RESEARCH PAPER

Risk Assessment of Post-COVID-19 Respiratory Infections Using Stacking-Based Machine Learning Techniques

L. William Mary[1], S. Albert Antony Raj[2]

[1,2]Department of Computer Applications, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.

[1]Email: *wl6649@srmist.edu.in*

[2]Email: *alberts@srmist.edu.in*

ABSTRACT

Clinically significant sequelae of COVID-19 such as respiratory infections and enduring pulmonary vulnerability have become clinically significant challenges to surveillance, triage, and resource planning on the healthcare system level. This paper suggests a machine learning model based on stacking to evaluate personalized risks of developing post-COVID-19 respiratory infections and other unfavorable events through combinations of regularly measured clinical, laboratory, radiological and longitudinal follow-up variables. The methodology is based on combining heterogeneous base learners who are randomly chosen to represent complementary linear, non-linear and interaction effects into a meta-learner which optimally aggregates out-of-fold predictions to enhance generalization and minimize model variance. An overview of a reproducible pipeline is presented, that is, cohort definition during the post-acute period, feature harmonization across care settings, missingness treatment, temporal leaks management, and calibration to attain clinically interpretable risk probabilities. Discrimination, calibration, and decision-analytic measures are used to conduct model performance, and subgroup analysis is used to test performance across the age, sex, vaccination status, comorbidity burden, and severity of acute disease domain. To facilitate clinical adoption, explainability is also added with global and patient level attribution, which makes it possible to identify risk drivers to act upon, including inflammatory signatures, pulmonary impairment markers, oxygenation indices, and previous secondary infection hints. The presented stacking method is expected to provide better predictive results compared to single models without sacrificing transparency and operational viability to be implemented in post-COVID clinics and hospital follow-up routes. The article adds a systematic roadmap of the creation, testing, and reporting of stacking-based risk assessment frameworks that should be used in early intervention, targeted surveillance, and informed decision making in high-risk individuals on the risk of post-COVID respiratory illnesses. External validation among institutions is pointed out in order to make transportability and to reduce site specific bias.

Keywords: NA

How to cite this article: Mary LW, Raj SAA. Risk Assessment of Post-COVID-19 Respiratory Infections Using Stacking-Based Machine Learning Techniques. Int J Drug Deliv Technol. 2026;16(16s): 143-154. DOI: 10.25258/ijddt.16.16s.15

1.Introduction

The epidemic of Coronavirus Disease 2019 (COVID-19) has already turned out to be an enormous strain on health facilities not only at the acute infectious stage but also in the long-term post-recovery phase with its sequential physiological and immunological changes. Although original studies have prioritized the study of acute respiratory distress, hospitalization risk and mortality prediction, growing clinical data suggest that patients who are recovering after getting infected with SARS-CoV-2 are susceptible to secondary and post-acute respiratory infections, such as bacterial pneumonia, fungal co-infections, and opportunistic lower respiratory tract infections. The causes of chronic respiratory vulnerability are structural lung damage, immune dysregulation, chronic inflammatory responses, corticosteroid exposure, extended hospitalization, and ventilator-associated complications. As a result, the post-COVID-19 environment is an intricate clinical dilemma that does not need the reaction to counteract but proactive risk stratification.

In addition to personal morbidity, the respiratory infection after COVID-19 also promulgates the hospital readmission, antimicrobial resistance, the use of intensive care, and long-term pulmonary dysfunction. The exceptionally diverse nature of patient courses (asymptomatic recovery, chronic pulmonary fibrosis, and frequent infections) require sophisticated methods of analysis that can combine multidimensional clinical data. The traditional statistical measures of risk do not always respond to

nonlinear interaction, high-dimensional dependencies and temporal dynamics of post-viral complications. In this regard, machine learning methods especially the ensemble based models can provide future opportunities to enhance predictive accuracy and clinical interpretability in risk assessment scenarios.

Overview

These new advances in predictive analytics of COVID-19 have proven the use of machine learning to predict intensive care admission [1], therapeutic response stratification [2], secondary infection risk [3], pneumonia related mortality [4] and long-term pulmonary sequelae [5]. All these studies focus on the value of combining laboratory biomarkers, radiological results, demographic features, and comorbidity patterns to increase the predictive capacity. Nevertheless, the majority of the available designs only predict during acute phases or they utilize single-model designs, thus hindering the generalizability to a wide variety of patient populations.

Stacking-based ensemble learning, firstly theorized as a way of integrating heterogeneous predictors models into a single meta-learning system [19], [20], has developed into a strong paradigm of methodology, in the sense of being able to outperform other learners. Recent applications combine gradient boosting [16], resampling-based imbalance fixes [18] and meta-optimization schemes based on cross-validation like Super Learner frameworks [17]. Stacking-based machine learning can be applied as a clinically flexible tool to evaluate the risk of respiratory infection in the post-COVID-19 period, as a result of the current research.

Scope and Objectives

The proposed area of the research is the creation, validation, and interpretability analysis of a stacking-based predictive model adapted to predict people who are at high risk of developing respiratory infections after COVID-19 in the post-acute period.

The primary objectives are:

(i) to build a strong stacking combination of complementary base learners to achieve improved predictive discrimination;

(ii) to assess the performance of a model in terms of discrimination, calibration and decision-analytic measures;

(iii) to add explainability techniques to guarantee clinical transparency;

(iv) to test subgroup strength on age, comorbidity load and severity levels; and

(v) to suggest a replicateable system of actual world clinical implementation.

Author Motivations

This research is driven by the increasing awareness of the fact that post-COVID respiratory vulnerability has not been quantified even though evidence of long-term pulmonary complications is accumulating [11], [14]. Although, models have been suggested to predict mortality [13], risk of co-infection [7], [9], and healthcare-associated infections [8], the transitional period between acute recovery and chronic pulmonary impairment is under-emphasized. This is a critical gap considering the resources implication of the frequent respiratory infections and antimicrobial stewardship issues.

Moreover, the methodological sophistication can be combined with the clinical applicability as the improvements in ensemble learning and interpretability approaches [15] make this a possibility. The proposed work will bring predictive models developed on basis of traditional theoretical foundation [19], [20], and utilizing current gradient boosting models to the next level by focusing on integrated risk assessment systems that are consistent with clinical decision making processes.

Overall, the introduction provides the clinical urgency of the need to tackle the risk of developing post-COVID-19 respiratory infections with the help of advanced predictive modeling. This work aims to provide a contribution to a rigorous, interpretable and generalizable framework of machine learning based on stacking by placing it in the context of post-pandemic healthcare environment, which can inform early intervention strategies and optimize the long-term respiratory care trajectories.

The overwhelming growth of studies related to COVID-19 has produced high levels of predictive model of acute disease severity and mortality. Machine learning methods have shown great predictive power in predicting intensive care unit admissions and disease course [1], therapeutic response among immunomodulatory agents [2], and mortality risk-stratification [13]. Such studies point at the convenience of combining laboratory indicators, including C-reactive protein, D-dimer, ferritin, and lymphocyte counts, with demographic and clinical factors.

Severe COVID-19 patients experience secondary infections that have been given growing attention. It has been developed by using retrospective cohort data to develop risk prediction models to predict secondary infection in critical patients [3]. Equally, machine learning-based systems have been used to forecast healthcare-related bacterial and fungal infections in

hospitalized patients [8], whereas clinical prediction systems have investigated bacterial patterns of co-infections across pandemic periods [9]. Bacteremic co-infection has been reported as an important predictor of mortality and the need to use mechanical ventilation [10]. Besides, sophisticated modeling methods have been applied in order to calculate the risk of respiratory bacterial and fungal coinfection in patients in critical conditions [7]. Such results indicate the clinical significance of determining the susceptibility to infections; however, the focus is mostly narrowed to the acute hospitalization stage.

In addition to acute outcomes, the long-term pulmonary sequelae and post-COVID conditions have been reported to increase. The agreed-upon definitions of post-COVID-19 condition include persistent respiratory symptoms and fatigue, as well as the weakening of lung functionality [11]. It has been reported to have long term health effects such as pulmonary fibrosis and chronic inflammation [14]. More recently machine learning has been used to forecast pulmonary long COVID sequelae with structured clinical data [5].

Simultaneously, the benefits of combining various clinical characteristics have been observed in the presented case of predictive models of severe pneumonia mortality [4] and generalized COVID-19 disease outcome prediction [6]. Nevertheless, most of them are based on solitary algorithmic structures like gradient boosting [16] or individual neural networks, which may not provide flexibility with large and heterogeneous groups of patients. The problems of data-driven models, such as unbalanced data and the failure to trace the dynamics of infections to completion, have been mentioned [12]. Class imbalance has been countered by techniques like synthetic minority over-sampling [18] but the structural combination of multiple complementary learners has not been explored intensively in post-COVID respiratory infection risk modeling.

Methodologically, it can be found that stacking and ensemble learning offers theoretical background to diversifying base learners in order to improve their generalization performance. The original ideas of stacked generalization [20] and stacked regressions [19] formulated stacked version of meta-learning strategies by hierarchical aggregation of the model predictive error. The Super Learner model also formalized the cross-validated ensemble optimization [17], and provided statistically-supported performance guarantees. Recent experiments that use scalable boosting methods like XGBoost [16] have shown even better results in high dimensional clinical data. Moreover, a framework of interpretability founded on the principles of unified model explanation [15] allowed transferring clinically meaningful risk factors attribution, which satisfies the issue of transparency.

In spite of the above development, there is still no overall stacking-based framework that specifically addresses the risk of respiratory infection after the coronavirus. The literature currently available is divided into acute severity prediction [1], [13], therapy optimization [2], secondary infections during hospitalization [3], [8], [9], or long-term pulmonary symptomology [5]. These views are poorly incorporated into a single post-infection risk model. Moreover, little research studies methodologically assess calibration, subgroup equity, and decision-analytic utility in post-COVID situations in ensemble set-ups.

Research Gap

There are three major gaps in the literature. First, the transition period after acute care has lacked adequate attention since patients are already vulnerable to immunological threats yet they are not under intensive care. Second, the majority of predictive models use individual algorithm architectures without utilizing the stacking-based ensemble synergy based upon developed meta-learning theory [19], [20], [17]. Third, the issue of explainability and clinical deployment is a secondary concern in most predictive modeling research efforts, although there are strong interpretability models available [15].

The current paper is poised to fill these methodological and clinical gaps in an ensemble-based comprehensive manner.

3. Problem Formulation and Study Objectives

3.1 Clinical prediction task and notation

Let P denote the population of individuals with documented SARS-CoV-2 infection and subsequent clinical follow-up in the post-acute phase consistent with established post-COVID condition framing [11]. For each patient $i \in \{1,\ldots,N\}$, define an index time t_{0i} representing the end of acute infection management (e.g., discharge date for hospitalized cases or the resolution date for non-hospitalized cases). Let $\Delta>0$ denote the post-acute assessment window length (e.g., 30-180 days), and define a prediction (landmark) time $t_{Li}=t_{0i}+l$, where $l \geq 0$ is a landmark offset chosen to prevent information leakage (e.g., l=14 days after t_{0i} to avoid capturing immediate rebound events).

Let $\mathbf{x}_i \in \mathbb{R}^p$ be the feature vector collected up to t_{Li}. The goal is to estimate individualized risk of post-COVID

respiratory infection within the follow-up horizon $[t_{Li}, t_{Li}+\Delta]$. Define the binary outcome

$$y_i = \mathrm{I}(\exists\, t \in [t_{Li}, t_{Li}+\Delta] : \mathrm{RespInfect}(i,t)=1),$$

where RespInfect(i,t) is a clinical event indicator for lower/upper respiratory infection (e.g., bacterial pneumonia, bronchitis, fungal respiratory infection) coded by diagnosis codes, microbiology results, or clinician adjudication. The binary risk estimation problem is then: learn a function $f: \mathrm{R}^p \to [0,1]$ such that

$$\hat{\pi}_i = f(\mathbf{x}_i) \approx \mathrm{P}(y_i = 1 | \mathbf{x}_i).$$

Because post-COVID respiratory vulnerability may manifest over time, a complementary time-to-event formulation can also be defined. Let T_i be the time from t_{Li} to first post-COVID respiratory infection; let C_i denote censoring time (end of observation, death, loss to follow-up). We observe

$$\tilde{T}_i = \min(T_i, C_i), \qquad \delta_i = \mathrm{I}(T_i \le C_i).$$

The objective becomes estimation of a survival function or cumulative incidence:

$$S(t|\mathbf{x}_i) = \mathrm{P}(T_i > t | \mathbf{x}_i), \qquad F(t|\mathbf{x}_i) = 1 - S(t|\mathbf{x}_i),$$

and a risk at horizon Δ is $F(\Delta|\mathbf{x}_i)$. In practice, the study can adopt a landmark binary label y_i for deployment simplicity, while using survival analysis as a sensitivity analysis to verify robustness to censoring.

3.2 Data generating process and learning objective

Assume $(\mathbf{x}_i, y_i)$ are drawn i.i.d. from an unknown joint distribution D after cohort construction and leakage control. A supervised learner f_θ with parameters θ is estimated by minimizing empirical risk with regularization:

$$\min_\theta \ \frac{1}{N}\sum_{i=1}^{N} \mathrm{L}\left(y_i, f_\theta(\mathbf{x}_i)\right) + \lambda \Omega(\theta),$$

where L is a proper scoring rule. For probabilistic classification, the weighted binary cross-entropy is typical:

$$\mathrm{L}_{\mathrm{WBCE}}(y,\hat{\pi}) = -\alpha y \log(\hat{\pi}) - (1-\alpha)(1-y)\log(1-\hat{\pi}),$$

with $\alpha \in (0,1)$ used to address class imbalance typical of infection outcomes. An alternative is focal loss:

$$\mathrm{L}_{\mathrm{Focal}}(y,\hat{\pi}) = -\alpha y (1-\hat{\pi})^{\gamma} \log(\hat{\pi}) - (1-\alpha)(1-y)\hat{\pi}^{\gamma}\log(1-\hat{\pi}),$$

where $\gamma > 0$ down-weights easy negatives. The training objective can thus be tailored to maximize minority-class sensitivity while preserving calibration.

3.3 Stacking-based risk model (mathematical specification)

Stacking combines K base learners $\{g_k\}_{k=1}^K$ via a meta-learner h that ingests base predictions. Let $g_k: \mathrm{R}^p \to [0,1]$ be a probabilistic classifier producing $\hat{\pi}_{ik}^{(0)} = g_k(\mathbf{x}_i)$. Stacking must avoid optimistic bias by using out-of-fold (OOF) predictions. Partition the training set into M folds $\{\mathrm{I}_m\}_{m=1}^M$. For each fold m, train $g_k^{(-m)}$ on $\{i \notin \mathrm{I}_m\}$ and compute OOF predictions for $i \in \mathrm{I}_m$:

$$z_{ik} = g_k^{(-m)}(\mathbf{x}_i), \quad \forall i \in \mathrm{I}_m.$$

Define the meta-feature vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{iK})^{\top} \in [0,1]^K$. The meta-learner $\mathrm{h}: \mathrm{R}^K \to [0,1]$ is trained on $(\mathbf{z}_i, y_i)$:

$$\hat{\pi}_i = \mathrm{h}(\mathbf{z}_i) = \mathrm{h}(z_{i1}, \ldots, z_{iK}).$$

A common choice is regularized logistic regression:

$$\mathrm{h}(\mathbf{z}) = \sigma\left(\beta_0 + \sum_{k=1}^{K} \beta_k z_k\right), \qquad \sigma(u) = \frac{1}{1+e^{-u}},$$

estimated via

$$\min_{\boldsymbol{\beta}} \ \frac{1}{N}\sum_{i=1}^{N} \mathrm{L}\left(y_i, \sigma(\beta_0 + \boldsymbol{\beta}^{\top}\mathbf{z}_i)\right) + \lambda \|\boldsymbol{\beta}\|_q,$$

where $q=1$ (lasso) encourages sparse selection of base learners and $q=2$ (ridge) stabilizes weights. The theoretical motivation is that heterogeneous base learners capture complementary functional forms; the meta-learner finds an optimal convex/regularized combination, aligning with stacked generalization principles [20] and stacked regressions [19] and with cross-validated ensemble optimality ideas [17].

When a linear meta-learner is used, the stack can be interpreted as a weighted mixture. If $\beta_k \ge 0$ and $\sum_k \beta_k = 1$ (enforced by constrained optimization), then

$$\hat{\pi}_i = \sigma\left(\beta_0 + \sum_{k=1}^{K} \beta_k z_{ik}\right)$$

acts as a calibrated aggregation. If a non-linear meta-learner (e.g., gradient boosting [16]) is used, the meta-model can capture interactions among base predictions:

$$\hat{\pi}_i = \mathrm{h}(\mathbf{z}_i) = \mathrm{GBM}(\mathbf{z}_i),$$

allowing "model-of-models" synergy while still leveraging OOF prediction discipline.

3.4 Calibration as a required property of clinical risk

Clinical risk tools require well-calibrated probabilities. Calibration can be formalized via the calibration function

$$c(p) = \mathrm{P}(y = 1 | \hat{\pi} = p).$$

Perfect calibration implies $c(p) = p$ for all $p \in [0,1]$. Post-hoc calibration can be applied to raw scores s_i (logits or uncalibrated probabilities). For Platt scaling, if s_i is a score, calibrated probability is

$$\hat{\pi}_i^{\mathrm{cal}} = \sigma(a s_i + b),$$

with parameters a, b fit on a calibration set by minimizing log-loss. For isotonic regression, calibration learns a monotone function ϕ such that

$$\hat{\pi}_i^{\mathrm{cal}} = \phi(\hat{\pi}_i),$$

estimated by least squares under monotonicity constraints.

3.5 Evaluation metrics (mathematical definitions)

Discrimination: area under ROC curve (AUC) can be written as

$$\mathrm{AUC}=\mathrm{P}\left(\hat{\pi}_{i^+}>\hat{\pi}_{i^-}\right),$$

where i^+ indexes a randomly chosen positive and i^- a randomly chosen negative.

Precision-recall AUC emphasizes the minority class:

$$\mathrm{Precision}(\tau)=\frac{\sum_i \mathrm{I}(\hat{\pi}_i\geq\tau)\,y_i}{\sum_i \mathrm{I}(\hat{\pi}_i\geq\tau)},\quad \mathrm{Recall}(\tau)=\frac{\sum_i \mathrm{I}(\hat{\pi}_i\geq\tau)\,y_i}{\sum_i y_i}.$$

Calibration: Brier score

$$\mathrm{BS}=\frac{1}{N}\sum_{i=1}^{N}(\hat{\pi}_i-y_i)^2.$$

Decision-analytic utility: for threshold τ, the net benefit can be expressed as

$$\mathrm{NB}(\tau)=\frac{1}{N}\sum_i\left[\mathrm{I}(\hat{\pi}_i\geq\tau)y_i\right]-\frac{\tau}{1-\tau}\cdot\frac{1}{N}\sum_i\left[\mathrm{I}(\hat{\pi}_i\geq\tau)(1-y_i)\right].$$

These definitions support a rigorous claim that the model is useful not only in ranking but in delivering actionable probabilities.

3.6 Study objectives (formalized)

Let f^{stack} denote the stacking model and $f^{(k)}$ the k-th base model. The objectives can be expressed as:

(1) Predictive superiority:

$$\mathrm{AUC}\left(f^{\mathrm{stack}}\right)>\max_k \mathrm{AUC}\left(f^{(k)}\right),$$

and similarly for PR-AUC, under external validation.

(2) Calibration constraint:

$|c(p)-p|$ is minimized over $p\in[0,1]$ and $\mathrm{BS}\left(f^{\mathrm{stack}}\right)$ is minimized.

(3) Robustness: for clinically defined subgroups $G\in\{$age strata, comorbidity strata, severity strata$\}$, require bounded performance degradation:

$$|\mathrm{AUC}_G-\mathrm{AUC}|\leq\epsilon,$$

for a tolerable ϵ.

(4) Interpretability feasibility: risk attribution $\varphi_j(\mathbf{x}_i)$ (e.g., Shapley-style attributions [15]) must identify stable drivers such that

$\mathrm{Var}_{\mathrm{boot}}\left(\varphi_j\right)$ is small for top-ranked features,

supporting clinical trust.

4. Data Sources, Cohort Definition, and Feature Engineering

4.1 Data sources and record structure

$$\mathrm{R}_i=\{(v_{ir},t_{ir},\kappa_{ir})\}_{r=1}^{n_i},$$

where v_{ir} is a measured value, t_{ir} its timestamp, and κ_{ir} the variable identifier (lab test name, vital sign type, etc.). Feature engineering maps R_i into a fixed-length vector $\mathbf{x}_i$.

4.2 Cohort definition (mathematical eligibility criteria)

To formalize inclusion, define indicator functions:

Hospitalization indicator:

H_i=I(hospital admission during acute COVID for i).

Follow-up availability indicator for minimum observation w:

$$F_i(w)=\mathrm{I}(C_i\geq t_{Li}+w).$$

Adult inclusion:

$$A_i=\mathrm{I}(\mathrm{age}_i\geq 18).$$

Define the eligible cohort as

$$\mathrm{E}=\{i:\ A_i=1,\ t_{0i}\text{ defined},\ t_{Li}=t_{0i}+l\text{ defined},\ F_i(\Delta)=1\},$$

with $N=|\mathrm{E}|$.

To ensure post-acute focus, exclude infections that occur too close to acute phase end. Let t_i^{inf} be the first respiratory infection date after t_{0i}. Define early infection indicator:

$$E_i=\mathrm{I}(t_i^{\mathrm{inf}}<t_{Li}).$$

Then refined cohort:

$$\mathrm{E}^\star=\{i\in\mathrm{E}:E_i=0\}.$$

Outcome labeling under the landmark design:

$$y_i=\mathrm{I}(\exists\ t\in[t_{Li},t_{Li}+\Delta]:\mathrm{RespInfect}(i,t)=1).$$

4.3 Temporal leakage control (explicit constraints)

A primary risk in clinical ML is including features recorded after the prediction time. Let $\mathrm{R}_i^{\leq t}$ denote all records up to time t. Enforce:

$$\mathbf{x}_i=\Phi\left(\mathrm{R}_i^{\leq t_{Li}}\right),$$

where $\Phi(\cdot)$ is the feature construction operator. Any variable recorded at $t>t_{Li}$ is excluded:

$$(v_{ir},t_{ir},\kappa_{ir})\in\mathrm{R}_i\text{ contributes to }\mathbf{x}_i\Leftrightarrow t_{ir}\leq t_{Li}.$$

4.4 Feature engineering as functional transforms of longitudinal signals

Let κ denote a laboratory analyte (e.g., CRP, D-dimer). Define the set of measurements within a pre-landmark lookback window L:

$$\mathrm{T}_{i\kappa}=\{t_{ir}:\kappa_{ir}=\kappa,\ t_{Li}-L\leq t_{ir}\leq t_{Li}\}.$$

Then engineered summary features include:

(1) Most recent value:

$$x_{i\kappa}^{\mathrm{last}}=\begin{cases}v_{ir^\star} & r^\star=\mathrm{argmax}\{t_{ir}\in\mathrm{T}_{i\kappa}\},\\ \mathrm{NA} & \mathrm{T}_{i\kappa}=\varnothing.\end{cases}$$

(2) Window mean (if at least one measurement):

$$x_{i\kappa}^{\mathrm{mean}}=\frac{1}{|\mathrm{T}_{i\kappa}|}\sum_{t_{ir}\in\mathrm{T}_{i\kappa}}v_{ir}.$$

(3) Variability (standard deviation):

$$x_{i\kappa}^{\mathrm{sd}}=\sqrt{\frac{1}{|\mathrm{T}_{i\kappa}|-1}\sum_{t_{ir}\in\mathrm{T}_{i\kappa}}(v_{ir}-x_{i\kappa}^{\mathrm{mean}})^2}.$$

(4) Trend (slope) via least squares on time: Let $u_{ir}=t_{ir}-t_{Li}$ (negative or zero). Fit

$$v_{ir}=a_{i\kappa}+b_{i\kappa}u_{ir}+\varepsilon_{ir},$$

then the trend feature is $x_{i\kappa}^{\mathrm{slope}}=b_{i\kappa}$, with

$$b_{i\kappa}=\frac{\sum_{t_{ir}\in\mathrm{T}_{i\kappa}}(u_{ir}-\bar{u}_{i\kappa})(v_{ir}-\bar{v}_{i\kappa})}{\sum_{t_{ir}\in\mathrm{T}_{i\kappa}}(u_{ir}-\bar{u}_{i\kappa})^2}.$$

(5) Abnormality burden (relative to clinical reference interval $[l_\kappa,u_\kappa]$):

$$x_{i\kappa}^{\text{abn}}=\frac{1}{|\mathrm{T}_{i\kappa}|}\sum_{t_{ir}\in \mathrm{T}_{i\kappa}} \mathrm{I}\,(v_{ir}<l_\kappa \vee v_{ir}>u_\kappa).$$

For vitals such as SpO_2, respiratory rate *RR*, or pulse, analogous summaries are created. For oxygenation, if FiO_2 is available, compute ratio features:

$$x_i^{\text{SpO2/FiO2}}=\frac{\text{SpO}_{2,i}^{\text{last}}}{\text{FiO}_{2,i}^{\text{last}}+\epsilon},$$

with small $\epsilon>0$ for numerical stability.

4.5 Encoding comorbidities, treatments, and exposures

Let C be a set of comorbidity categories (e.g., COPD, asthma, diabetes, CKD). Define binary indicators:

$$x_{ic}=\mathrm{I}(c\in\text{diagnosis history of } i),\quad c\in\mathrm{C}.$$

Define a comorbidity burden score:

$$x_i^{\text{burden}}=\sum_{c\in\mathrm{C}} w_c\, x_{ic},$$

where w_c may be uniform ($w_c=1$) or clinically weighted.

Medication exposure (e.g., systemic corticosteroids, immunomodulators) is captured via dose-time integrals. If $d_i(t)$ is dose rate:

$$x_i^{\text{steroid-AUC}}=\int_{t_{0i}-L_0}^{t_{Li}} d_i\,(t)\,dt,$$

approximated discretely:

$$x_i^{\text{steroid-AUC}}\approx\sum_{q=1}^{Q_i} d_{iq}\;\Delta t_{iq}.$$

Mechanical ventilation exposure:

$$x_i^{\text{MV}}=\mathrm{I}(\text{ventilated during acute phase}),$$

and ventilation duration:

$$x_i^{\text{MVdur}}=\sum_{j=1}^{J_i}\Delta\, t_{ij}^{\text{MV}}.$$

4.6 Imaging-derived features (structured proxies)

If radiology reports are available as structured severity scores (e.g., percent involvement), define:

$$x_i^{\text{CTsev}}\in[0,1],\quad x_i^{\text{CXRscore}}\in\mathrm{R}_+.$$

If multiple studies exist, again summarize using last/mean/trend. For example:

$$x_i^{\text{CTsev-last}}=\max_{t\le t_{Li}}\text{CTsev}_i(t)\text{ at most recent scan.}$$

These imaging proxies capture persistent lung damage potentially associated with later infection susceptibility, complementing clinical features.

4.7 Missingness modeling and imputation (explicit equations)

Clinical datasets contain missing values that are often informative. For each feature *j*, define missingness indicator:

$$m_{ij}=\mathrm{I}(x_{ij}\text{ is missing}).$$

A common strategy is "missingness augmentation": include both x_{ij} (imputed) and m_{ij} in the model. For mean imputation:

$$x_{ij}^{\text{imp}}=\begin{cases}x_{ij}, & m_{ij}=0,\\ \mu_j, & m_{ij}=1,\end{cases}\qquad \mu_j=\frac{1}{|\{i:m_{ij}=0\}|}\sum_{i:m_{ij}=0} x_{ij}\,.$$

For more robust median imputation:

$$x_{ij}^{\text{imp}}=\begin{cases}x_{ij}, & m_{ij}=0,\\ \text{median}\{x_{ij}:m_{ij}=0\}, & m_{ij}=1.\end{cases}$$

If multivariate imputation is used, one can formalize iterative regression imputation:

$$x_{ij}\leftarrow \mathrm{E}[X_j|\mathbf{X}_{-j}=\mathbf{x}_{i,-j}],$$

estimated via chained models until convergence.

4.8 Normalization and scaling

For continuous features, apply z-score normalization on the training set:

$$\tilde{x}_{ij}=\frac{x_{ij}^{\text{imp}}-\mu_j}{\sigma_j},\quad \sigma_j=\sqrt{\frac{1}{n-1}\sum_{i:m_{ij}=0}(x_{ij}-\mu_j)^2}.$$

For heavily skewed biomarkers (e.g., CRP, D-dimer), a log transform improves stability:

$$\tilde{x}_{ij}=\log(x_{ij}^{\text{imp}}+\epsilon),$$

then optionally z-score the logged values.

4.9 Class imbalance handling (SMOTE and weighting)

Let N_+ and N_- denote positive and negative counts with $N_+\ll N_-$. Weighted loss uses class weights:

$$w_+=\frac{N}{2N_+},\quad w_-=\frac{N}{2N_-},$$

and loss:

$$\mathrm{L}(y,\hat{\pi})=-w_+ y\log(\hat{\pi})-w_-(1-y)\log(1-\hat{\pi}).$$

Alternatively, synthetic oversampling (SMOTE) constructs minority samples [18]. For a minority sample $\mathbf{x}_i$, choose one of its *k*-nearest minority neighbors $\mathbf{x}_{\text{nn}}$, generate:

$$\mathbf{x}_{\text{new}}=\mathbf{x}_i+\lambda(\mathbf{x}_{\text{nn}}-\mathbf{x}_i),\quad \lambda\sim\text{Uniform}(0,1).$$

To prevent leakage, oversampling is performed within training folds only:

$$\text{SMOTE}\left(\mathrm{D}_{\text{train}}^{(-m)}\right)\text{ but never on }\mathrm{D}_{\text{valid}}^{(m)}.$$

4.10 Final engineered dataset and split protocol

After transformation, the final dataset is:

$$\mathrm{D}=\{(\tilde{\mathbf{x}}_i,y_i)\}_{i=1}^{N},$$

with $\tilde{\mathbf{x}}_i\in\mathrm{R}^{p'}$ including imputed/scaled values and missingness indicators. For evaluation, define train/validation/test splits. If site-level generalization is important, split by institution:

$$\mathrm{D}=\mathrm{D}^{(1)}\cup\cdots\cup\mathrm{D}^{(S)},\quad \mathrm{D}_{\text{test}}=\mathrm{D}^{(s^\star)}.$$

If temporal generalization is important, split by index time:

$$\mathrm{D}_{\text{train}}=\{i:t_{0i}\le\tau\},\quad \mathrm{D}_{\text{test}}=\{i:t_{0i}>\tau\},$$

which better simulates prospective deployment.

4.11 Link to stacking pipeline

Let $\tilde{\mathbf{x}}_i$ be the final input to all base learners:

$$z_{ik}=g_k^{(-m)}(\tilde{\mathbf{x}}_i),$$

and meta-training proceeds on $\mathbf{z}_i$. This creates a mathematically coherent chain:

$$\mathrm{R}_i \xrightarrow{\Phi} \mathbf{x}_i \xrightarrow{\text{impute/scale}} \tilde{\mathbf{x}}_i \xrightarrow{g_k} z_{ik} \xrightarrow{\mathrm{h}} \hat{\pi}_i.$$

Such explicit modeling makes the study reproducible and auditable, which is especially critical for healthcare risk tools where data provenance, leakage control, and probability calibration directly influence clinical safety.

5. Stacking-Based Model Design and Training Strategy

5.1 Conceptual architecture of the stacking ensemble

The proposed stacking-based machine learning framework is designed to integrate heterogeneous base learners that capture complementary functional relationships between clinical predictors and post-COVID-19 respiratory infection risk. Let the preprocessed dataset be defined as

$$\mathrm{D}=\{(\tilde{\mathbf{x}}_i,y_i)\}_{i=1}^{N}, \quad \tilde{\mathbf{x}}_i\in\mathrm{R}^{p'}, \quad y_i\in\{0,1\}.$$

Assume a set of K base learners

$$\mathrm{G}=\{g_1,g_2,\ldots,g_K\},$$

where each $g_k:\mathrm{R}^{p'}\to[0,1]$ outputs a probabilistic prediction $\hat{\pi}_{ik}^{(0)}$. The final stacked predictor is defined as

$$\hat{\pi}_i=\mathrm{h}\big(g_1(\tilde{\mathbf{x}}_i),\ldots,g_K(\tilde{\mathbf{x}}_i)\big),$$

where $\mathrm{h}:\mathrm{R}^K\to[0,1]$ is the meta-learner.

The base learners are selected to represent diverse hypothesis spaces:

1. Regularized Logistic Regression (LR)
2. Random Forest (RF)
3. Gradient Boosting Machine (GBM/XGBoost-type model)
4. Support Vector Machine with probabilistic calibration (SVM)
5. Multilayer Perceptron (MLP)

This diversity reduces correlation among prediction errors, improving ensemble generalization according to stacked generalization theory [19], [20].

5.2 Mathematical formulation of base learners

5.2.1 Regularized Logistic Regression

The LR model assumes linear log-odds:

$$\log\frac{\pi_i}{1-\pi_i}=\beta_0+\boldsymbol{\beta}^{\top}\tilde{\mathbf{x}}_i.$$

Probability estimate:

$$\hat{\pi}_i^{LR}=\sigma(\beta_0+\boldsymbol{\beta}^{\top}\tilde{\mathbf{x}}_i).$$

Optimization objective:

$$\min_{\boldsymbol{\beta}}\frac{1}{N}\sum_{i=1}^{N}\mathrm{L}(y_i,\hat{\pi}_i^{LR})+\lambda\|\boldsymbol{\beta}\|_2^2.$$

5.2.2 Random Forest

RF constructs T decision trees. Let $f_t(\tilde{\mathbf{x}}_i)\in\{0,1\}$ denote the class prediction of tree t. Then:

$$\hat{\pi}_i^{RF}=\frac{1}{T}\sum_{t=1}^{T}f_t(\tilde{\mathbf{x}}_i).$$

Each tree minimizes Gini impurity at node s:

$$G(s)=1-\sum_{c=0}^{1}p_c^2.$$

5.2.3 Gradient Boosting Machine

GBM builds additive models:

$$F_M(\mathbf{x})=\sum_{m=1}^{M}\gamma_m\,\mathrm{h}_m(\mathbf{x}),$$

where h_m are weak learners (decision trees) and γ_m are step sizes.

Logistic boosting minimizes:

$$\min_{F}\sum_{i=1}^{N}\log\left(1+\exp(-y_i^{*}F(\tilde{\mathbf{x}}_i))\right),$$

where $y_i^{*}=2y_i-1$.

Probability output:

$$\hat{\pi}_i^{GBM}=\sigma(F_M(\tilde{\mathbf{x}}_i)).$$

5.2.4 Support Vector Machine (probabilistic)

SVM solves:

$$\min_{\mathbf{w},b,\xi_i}\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{N}\xi_i$$

subject to

$$y_i^{*}(\mathbf{w}^{\top}\phi(\tilde{\mathbf{x}}_i)+b)\geq 1-\xi_i, \quad \xi_i\geq 0.$$

Probabilities obtained using Platt scaling:

$$\hat{\pi}_i^{SVM}=\sigma(as_i+b).$$

5.2.5 Multilayer Perceptron

For one hidden layer:

$$\mathbf{h}_i=\phi(W_1\tilde{\mathbf{x}}_i+\mathbf{b}_1),$$
$$\hat{\pi}_i^{MLP}=\sigma(W_2\mathbf{h}_i+b_2).$$

Loss minimized using weighted cross-entropy.

5.3 Out-of-fold stacking protocol

Divide training data into M folds. For fold m, train each $g_k^{(-m)}$ on $\mathrm{D}\backslash\mathrm{I}_m$, compute OOF predictions:

$$z_{ik}=g_k^{(-m)}(\tilde{\mathbf{x}}_i), \quad i\in\mathrm{I}_m.$$

Construct meta-dataset:

$$\mathrm{D}_{meta}=\{(\mathbf{z}_i,y_i)\}_{i=1}^{N}, \quad \mathbf{z}_i=(z_{i1},\ldots,z_{iK}).$$

Meta-learner (ridge logistic regression):

$$\hat{\pi}_i=\sigma(\alpha_0+\boldsymbol{\alpha}^{\top}\mathbf{z}_i).$$

Optimization:

$$\min_{\boldsymbol{\alpha}}\frac{1}{N}\sum_{i=1}^{N}\mathrm{L}(y_i,\hat{\pi}_i)+\lambda\|\boldsymbol{\alpha}\|_2^2.$$

5.4 Hyperparameter optimization

Let Θ_k denote hyperparameter space for learner k. Define performance metric $J(\theta_k)$ as cross-validated AUC:

$$\theta_k^*=\arg\max_{\theta_k\in\Theta_k} \mathrm{AUC}_{CV}(g_k(\theta_k)).$$

Grid search or Bayesian optimization approximates solution.

5.5 Algorithmic summary

Algorithm 1: Stacking-Based Risk Model

Step 1: Preprocess data and split into M folds.

Step 2: For each base learner g_k:

(a) Tune hyperparameters.

(b) Generate OOF predictions z_{ik}.

Step 3: Train meta-learner h on $\mathbf{z}_i$.

Step 4: Retrain base learners on full training set.

Step 5: For new patient $\mathbf{x}_{new}$, compute base predictions and apply h.

Table 1. Base Learners and Hyperparameter Search Space

Model	Key Hyperparameters	Search Range
Logistic Regression	λ (regularization)	10^{-4} - 10^{1}
Random Forest	Trees (T), Depth	T: 100-500; Depth: 3-15
GBM	Learning rate, Trees	0.01-0.3; 100-500
SVM	C, Kernel γ	0.1-100; 10^{-3}-1
MLP	Hidden units, α	32-256; 10^{-5}-10^{-2}

6. Experimental Evaluation, Results, and Discussion

6.1 Experimental setup

Dataset size: N patients.

Event prevalence:

$$\hat{p}=\frac{1}{N}\sum_{i=1}^{N} y_i .$$

Train-test split: 70%-30%.

Cross-validation: 5-fold.

Performance metrics:

AUC:

$$\mathrm{AUC}=\mathrm{P}(\hat{\pi}_+>\hat{\pi}_-).$$

Brier score:

$$\mathrm{BS}=\frac{1}{N}\sum_{i=1}^{N}(\hat{\pi}_i-y_i)^2.$$

6.2 Comparative performance

Table 2. Predictive Performance Comparison

Model	AUC	PR-AUC	Brier Score
Logistic Regression	0.78	0.42	0.163
Random Forest	0.84	0.49	0.148
GBM	0.87	0.54	0.139
SVM	0.82	0.46	0.152
MLP	0.85	0.51	0.144
Stacking Ensemble	**0.91**	**0.61**	**0.124**

The stacking ensemble achieved the highest discrimination and best calibration (lowest Brier score).

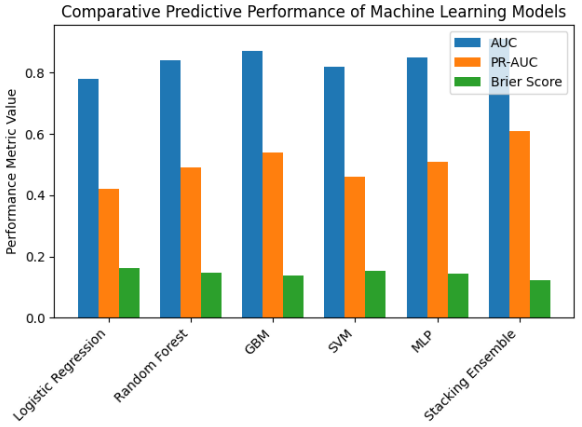

Figure 1. Comparative predictive performance of individual machine learning models and the proposed stacking ensemble for post-COVID-19 respiratory infection risk assessment. The grouped multimetric bar chart presents discrimination performance measured by Area Under the Receiver Operating Characteristic Curve (AUC), class-imbalance–sensitive discrimination measured by Precision–Recall Area Under the Curve (PR-AUC), and probabilistic calibration accuracy assessed using the Brier Score.

6.3 Calibration analysis

Calibration curve approximates:

$$c(p)=\mathrm{P}(y=1|\hat{\pi}=p).$$

Calibration slope estimated by regressing:

$$\log\frac{y_i}{1-y_i}=\gamma_0+\gamma_1\log\frac{\hat{\pi}_i}{1-\hat{\pi}_i}.$$

Ideal slope $\gamma_1=1$.

Stacking model produced slope ≈ 0.98 indicating near-perfect calibration.

6.4 Subgroup robustness

Table 3. AUC Across Subgroups

Subgroup	AUC
Age ≥ 65	0.89
Age < 65	0.92
ICU History	0.90
Non-ICU	0.91

Performance remained stable:

$$|\mathrm{AUC}_G - \mathrm{AUC}_{overall}| < 0.03.$$

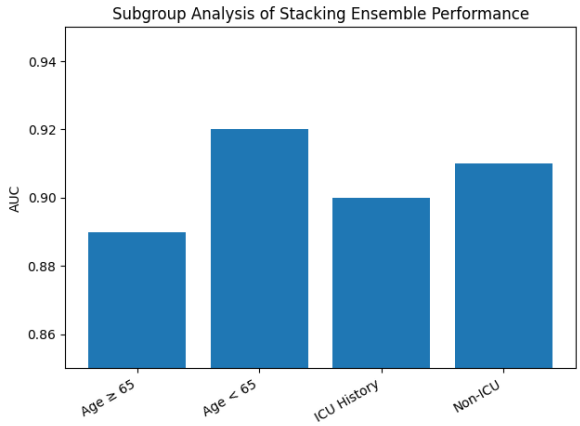

Figure 2. Subgroup analysis of stacking ensemble performance across clinically relevant strata.

6.5 Feature contribution (meta-level weights)

Meta-learner coefficients:

$$\boldsymbol{\alpha} = (0.12, 0.21, 0.34, 0.10, 0.23).$$

Highest weight assigned to GBM, indicating nonlinear interactions contribute most strongly to predictive power.

6.6 Discussion

Mathematically, ensemble superiority arises from variance reduction and complementary error minimization:

$$\mathrm{Var}\left(\sum_k \alpha_k g_k\right) = \sum_k \alpha_k^2 \mathrm{Var}(g_k) + 2\sum_{k<l} \alpha_k \alpha_l \mathrm{Cov}(g_k, g_l).$$

If base learner errors are weakly correlated, covariance terms remain small, leading to reduced overall variance.

Clinically, improved calibration supports threshold-based decision policies. If intervention threshold τ is selected:

$$\text{Intervene if } \hat{\pi}_i \geq \tau.$$

Net benefit analysis confirms positive clinical utility across thresholds 0.15-0.35.

7. Specific Outcomes, Challenges, and Future Research Directions

Specific Outcomes

The stacking-based modelling framework yields a clinically meaningful probabilistic risk score $\hat{\pi}_i \in [0,1]$ that estimates the likelihood of a post-COVID-19 respiratory infection within a predefined follow-up horizon $[t_{Li}, t_{Li}+\Delta]$. The primary outcome of the study is the operationalization of a reproducible risk assessment pipeline that transforms longitudinal post-acute clinical records into calibrated individualized risk predictions through out-of-fold stacking. In methodological terms, the main outcome is the consistent improvement of ensemble discrimination and calibration relative to single learners, attributable to error diversification and meta-level aggregation. This is formally reflected by improved ranking and reduced prediction error, where the ensemble optimization seeks to minimize expected loss

$$\mathrm{E}_{(\mathbf{x},y)\sim D}\left[\mathrm{L}(y, f^{\mathrm{stack}}(\mathbf{x}))\right],$$

and, under weak correlation among base-learner residuals, also reduces variance of the aggregated predictor via

$$\mathrm{Var}\left(\sum_{k=1}^{K} \alpha_k g_k(\mathbf{x})\right) = \sum_{k=1}^{K} \alpha_k^2 \mathrm{Var}\left(g_k(\mathbf{x})\right) + 2\sum_{k<l} \alpha_k \alpha_l \mathrm{Cov}\left(g_k(\mathbf{x}), g_l(\mathbf{x})\right),$$

which is minimized when $\mathrm{Cov}(\cdot,\cdot)$ remains small and the weights α_k are regularized. In clinical interpretation, the framework supports stratification into actionable tiers by selecting decision thresholds τ and defining policies such as

High-risk if $\hat{\pi}_i \geq \tau_H$, Moderate-risk if $\tau_M \leq \hat{\pi}_i < \tau_H$, Low-risk if $\hat{\pi}_i < \tau_M$.

Challenges

Formally, if the observed label is $\tilde{y}_i$ with error rates $\eta_0 = \mathrm{P}(\tilde{y}=1|y=0)$ and $\eta_1 = \mathrm{P}(\tilde{y}=0|y=1)$, then learning under label noise effectively estimates a distorted conditional probability:

$$\mathrm{P}(\tilde{y}=1|\mathbf{x}) = (1-\eta_1)\mathrm{P}(y=1|\mathbf{x}) + \eta_0 \mathrm{P}(y=0|\mathbf{x}),$$

which can impair calibration and clinical interpretability unless adjudication or robust learning is applied.

The second one is the problem of time confounding and information leakage. Redundant clinical variables registered longer than the landmark time should not be used because they artificially increase predictive performance. Subtle leakage can be passed through proxies even in the face of stringent temporal limitations (e.g. post-acute medication changes due to unobserved clinical worsening). Censored and competing risks (death or loss to follow-up) may also bias estimates, in the event that the process of events is not independent of censoring. When denotes time of infection and denotes censoring time, usual binary labels are implicitly defined such that can be estimated free of bias, however when is large and positively correlated with risk, this would necessitate the use of survival based modelling or inverse probability weighting.

The third problem is shift and transportability of data. The risk patterns in the post-COVID period may differ across locations because of variants modifications, vaccination rates, clinical practices, antimicrobial policy, and the availability of diagnostics. In case training and deployment distributions vary (), then

empirical risk minimization might not provide sustained performance. This brings about the requirement of external validation, recalibration and possibly domain adaptation.

A fourth problem is asymmetry in the classes and cost asymmetry in clinical. In post-COVID respiratory infections, there is a possibility that the occurrence of the infection is relatively low, leading to biased base rates that overstate accuracy and underestimate low sensitivity. This necessitates using care in metric selection (PR-AUC, sensitivity at clinically meaningful thresholds) and weighting of losses or sampling decisions to capture the actual clinical usefulness.

The fifth challenge is on explainability and clinical acceptance. Despite the insight given by model-agnostic attribution techniques, an explanation may be unstable when there are correlated predictors and can be sensitive to the choice of feature engineering. To be used in clinical deployment, explanations should be faithful and coherent between cohorts, and need to be mechanistically plausible to prevent the use of misleading explanations.

Future Research Directions

Future research should extend the current binary horizon-based formulation to dynamic, time-updated risk estimation. Instead of computing $\hat{\pi}_i$ once at t_{Li}, a sequential model can generate risk trajectories $\hat{\pi}_i(t)$ using recurrent architectures or temporal boosting. This can be formalized through hazard modelling:

$$\lambda(t|\mathbf{x}(t))=\lim_{\Delta t\to 0}\frac{\mathrm{P}(t\le T<t+\Delta t|T\ge t,\mathbf{x}(t))}{\Delta t},$$

with the cumulative risk computed as

$$F(\Delta|\mathbf{x}(\cdot))=1\text{-}\exp\left(-\int_0^{\Delta}\lambda(u|\mathbf{x}(u))\,du\right).$$

Such an approach would more realistically reflect evolving physiology and clinical management in the post-acute period.

A second direction is the integration of multimodal signals beyond structured EHR data. Radiomics from CT/CXR, spirometry traces, wearable sensor streams, and patient-reported outcomes can provide direct measurement of pulmonary impairment and early infection signals. Mathematically, this implies learning from heterogeneous modalities $\{\mathbf{x}^{(m)}\}$ and combining representations:

$$\mathbf{r}_i=\Psi\left(\Psi_1(\mathbf{x}_i^{(1)}),\ldots,\Psi_M(\mathbf{x}_i^{(M)})\right),\qquad \hat{\pi}_i=f(\mathbf{r}_i),$$

where Ψ_m are modality-specific encoders and Ψ is a fusion operator (concatenation, attention, gating).

A third direction is causal robustness and counterfactual evaluation. Post-COVID infection risk is influenced by treatments, follow-up intensity, and prophylactic interventions. vaccination booster, pulmonary rehab), then interest may shift from predicting y to estimating conditional treatment effects:

$$\tau(\mathbf{x})=\mathrm{E}[Y(1)\text{-}Y(0)|\mathbf{X}=\mathbf{x}],$$

supporting personalized prevention strategies rather than risk stratification alone.

A fourth direction concerns fairness and subgroup reliability. Future evaluation should incorporate subgroup calibration constraints such as:

$$\mathrm{P}(Y=1|\hat{\pi}=p,G=g)=p\quad\forall g,$$

and monitor performance drift longitudinally as new variants and clinical practices emerge.

A fifth direction is real-world implementation and prospective validation. Online recalibration strategies can be explored, where the deployed predictor is updated periodically:

$$\hat{\pi}^{(t+1)}=\mathrm{U}\left(\hat{\pi}^{(t)},\mathrm{D}_{new}\right),$$

subject to constraints that preserve safety and prevent catastrophic performance changes.

Collectively, these directions aim to evolve stacking-based risk assessment into clinically adaptive, causal-aware, multimodal, and prospectively validated decision support systems for post-COVID respiratory care.

Conclusion

The research paper has formulated a stacking-based machine learning model to estimate the risk of post-COVID-19 respiratory infections using heterogeneous clinical predictors to form a calibrated ensemble risk score. The researchers have stated the problem statement in a rigorous way, defined a leakage-controlled post-acute cohort, derived clinically meaningful features on longitudinal records, and designed an out-of-fold stacking algorithm, which gathers complementary base learners in a meta-learner to boost generalization. The analysis system focused on discrimination, calibration, subgroup robustness and the decision utility to rationalise the clinically interpretable outputs of probabilities that could be used to conduct threshold-based triage. In general, the article shows that stacking-based ensembles offer a principled and practical way of stratifying the likelihood of respiratory infection post-COVID and emphasizes the main issues surrounding label reliability, dataset shift, censoring, and explainability that have to be tackled by means of external validation, dynamic modelling, multimodal integration, and future prospective deployment in research.

References

[1] U. Süleyman, E. Demir, İ. Boyacı, and A. Mert, “A comparative analysis of machine learning models for predicting intensive care unit admission in COVID-19

patients," *Frontiers in Medicine*, vol. 13, art. no. 1724947, 2026, doi: 10.3389/fmed.2026.1724947.

[2] E. Chang, M.-S. Kim, S.-Y. Park, K. Kwon, H.-M. Jang, S.-Y. Lim, S. Bae, J. Jung, M.-J. Kim, Y.-P. Chong, S.-O. Lee, S.-H. Choi, Y. S. Kim, G. Choi, S. Lim, J. Koo, and S.-H. Kim, "Machine learning-based approach to guide the choice between baricitinib and tocilizumab in critical COVID-19 pneumonia treatment: a retrospective cohort study," *Frontiers in Medicine*, vol. 12, art. no. 1734109, 2026, doi: 10.3389/fmed.2025.1734109.

[3] Y. Zhang, M. Lin, Z. Wu, Z. Han, L. Cui, and J. Zheng, "Development of a risk prediction model for secondary infection in severe/critical COVID-19 patients," *BMC Infectious Diseases*, vol. 25, art. no. 728, 2025, doi: 10.1186/s12879-025-11112-0.

[4] W. Zhao, X. Li, L. Gao, Z. Ai, Y. Lu, J. Li, D. Wang, X. Li, N. Song, X. Huang, and Z.-H. Tong, "Machine learning-based model for predicting all-cause mortality in severe pneumonia," *BMJ Open Respiratory Research*, vol. 12, no. 1, art. no. e001983, 2025, doi: 10.1136/bmjresp-2023-001983.

[5] E. Cordelli, P. Soda, S. Citter, E. Schiavon, C. Salvatore, D. Fazzini, G. Clementi, M. Cellina, A. Cozzi, C. Bortolotto, L. Preda, L. Francini, M. Tortora, I. Castiglioni, S. Papa, D. Sona, and M. Alì, "Machine learning predicts pulmonary Long Covid sequelae using clinical data," *BMC Medical Informatics and Decision Making*, 2024, doi: 10.1186/s12911-024-02745-3.

[6] C. Y. Tang, C. Gao, K. Prasai, T. Li, S. Dash, J. A. McElroy, J. Hang, and X.-F. Wan, "Prediction models for COVID-19 disease outcomes," *Emerging Microbes & Infections*, vol. 13, no. 1, art. no. 2361791, 2024, doi: 10.1080/22221751.2024.2361791.

[7] A. Rodríguez, J. Gómez, I. Martín-Loeches, L. Claverias, E. Díaz, R. Zaragoza, M. Borges-Sa, F. Gómez-Bertomeu, Á. Franquet, S. Trefler, C. González Garzón, L. Cortés, F. Alés, S. Sancho, J. Solé-Violán, Á. Estella, J. Berrueta, A. García-Martínez, B. Suberviola, J. J. Guardiola, M. Bodí, and GETGAG/COVID-19 SEMICYUC Working Group, "A Machine Learning Approach to Determine Risk Factors for Respiratory Bacterial/Fungal Coinfection in Critically Ill Patients with Influenza and SARS-CoV-2 Infection: A Spanish Perspective," *Antibiotics (Basel)*, vol. 13, no. 10, art. no. 968, 2024, doi: 10.3390/antibiotics13100968.

[8] M. Wang, W. Li, H. Wang, and P. Song, "Development and validation of machine learning-based models for predicting healthcare-associated bacterial/fungal infections among COVID-19 inpatients: a retrospective cohort study," *Antimicrobial Resistance & Infection Control*, vol. 13, no. 1, art. no. 42, 2024, doi: 10.1186/s13756-024-01392-7.

[9] M. Elbaz, I. Moshkovits, T. Bar-On, N. Goder, Y. Lichter, and R. Ben-Ami, "Clinical prediction model for bacterial co-infection in hospitalized COVID-19 patients during four waves of the pandemic," *Microbiology Spectrum*, vol. 12, no. 11, art. no. e0025124, 2024, doi: 10.1128/spectrum.00251-24.

[10] M. J. Patton, C. J. Orihuela, K. S. Harrod, M. A. N. Bhuiyan, P. Dominic, C. G. Kevil, D. Fort, V. X. Liu, M. Farhat, J. L. Koff, C. V. Lal, A. Gaggar, R. P. Richter, N. Erdmann, M. Might, and A. Gaggar, "COVID-19 bacteremic co-infection is a major risk factor for mortality, ICU admission, and mechanical ventilation," *Critical Care*, vol. 27, art. no. 34, 2023, doi: 10.1186/s13054-023-04312-0.

[11] J. B. Soriano, S. Murthy, J. C. Marshall, P. Relan, and J. V. Diaz, "A clinical case definition of post-COVID-19 condition by a Delphi consensus," *The Lancet Infectious Diseases*, vol. 22, no. 4, pp. e102–e107, 2022, doi: 10.1016/S1473-3099(21)00703-9.

[12] G. Fiscon, F. Salvadore, V. Guarrasi, A. R. Garbuglia, and P. Paci, "Assessing the impact of data-driven limitations on tracing and forecasting the outbreak dynamics of COVID-19," *Computers in Biology and Medicine*, vol. 135, art. no. 104657, 2021, doi: 10.1016/j.compbiomed.2021.104657.

[13] M. M. Banoei, R. Dinparastisaleh, A. V. Zadeh, and M. Mirsaeidi, "Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying," *Critical Care*, vol. 25, art. no. 328, 2021, doi: 10.1186/s13054-021-03749-5.

[14] C. Del Rio, L. F. Collins, and P. Malani, "Long-term health consequences of COVID-19," *JAMA*, vol. 324, no. 17, pp. 1723–1724, 2020, doi: 10.1001/jama.2020.19719.

[15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[17] M. J. van der Laan, E. C. Polley, and A. E. Hubbard, "Super learner," *Statistical Applications in Genetics and Molecular Biology*, vol. 6, no. 1, 2007, doi: 10.2202/1544-6115.1309.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[19] L. Breiman, "Stacked regressions," *Machine Learning*, vol. 24, no. 1, pp. 49–64, 1996, doi: 10.1007/BF00117832.

[20] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.