, Timm J, Ventura-Pereira J, de Angelis C), University Hospital Erlangen, Erlangen (Kraska D, Kremer AE, Leppkes M, Mang J, Neurath MF, Prokosch HU, Schmid J, Vetter M, Willam C, Wolf K), University Hospital Frankfurt, Frankfurt am Main (Arendt C, Bellinghausen C, Cremer S, Groh A, Gruenewaldt A, Khodamoradi Y, Klinsing S, Rohde G, Vehreschild M, Vogl T), University Hospital Giessen & Marburg, Giessen (Halfar HH, Herold S, Kryvenko V, Larsen CM, Ruppert C, Scheunemann M, Schnell E, Seeger W, Uribe Munoz A, Vadasz I), University Hospital Hamburg-Eppendorf, Hamburg (Addo M, Almahfoud M, Engels ALF, Jarczak D, Kerinn M, Kluge S, Kobbe R, Petereit S, Schlesner C, Zeller T), University Hospital Jena, Jena (Eibner C, Kiehntopf M, Kiesewetter H, Knutti N, Koehler J, Niederlechner H, Quickert S, Reuken P, Ritzmann H, Stallmach A), University Hospital LMU Munich, Munich (Frank S, Hellmuth JC, Keppler OT, Mandel C, Mueller S, Muenchhoff M, Reeh L, Stubbe H, Zwissler B, von Bergwelt M), University Hospital Leipzig, Leipzig (Baber R, Bercker S, Krug N, Mueller SD, Wirtz H), University Hospital RWTH Aachen, Aachen (Dahl E, Dreher M, Marx N, Mueller-Wieland D, Wipperfuerth J), University Hospital Regensburg, Regensburg (Albig H, Burkhardt R, Feustel M, Fisser C, Hansch S, Hanses F, Hitzenbichler F, Niedermair T, Schuster P, Tran T), University Hospital Schleswig-Holstein Kiel, Kiel (Figuera Basso ME, Franzenburg J, Heyckendorf J, Pape D, Scheer AM), University Hospital Schleswig-Holstein Luebeck, Luebeck (Bohnet S, Droemann D, Franzen KF, Hoerster R, Parschke P), University Hospital Technical University Munich, Munich (Barkey W, Erber J, Fricke L, Gladis S, Michler T, Mueller L, Schneider J, Spinner C, Winter C), University Hospital Tuebingen, Tuebingen (Bitzer M, Bunk S, Göpel S, Haeberle H, Kienzle K, Mahrhofer H, Malek N, Rosenberger P, Struemper C, Trauner F), University Hospital Wuerzburg, Wuerzburg (Einsele H, Frantz S, Frey A, Haas K, Haertel C, Herrmann J, Meybohm P, Schmidt J, Schulze P, Volkmann J), University Hospital of Cologne, Cologne (Adam M, Cornely OA, Faust M, Grothe J, Koehler P, Nacov J, Persegihl T, Rybniker J, Stemler J, Steven P), University Hospitals of the Ruhr University Bochum, Bochum (Brinkmann F, Brueggemann Y, Gambichler T, Hellwig K, Luecke T, Reinacher-Schick A, Schmidt WE, Schuette C, Steinmann E, Torres Reyes C), University Medical Center Freiburg, Freiburg (August D, Barrera M, Goetz V, Imhof A, Koch S, Nieters A, Peyerl-Hoffmann G, Rieg SR), University Medical Center Goettingen, Central Biobank, Goettingen (Nussbeck SY), University Medical Center Goettingen, Central Laboratory, Goettingen (Hafke A), University Medical Center Goettingen, Emergency Department, Goettingen (Blaschke S, Hermanns G, Santibanez-Santana M, Zeh S), University Medicine Essen, Essen (Brochhagen L, Dolff S, Elsner C, Krawczyk A, Madel RJ, Otte M, Thuemmler L, Witzke O), University Medicine Greifswald, Greifswald (Becker K, Doerr M, Lehnert K, Nauck M, Piasta N, Schaefer C, Schaefer E, Schattschneider M, Scheer C, Stahl D), University Medicine Oldenburg, Oldenburg (Arlt A, Griesinger F, Guenther U, Hamprecht A, Juergens K, Kluge A, Meinhardt C, Meinhardt K, Petersmann A, Prenzel R).We gratefully thank all participating NAPKON and NUKLEUS infrastructures that contributed to this analysis. The representatives of these NAPKON and NUKLEUS infrastructures are (alphabetical order): Charite - Universitaetsmedizin Berlin, Berlin (Balzuweit B, Berger S, Krannich A, Kurth F, Lienau J, Lorbeer R, Pley C, Schaller J, Schmidt S, Thibeault C, Witzenrath M, Zoller T), Helmholtz Center Munich, Munich (Kraus M, Lorenz-Depiereux B), University Hospital Cologne, Cologne (Brechtel M, Broehl I, Fiedler K, Hopff SM, Knaub K, Lee C, Nunes de Miranda S, Sauer G, Vehreschild JJ, Wagner P), University Hospital Frankfurt, Frankfurt (Appel K, Geisler R, Raquib S, Scherer M, Vehreschild JJ, Weirauch T, Wolf L), University Hospital Wuerzburg, Wuerzburg (Guenther K, Haug F, Haug J), University Hospital Wuerzburg and University of Wuerzburg, Wuerzburg (Fiessler C, Heuschmann PU, Miljukov O, Nuernberger C, Reese JP, Schmidbauer L), University Medicine Goettingen, Goettingen (Chaplinskaya I, Hanss S, Krefting D, Pape C, Rainers M, Schoneberg A, Weinert N), University Medicine Greifswald, Greifswald (Bahls T, Hoffmann W, Nauck M, Schaefer C, Schattschneider M, Stahl D, Valentin H), University of Wuerzburg, Wuerzburg (Jiru-Hillmann S).We gratefully thank the NAPKON Steering Committee: University Hospital Giessen and Marburg, Giessen (Herold S), University of Würzburg, Würzburg (Heuschmann P), Charité - Universitätsmedizin Berlin, Berlin (Heyder R), University Medicine Greifswald, Greifswald (Hoffmann W), Hannover Unified Biobank, Hannover Medical School, Hannover (Illig T), University Hospital Schleswig-Holstein, Kiel (Schreiber S), University Hospital Cologne and University Hospital Frankfurt, Cologne and Frankfurt (Vehreschild JJ), Charité - Universitätsmedizin Berlin, Berlin (Witzenrath M).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbih.2026.101267.

Data availability

The authors do not have permission to share data.

References

Ballering, A.V., van Zon, S.K., olde Hartman, T.C., Rosmalen, J.G., 2022. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. Lancet 400 (10350), 452–461.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Cao, X., 2020. COVID-19: immunopathology and its implications for therapy. Nat. Rev. Immunol. 20 (5), 269–270.
Carfi, A., Bernabei, R., Landi, F., 2020. Persistent symptoms in patients after acute COVID-19. JAMA 324 (6), 603–605.
Cella, D., Choi, S.W., Condon, D.M., Schalet, B., Hays, R.D., Rothrock, N.E., Yount, S., Cook, K.F., Gershon, R.C., Amtmann, D., 2019. PROMIS® adult health profiles: efficient short-form measures of seven health domains. Value Health 22 (5), 537–544.
Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., 2010. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J. Clin. Epidemiol. 63 (11), 1179–1194.
Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rüsch, N., Brown, J.S., Thornicroft, G., 2015. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol. Med. 45 (1), 11–27.
Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. J. British Surg. 102 (3), 148–158.
Davis, H.E., McCorkell, L., Vogel, J.M., Topol, E.J., 2023. Long COVID: major findings, mechanisms and recommendations. Nat. Rev. Microbiol. 21 (3), 133–146.
Davydow, D.S., Gifford, J.M., Desai, S.V., Needham, D.M., Bienvenu, O.J., 2008. Posttraumatic stress disorder in general intensive care unit survivors: a systematic review. Gen. Hosp. Psychiatry 30 (5), 421–434.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Geense, W.W., Zegers, M., Peters, M.A., Ewalds, E., Simons, K.S., Vermeulen, H., van der Hoeven, J.G., van den Boogaard, M., 2021. New physical, mental, and cognitive problems 1 year after ICU admission: a prospective multicenter study. Am. J. Respir. Crit. Care Med. 203 (12), 1512–1521.
Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63 (1), 3–42.
Greenhalgh, T., Knight, M., Buxton, M., Husain, L., 2020. Management of post-acute covid-19 in primary care. BMJ 370.
Hays, R.D., Spritzer, K.L., Schalet, B.D., Cella, D., 2018. PROMIS®-29 v2. 0 profile physical and mental health summary scores. Qual. Life Res. 27 (7), 1885–1891.
Henderson, C., Evans-Lacko, S., Thornicroft, G., 2013. Mental illness stigma, help seeking, and public health programs. Am. J. Publ. Health 103 (5), 777–780.
Herridge, M.S., Tansey, C.M., Matté, A., Tomlinson, G., Diaz-Granados, N., Cooper, A., Guest, C.B., Mazer, C.D., Mehta, S., Stewart, T.E., 2011. Functional disability 5 years after acute respiratory distress syndrome. N. Engl. J. Med. 364 (14), 1293–1304.
Hippisley-Cox, J., Coupland, C.A., Mehta, N., Keogh, R.H., Diaz-Ordaz, K., Khunti, K., Lyons, R.A., Kee, F., Sheikh, A., Rahman, S., 2021. Risk prediction of covid-19 related death and hospital admission in adults after covid-19 vaccination: national prospective cohort study. BMJ 374.
Hopff, S.M., Appel, K.S., Miljukov, O., Schneider, J., Addo, M.M., Bals, R., Bercker, S., Blaschke, S., Bröhl, I., Büchner, N., 2024. Comparison of post-COVID-19 symptoms in patients infected with the SARS-CoV-2 variants delta and omicron—results of the cross-sectoral platform of the German national pandemic cohort network (NAPKON-SUEP). Infection 1–15.
Huang, C., Huang, L., Wang, Y., Li, X., Ren, L., Gu, X., Kang, L., Guo, L., Liu, M., Zhou, X., 2023. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet 401 (10393), e21–e33.
Huang, C.-W., Park, J.S., Song, H., Khang, V.K., Yu, A.S., Nguyen, H.Q., Lee, J.S., Shen, E., 2022. Disease-specific factors associated with readmissions or mortality after hospital discharge in COVID-19 patients: a retrospective cohort study. J. Gen. Intern. Med. 37 (15), 3973–3978.
Katon, W.J., 2011. Epidemiology and treatment of depression in patients with chronic medical illness. Dialogues Clin. Neurosci. 13 (1), 7–23.
Jackson, J.C., Pandharipande, P.P., Girard, T.D., Brummel, N.E., Thompson, J.L., Hughes, C.G., Pun, B.T., Vasilevskis, E.E., Morandi, A., Shintani, A.K., 2014. Depression, post-traumatic stress disorder, and functional disability in survivors of critical illness in the BRAIN-ICU study: a longitudinal cohort study. Lancet Respir. Med. 2 (5), 369–379.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Madley-Dowd, P., Hughes, R., Tilling, K., Heron, J., 2019. The proportion of missing data should not be used to guide decisions on multiple imputation. J. Clin. Epidemiol. 110, 63–73.
Mazza, M.G., De Lorenzo, R., Conte, C., Poletti, S., Vai, B., Bollettini, I., Melloni, E.M.T., Furlan, R., Ciceri, F., Rovere-Querini, P., 2020. Anxiety and depression in COVID-19 survivors: role of inflammatory and clinical predictors. Brain Behav. Immun. 89, 594–600.
Mitchell, A.J., Vaze, A., Rao, S., 2009. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374 (9690), 609–619.
Moons, K.G., Altman, D.G., Reitsma, J.B., Ioannidis, J.P., Macaskill, P., Steyerberg, E.W., Vickers, A.J., Ransohoff, D.F., Collins, G.S., 2015. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann. Intern. Med. 162 (1), W1–W73.
Nalbandian, A., Sehgal, K., Gupta, A., Madhavan, M.V., McGroder, C., Stevens, J.S., Cook, J.R., Nordvig, A.S., Shalev, D., Sehrawat, T.S., 2021. Post-acute COVID-19 syndrome. Nat. Med. 27 (4), 601–615.
Needham, D.M., Davidson, J., Cohen, H., Hopkins, R.O., Weinert, C., Wunsch, H., Zawistowski, C., Bemis-Dougherty, A., Berney, S.C., Bienvenu, O.J., 2012. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. Crit. Care Med. 40 (2), 502–509.
Okifuji, A., Hare, B.D., 2015. The association between chronic pain and obesity. J. Pain Res. 399–408.
Pandharipande, P.P., Girard, T.D., Jackson, J.C., Morandi, A., Thompson, J.L., Pun, B.T., Brummel, N.E., Hughes, C.G., Vasilevskis, E.E., Shintani, A.K., 2013. Long-term cognitive impairment after critical illness. N. Engl. J. Med. 369 (14), 1306–1316.
Peluso, M.J., Deitchman, A.N., Torres, L., Iyer, N.S., Munter, S.E., Nixon, C.C., Donatelli, J., Thanh, C., Takahashi, S., Hakim, J., 2021. Long-term SARS-CoV-2-specific immune and inflammatory responses in individuals recovering from COVID-19 with and without post-acute symptoms. Cell Rep. 36 (6).
Pfaff, E.R., Girvin, A.T., Bennett, T.D., Bhatia, A., Brooks, I.M., Deer, R.R., Dekermanjian, J.P., Jolley, S.E., Kahn, M.G., Kostka, K., 2022. Identifying who has long COVID in the USA: a machine learning approach using N3C data. Lancet Digit. Health 4 (7), e532–e541.
Phetsouphanh, C., Darley, D.R., Wilson, D.B., Howe, A., Munier, C., Patel, S.K., Juno, J. A., Burrell, L.M., Kent, S.J., Dore, G.J., 2022. Immunological dysfunction persists for 8 months following initial mild-to-moderate SARS-CoV-2 infection. Nat. Immunol. 23 (2), 210–216.
Rothrock, N.E., Hays, R.D., Spritzer, K., Yount, S.E., Riley, W., Cella, D., 2010. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). J. Clin. Epidemiol. 63 (11), 1195–1204.
Schons, M., Pilgram, L., Reese, J.-P., Stecher, M., Anton, G., Appel, K.S., Bahmer, T., Bartschke, A., Bellinghausen, C., Bernemann, I., 2022. The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. Eur. J. Epidemiol. 37 (8), 849–870.
Scott, K.M., Lim, C., Al-Hamzawi, A., Alonso, J., Bruffaerts, R., Caldas-de-Almeida, J.M., Florescu, S., De Girolamo, G., Hu, C., De Jonge, P., 2016. Association of mental disorders with subsequent chronic physical conditions: world mental health surveys from 17 countries. JAMA Psychiatry 73 (2), 150–158.

Siontis, G.C., Tzoulaki, I., Castaldi, P.J., Ioannidis, J.P., 2015. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J. Clin. Epidemiol. 68 (1), 25–34.

Steinbeis, F., Thibeault, C., Steinbrecher, S., Ahlgrimm, Y., Haack, I.A., August, D., Balzuweit, B., Bellinghausen, C., Berger, S., Chaplinskaya-Sobol, I., 2024. Analysis of acute COVID-19 including chronic morbidity: protocol for the deep phenotyping National Pandemic Cohort Network in Germany (NAPKON-HAP). Infection 52 (1), 93–104.

Steyerberg, E.W., Harrell Jr, F.E., 2015. Prediction models need appropriate internal, internal-external, and external validation. J. Clin. Epidemiol. 69, 245.

Su, Y., Yuan, D., Chen, D.G., Ng, R.H., Wang, K., Choi, J., Li, S., Hong, S., Zhang, R., Xie, J., 2022. Multiple early factors anticipate post-acute COVID-19 sequelae. Cell 185 (5), 881–895. e820.

Subramanian, A., Nirantharakumar, K., Hughes, S., Myles, P., Williams, T., Gokhale, K. M., Taverner, T., Chandan, J.S., Brown, K., Simms-Williams, N., 2022. Symptoms and risk factors for long COVID in non-hospitalized adults. Nat. Med. 28 (8), 1706–1714.

Sudre, C.H., Murray, B., Varsavsky, T., Graham, M.S., Penfold, R.S., Bowyer, R.C., Pujol, J.C., Klaser, K., Antonelli, M., Canas, L.S., 2021. Attributes and predictors of long COVID. Nat. Med. 27 (4), 626–631.

Taquet, M., Geddes, J.R., Husain, M., Luciano, S., Harrison, P.J., 2021. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. Lancet Psychiatry 8 (5), 416–427.

Thompson, E.J., Williams, D.M., Walker, A.J., Mitchell, R.E., Niedzwiedz, C.L., Yang, T. C., Huggins, C.F., Kwong, A.S., Silverwood, R.J., Di Gessa, G., 2022. Long COVID burden and risk factors in 10 UK longitudinal studies and electronic health records. Nat. Commun. 13 (1), 3528.

Ul-Haq, Z., Mackay, D.F., Fenwick, E., Pell, J.P., 2013. Meta-analysis of the association between body mass index and health-related quality of life among adults, assessed by the SF-36. Obesity 21 (3), E322–E327.

Walsh, T.P., Arnold, J.B., Evans, A.M., Yaxley, A., Damarell, R.A., Shanahan, E.M., 2018. The association between body fat and musculoskeletal pain: a systematic review and meta-analysis. BMC Muscoskelet. Disord. 19 (1), 233.

Weich, S., 2022. In: Mental Health After covid-19, 376. British Medical Journal Publishing Group.

Whitaker, M., Elliott, J., Chadeau-Hyam, M., Riley, S., Darzi, A., Cooke, G., Ward, H., Elliott, P., 2022. Persistent COVID-19 symptoms in a community study of 606,434 people in England. Nat. Commun. 13 (1), 1957.

Xie, Y., Xu, E., Bowe, B., Al-Aly, Z., 2022. Long-term cardiovascular outcomes of COVID-19. Nat. Med. 28 (3), 583–590.

Zhao, Y., Shi, L., Jiang, Z., Zeng, N., Mei, H., Lu, Y., Yang, J., Jin, F., Ni, S., Wu, S., 2023. The phenotype and prediction of long-term physical, mental and cognitive COVID-19 sequelae 20 months after recovery, a community-based cohort study in China. Mol. Psychiatr. 28 (4), 1793–1801.