RESEARCH ARTICLE

JOURNAL OF
MEDICAL VIROLOGY WILEY

# Long-term SARS-CoV-2 neutralizing antibody level prediction using multimodal deep learning: A prospective cohort study on longitudinal data in Wuhan, China

Cong Fang[1]  |  Weiming Yan[2]  |  Yuying Chen[2]  |  Zhiyong Dou[1]  |  Tingting Liu[2]  |
Fengning Luo[3]  |  Weiwei Chen[4]  |  Xitang Li[2]  |  Yajie Chen[1]  |  Wenhui Wu[2]  |
Zhize Yuan[2]  |  Yuxin Niu[2]  |  Peng Wang[2]  |  Wenzhen Zhu[4]  |  Xiaoping Luo[5]  |
Tao Chen[2]  |  Xiang Bai[6]  |  Xiaojing Wang[2]  |  Qin Ning[2] 🄶

[1]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China

[2]State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Department and Institute of Infectious Disease, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Computer Science, University of Toronto, Toronto, Canada

[4]State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[5]State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Department of Pediatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[6]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

**Correspondence**
Qin Ning, Xiaojing Wang, and Tao Chen, State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Department and Institute of Infectious Disease, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
Email: qning@vip.sina.com, wxjmoon@hotmail.com and chentao_tjh@vip.sina.com

Xiang Bai, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.
Email: xbai@hust.edu.cn

## Abstract

The ongoing epidemic of SARS-CoV-2 is taking a substantial financial and health toll on people worldwide. Assessing the level and duration of SARS-CoV-2 neutralizing antibody (Nab) would provide key information for government to make sound healthcare policies. Assessed at 3-, 6-, 12-, and 18-month postdischarge, we described the temporal change of IgG levels in 450 individuals with moderate to critical COVID-19 infection. Moreover, a data imputation framework combined with a novel deep learning model was implemented to predict the long-term Nab and IgG levels in these patients. Demographic characteristics, inspection reports, and CT scans during hospitalization were used in this model. Interpretability of the model was further validated with Shapely Additive exPlanation (SHAP) and Gradient-weighted Class Activation Mapping (GradCAM). IgG levels peaked at 3 months and remained stable in 12 months postdischarge, followed by a significant decline in 18 months postdischarge. However, the Nab levels declined from 6 months postdischarge. By training on the cohort of 450 patients, our long-term antibody prediction (LTAP) model could predict long-term IgG levels with relatively high area

under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1-score, which far exceeds the performance achievable by commonly used models. Several prognostic factors including FDP levels, the percentages of T cells, B cells and natural killer cells, older age, sex, underlying diseases, and so forth, served as important indicators for IgG prediction. Based on these top 15 prognostic factors identified in IgG prediction, a simplified LTAP model for Nab level prediction was established and achieved an AUC of 0.828, which was 8.9% higher than MLP and 6.6% higher than LSTM. The close correlation between IgG and Nab levels making it possible to predict long-term Nab levels based on the factors selected by our LTAP model. Furthermore, our model identified that coagulation disorders and excessive immune response, which indicate disease severity, are closely related to the production of IgG and Nab. This universal model can be used as routine discharge tests to identify virus-infected individuals at risk for recurrent infection and determine the optimal timing of vaccination for general populations.

**KEYWORDS**

antibody prediction, COVID-19, data imputation, deep learning, longitudinal data, multimodal data fusion, prospective study

# 1 | BACKGROUND

The novel coronavirus disease 2019 (COVID-19),[1,2] caused by SARS-CoV-2,[3] has lasted more than 2 years since the end of 2019 with no signs of ending in the near future. With high infection and fatality rates, the COVID-19 pandemic has resulted in more than 662 million cases of infection worldwide and over 6.4 million deaths by December 5, 2022.[4–7] Furthermore, a long-term follow-up of severe patients also revealed multiple organ damage following discharge.[8] Evaluating long-term immune responses of post-SARS-CoV-2 infection provides critical information for preventing recurrent infection. It additionally serves as a remarkable reference for determining the optimal timing of vaccination for the general population and SARS-CoV-2-infected individuals. It was reported that IgG responses occurred with a mean day of 12 days, peaked at about 25 days postinfection, and might decline at 60 days.[9] A recent study also pointed out that RBD-specific IgG and plasma neutralizing activity remained stable from 6 to 12 months.[10]

Despite extensive research efforts,[11,12] data are still lacking in understanding whether significant immunologic differences exist among patients with different demographic and clinical characteristics in large-scale cohorts over 1 year. Moreover, it is unclear whether the trend of neutralizing activities was similar to IgG levels, and whether the neutralizing activities can be roughly predicted using IgG levels. Finally, there is also a paucity of models to predict long-term Nab levels based on hospitalized patients' demographic and clinical characteristics. Finding such a model brings substantial economic cost savings for underdeveloped regions where antibody testing is unavailable.

Widely used in medical auxiliary diagnosis systems, methods based on deep learning have played an important role in the battle against the COVID-19 epidemic,[13,14] such as disease diagnosis,[15,16] lesion segmentation,[17,18] severe disease prediction,[19,20] and drug repurposing,[21,22] and so forth. Existing studies also reported basic models (e.g., linear regression and multiple linear regression) to predict the Nab longevity[23] and to find out the independent predictors of IgG levels.[24] However, a lack of robust models that accurately predict long-term Nab levels of COVID-19 patients has largely hindered the evaluation of SARS-CoV-2 protection and vaccination strategy. Concurrently, the pervasive implementation and accelerated advancement of interpretable models[25,26] have facilitated the development of tools to discern crucial factors influencing long-term antibody levels, thereby enabling further investigation of the pathophysiological mechanisms underlying antibody production. Nevertheless, several common pitfalls in real-world medical data and assisted diagnosis systems limit the development of reliable models, including high data missing rates, nonequidistant spacing of longitudinal data collection, and difficulty in fusing multimodal heterogeneous data.

To alleviate the above-mentioned problems, we developed a robust imputation framework to reliably transform the highly sparse datasets. We then utilized multimodal heterogeneous data, including demographic characteristics, past medical history, inspection reports, and CT scans during hospitalization to implement a novel deep learning model that accurately predicts the long-term Nab and IgG levels of COVID-19 patients in a large prospective cohort study from Wuhan, China. When taken as a whole, our study not only explained the temporal changes in IgG and neutralizing activities but also overcame technical barriers to create a long-term antibody prediction model that can be used in real-world settings. This will result in significant financial savings for the healthcare system.

## 2 | MATERIALS AND METHODS

### 2.1 | Study design and participants

A total of 3050 patients with laboratory-confirmed (RT-PCR positive) or suspected (based on epidemiological history, clinical manifestations, and laboratory results) COVID-19 were admitted to Tongji Hospital from February 10 to April 30, 2020. Unfortunately, 205 patients died with a mortality of 6.72% for the first breakout with over 85% of them being severe and critically ill cases in Tongji hospital. Of which, 1422 patients were successfully contacted after hospital discharge, and finally 450 were enrolled in the follow-up study. The reasons for none-nrolment included: 203 patients with suspected COVID-19 who tested negative for viral RNA during hospitalization; 1220 patients who were not contacted after discharge; 828 patients who declined to participate or were unwilling to comply with all study requirements; and 144 patients who did not have complete medication information. Of 450 enrolled patients, 450, 359, 350, and 77 completed 3-, 6-, 12-, and 18-month postdischarge follow-up visits, respectively (Figure 1).

We also measured Nab levels at V1-V3 in a representative group of 45 patients. These patients were categorized into three groups according to disease severity: moderate group, severe group and critically ill group. A trimatch propensity score analysis was applied to control potential confounders among the three groups by using R language software. A total of nine covariates including age, sex, disease severity and comorbidities (Hypertension, diabetes, coronary heart disease, cerebro vascular disease, hepatitis, and chronic obstructive pulmonary disease [COPD]) were included, and the caliper value was taken as 0.08. After matching, a total of 45 patients were included for analysis with 11, 16, 18 patients for moderate, severe and critical group, respectively.

### 2.2 | IgG detection

Based on nephelometric methods, siemens BNII human IgG assay was conducted on a BNII system (Dade Behring Canada Inc, Siemens Company). The use of polystyrene particles guaranteed a high sensitivity and specificity of antibody detection against IgG. Serum samples were used for testing. N protein standard SL (OQIM) was used for calibration. All of the calibrators, reagents, and controls were provided by Siemens and were according to the manufacturer's instructions.

### 2.3 | Pseudovirus neutralization assay

The SARS-CoV-2 Nab assay was conducted based on a vesicular stomatitis virus (VSV) pseudovirus system, which presents the SARS-CoV-2 spike protein on the surface of the VSV particle. The pseudoviruses were produced and titrated by the above methods.[27] The SARS-CoV-2 Nab titer was defined as 50% inhibitory dilution ($ID_{50}$), which equals the serum dilution. At this serum dilution, the relative light units (RLUs) were reduced by 50% compared with the virus control wells after subtracting the background RLUs from the control groups with only cells.
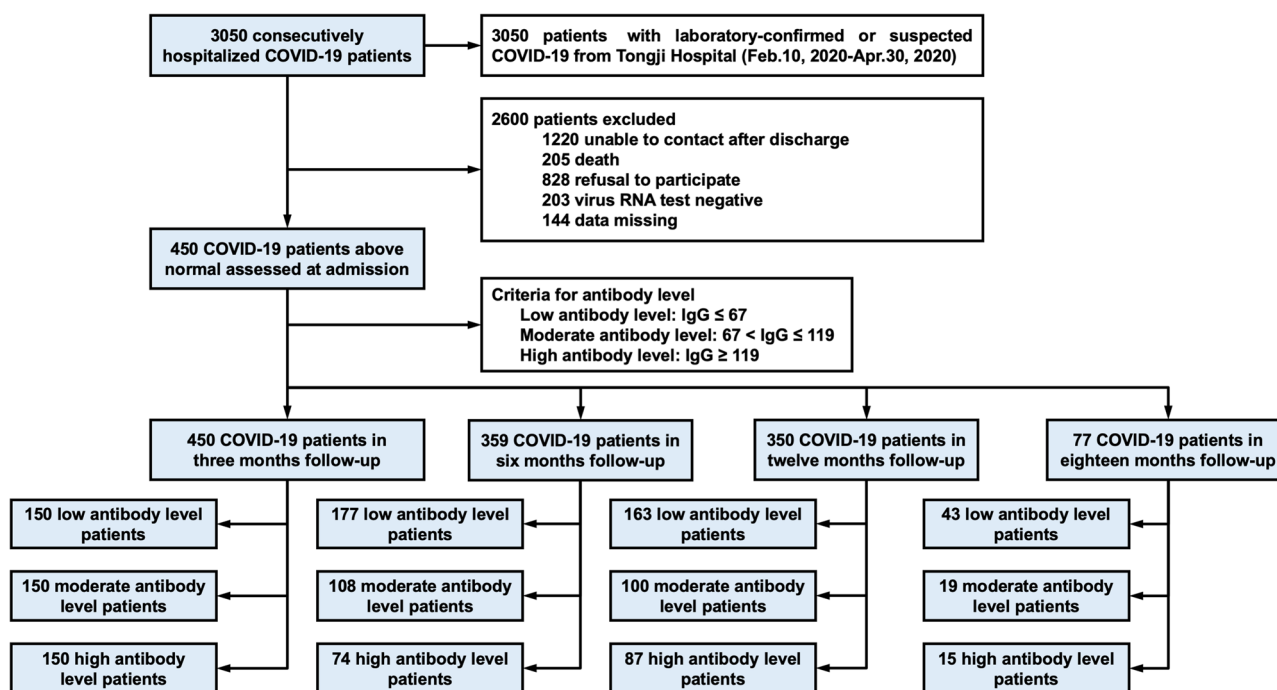


**FIGURE 1** Patient flowchart.

First, 20 μL test samples were added to 130 μL complete DMEM (GIBCO) medium in duplicate. Following eight serial dilutions in a 3-fold step-wise manner, pseudovirus at an optimized concentration as well as the cell and virus control were incubated together in a 96-well plate for 1 h in a 5% $CO_2$ environment at 37°C. Then, 100 μL $2 \times 10^5$ cells/mL freshly trypsinized Huh7 cells were added to each well. Following 24 h of incubation in a 5% $CO_2$ environment at 37°C, the culture supernatant was aspirated gently to leave 150 μL in each well. Then, 100 μL of luciferase substrate (Perkinelmer; 6066769) was added to each well. Following incubation at room temperature for 2 min, 150 μL of lysate was transferred to white solid 96-well plates (Biosharp; BS-MP-96W) for the detection of luminescence using a microplate lumi- nometer (PerkinElmer; Ensight). The $EC_{50}$ values were calculated with nonlinear regression, that is, log (inhibitor) versus response (four parameters), using GraphPad Prism 8 (GraphPad Software, Inc). The pseudovirus was purchased from National Institutes for Food and Drug Control. Huh7 cell line was kindly provided by Mr. Shuo Shen from Wuhan Institute of Biological Products Co. Ltd.

## 2.4 | The overall architecture of our LTAP model

Our LTAP model for COVID-19 long-term immunity prediction consists of two models, the first being the lung-lesion segmentation model, and the second being the antibody prediction model. The pretrained lung-lesion segmentation model takes a raw CT scan as the input and produces a lung-lesion map to calculate quantitative CT features. The antibody prediction model consists of the embedded layers, the multimodal attention module, and a multilayer perceptron (MLP) head (Figure 2).

In the antibody prediction model, the input data encompasses both static and longitudinal data. Static data refers to demographic characteristics, medical history, and patient status at admission, while longitudinal data includes multiple inspection reports and quantitative CT features during hospitalization, such as organ damage indicators, coagulation abnormalities, blood routine, blood biochemistry, inflammation markers, and so forth.

This static and longitudinal data are then fed into the corresponding embedding layers, yielding vectors of equal length. In this process, static data is mapped into a single vector known as the static embedding, while longitudinal data samples from individual time points are concurrently mapped into distinct vectors called report embeddings and CT embeddings. Notably, disparate temporal samples originating from an identical inspection report or quantitative CT features share a common embedding layer.

Next, the static embedding, report embeddings, and CT embeddings are input into the multimodal attention module for feature refinement and then passed through the transformer encoder along with the proposed interval embeddings. To perform prediction, we use the MLP head to output the probabilities of long-term IgG levels. The overall architecture of the LTAP model and its hyperparameters used in our experiments are shown in Figure 2 and Supporting Information: Table S1, respectively. We described each module of the LTAP model in detail (See Embedding layer, Feature-attention module, and Mutual-attention module of Supporting Information).
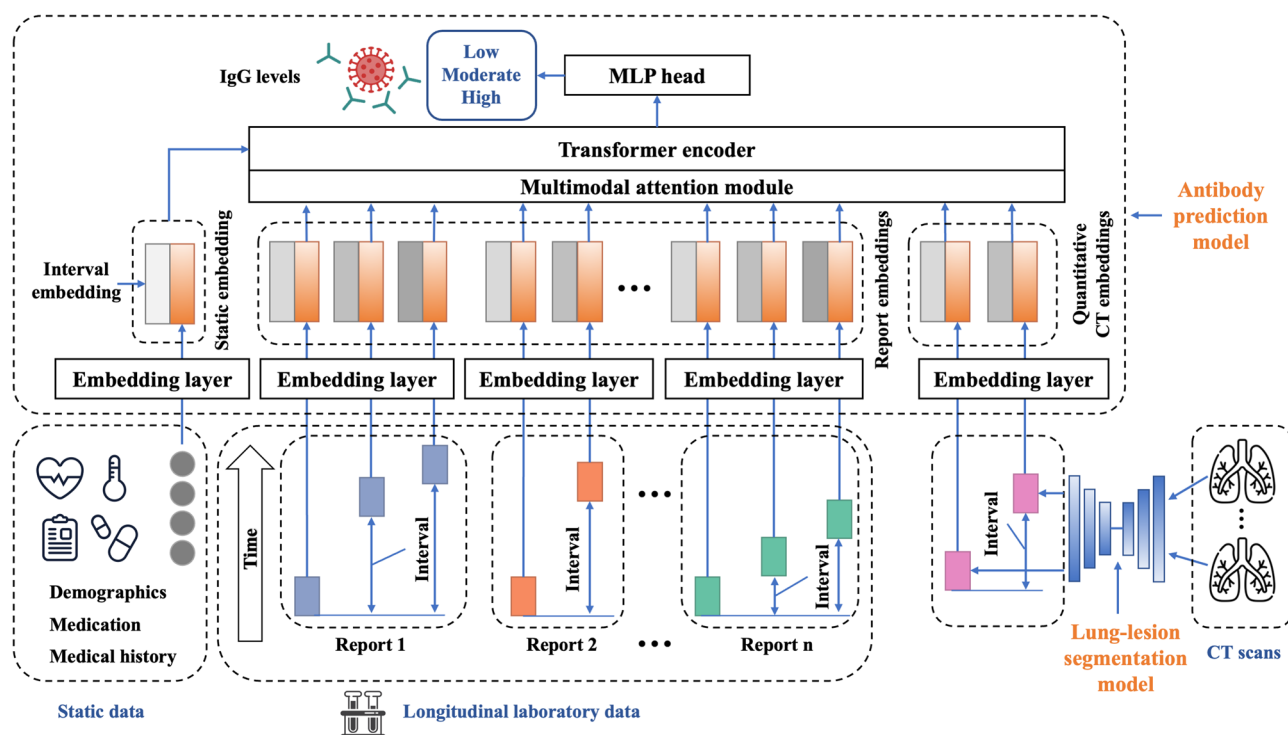


**FIGURE 2** Architecture of our LTAP model. MLP, multilayer perceptron.

## 2.5 | CT feature quantification

In the CT feature quantification stage, we first trained a popular medical image segmentation network, U-net,[28] on 507 CT scans from Tongji Hospital to produce a lung-lesion map for each slice of 1645 CT scans in the CALP (COVID-19 antibody level prediction) data set. Then, we downsampled each lung-lesion map from $512 \times 512$ to $128 \times 128$, cut off the slices with no infection at both ends of the lung area, resized the slice dimension of each truncated CT scan to 64, and got a lung-lesion cube of $128 \times 128 \times 64$ (length × width × slice). We calculated six different hand-crafted features in the left and right lung areas, separately, including median CT value, 90th percentile of CT value, mean CT value, max CT value, the volume of infected areas, and infected percentage of the lung area. Finally, we had a 768-dimension ($12 \times 64$) quantitative feature for each CT scan.

## 2.6 | Interval embedding

Since the time information is important to reveal long-term IgG levels of COVID-19 convalescent patients, interval embeddings are added to the static, report, and CT embeddings (Figure 2). Let $t$ be the time interval from the data measurement time to the prediction time in days, $\vec{p}_t^{(i)} \in \mathrm{R}^D$ be its corresponding interval embedding, and $D$ be the embedding dimension. Then $f : N \to \mathrm{R}^D$ will be the function that produces the output vector $\vec{p}_t$ and it is defined as follows:

$$\vec{p}_t^{(i)} = f(t)^{(i)} = \begin{cases} \sin(\omega_k t), & \text{if} \quad i = 2k \\ \cos(\omega_k t), & \text{if} \quad i = 2k + 1 \end{cases} \tag{1}$$

where $\omega_k = \frac{m}{10000^{2k/D}}$, $i$ is the dimension, $m$ is a parameter used to adjust the frequency of trigonometric functions varying with vector dimension. The overall interval embeddings (shown in Supporting Information: Figure S5) can be summarized as:

$$\begin{aligned} E_{\text{interval}} &= [\vec{p}_s; \vec{p}_1; \cdots; \vec{p}_N; \vec{p}_C], \vec{p}_s \in \mathrm{R}^D, \vec{p}_N \in \mathrm{R}^{L_N \times D}, \\ \vec{p}_C &\in \mathrm{R}^{L_C \times D}, E_{\text{interval}} \in \mathrm{R}^{(1 + L_1 + \cdots + L_N + L_C) \times D} \end{aligned} \tag{2}$$

where $\vec{p}_s$ is the static inverval embedding, $\vec{p}_N$ is the N-th report interval embedding, $\vec{p}_C$ is the CT interval embedding, $L_N$ is the length of the N-th report sequence, and $L_C$ is the length of the CT scan sequence. We also experimented with using learned positional embeddings and fixed sinusoidal version (Supporting Information: Figure S6)[29] instead.

## 2.7 | Multimodal attention module

The multimodal attention module incorporated the information from different modalities, whose network structure parameters are listed in Supporting Information: Table S1. It employed two sequential modules, namely the feature-attention module and the mutual-attention module. In the feature-attention module, we produced feature attention matrices by exploiting inter-feature relationships in the static embedding, report embeddings, and CT embeddings, respectively. Since each value in an embedding vector is regarded as a feature, the feature attention focuses on what is meaningful given an input embedding. To compute the feature attention efficiently, we squeezed the dimension of the input feature vector. On the other hand, the mutual-attention module was used to generate mutual-attention vectors by utilizing inter-modality relationships of the input embeddings. The mutual attention was operated on the temporal dimension to find the most informative time point, which was complementary to the feature attention. We described the detailed operations of the above submodules in Supporting Information (see Feature-attention module and Mutual-attention module of Supporting Information).

## 2.8 | Transformer-based classification network

The backbone of the antibody prediction model is a Transformer encoder,[29] whose detailed structure is shown in Supporting Information: Table S1. It consists of alternating layers of multiheaded self-attention with mask and MLP blocks. A residual connection is employed around each block, followed by layer normalization (LN).[30] The refined feature embeddings and the interval embeddings are fused by element-wise addition (Equation 3) and serve as input to the encoder. The classification head is implemented by an MLP with one hidden layer to the average encoder output along the time dimension.

$$z_0 = E'_{\text{feature}} \oplus E_{\text{interval}}, \tag{3}$$

where $E'_{\text{feature}}$ is the refined feature embedding of the multimodal attention module, $E_{\text{interval}}$ is the interval embedding, and $\oplus$ denotes element-wise addition.

## 2.9 | Data imputation

To deal with the problems of data sparsity and missing data, we proposed a data imputation net (DIN) as shown in Supporting Information: Figure S7. DIN is a deep neural network that takes the raw data matrix with missing data as input and outputs imputed data values of each missing data point.

We denote the input data matrix as $Y_{P \times F} = \{y_{11}, y_{12}, ..., y_{PF}\}$, where $P$ is the number of patients, $F$ is the number of features, and $y_{ij}$ is the value of the $i$-th patient with the $j$-th feature. In the input data matrix, we naturally take one-hot encodings of the patient ID and feature ID as the initial representation. Those patient and feature representations are highly sparse.

Right after the input layer is the embedding layer. It is a fully connected layer that projects the sparse representation to a dense vector. This procedure assigned fixed-size continuous vectors to each patient-ID and feature-ID as dense embeddings to alleviate the sparsity problem. The patient embeddings and feature embeddings are then fed into several interaction blocks to generate the final imputed data values.

The interaction block, following the embedding layer, is composed of a multilayer neural network and a matrix factorization component.[31] The multilayer neural network is composed of a sequence of fully connected layers, and the matrix factorization component is an element-wise product operation.

The final part of the model is the prediction layer followed by a loss function. The prediction layer is an MLP that projects the output of the previous interaction block to the predicted value, and the loss function minimizes the difference between the predicted value and the true value. We formulate the mean squared error (MSE) loss function as follows:

$$L_{mse}(y, \hat{y}) = \frac{1}{S} \sum_{(i,j) \in Y} (y_{ij} - \hat{y}_{ij})^2, \tag{4}$$

where $y_{ij}$ and $\hat{y}_{ij}$ are the true value and predicted value of the $i$-th patient with the $j$-th feature, respectively, $S$ is the number of total samples, and $Y$ is the input patient-feature matrix.

## 2.10 | Statistical analysis and implementation details

Continuous variables were expressed as mean and standard deviation (SD) and compared with the unpaired two-sided student's $t$ test. Categorical variables were expressed as number (%) and compared using the $\chi^2$ test or Fisher's exact test following the weight of frequencies between different groups.

The proposed LTAP model was implemented on an Intel(R) Xeon(R) CPU E5-2637 with an NVIDIA TITAN V GPU workstation based on the PyTorch[32] framework. The embedding dimension of the DIN model was set to 128 in the embedding layer. In the interaction block of the DIN model, the MLP has three hidden layers with 256, 128, and 64 units in each layer, respectively, and between every two hidden layers, there is a dropout layer[33] with $p = 0.5$. To optimize the DIN model, we used the Adam[34] optimizer with $\alpha = 0.0045, \beta_1 = 0.25$, and $\beta_2 = 0.5$ as the learning rate, and used MSE as the loss function. The LTAP model was trained by an SGD optimizer with a fixed learning rate of 0.005, and a batch size of 256 on a single NVIDIA TITAN V GPU. To optimize the LTAP model, we chose the cross entropy loss and $K = 5$ in the K-fold cross-validation[35] data set splitting. The data set was divided into five equal folds. The model was then trained on four of these folds and tested on the remaining fold. This process was repeated five times, with each fold serving as the test set exactly once. The final model performance was assessed by averaging the results obtained from each of the five folds.

## 3 | RESULTS

## 3.1 | Demographic and clinical characteristics of COVID-19 patients grouped by IgG level

We included 450 in-patients from Tongji Hospital between February 2020 and April 2020 with moderate to critical COVID-19 infection in our prospective longitudinal study. We analyzed the anti-IgG titers in serum samples of these patients following their discharge at 3 months, 6 months, 12 months, and 18 months, labeled as V1 through V4, respectively. Due to different measurement methods, there is no uniform threshold to distinguish the IgG level. Since deep-learning models can benefit from a balanced data set to achieve better robustness and effectiveness, we equally divided participants into three groups based on their IgG levels in V1: high (IgG > 119, $n = 150$), moderate ($67 < IgG < 119$, $n = 150$), and low (IgG ≤ 67, $n = 150$) IgG levels. We described the demographic and clinical characteristics of these patients in Table 1.

Overall, a significantly higher proportion of critical patients were found in the high IgG group compared with moderate and low IgG groups (18% vs. 13% and 7%; $p = 0.002$). Consistent with disease severity, patients in the high IgG group showed older age, higher ICU admission frequency, and lower SpO2 levels than those in the moderate and low groups. Meanwhile, higher WBC and neutrophil counts, as well as elevated levels of lactate dehydrogenase (LDH) and C-reactive protein (CRP) levels, were found in the high IgG group. Compared with moderate and low IgG groups, elevated pro-inflammatory cytokines, including IL-6 and IL-2R, were observed in the high IgG group. As shown in Table 1, the vast majority of laboratory parameters, sex composition, and underlying medical comorbidities showed no significant difference in these three groups.

Next, we measured Nab levels at V1-V3 in a representative group of 45 patients. Due to different measurement methods, there is no uniform quantitative threshold to distinguish the effective level of Nab. As previously reported,[36] we redeem the mean of convalescent level as threshold and then divided these participants into two groups based on mean Nab titers of V1 and V2: low (Nab titer (log10) ≤ 2.199, $n = 24$), high (Nab titer (log10) > 2.199, $n = 21$). As shown in Supporting Information: Table S2, age, the likelihood of ICU admission, SpO2, and SBP did not show significance when selected patients were grouped by Nab titer level. That is mainly due to the limited number of patients who underwent Nab testing. Importantly, critical patients were present more frequently in the high Nab group (61.9% vs. 20.8%; $p = 0.008$). Also, the levels of CRP, LDH, and IL-2R were significantly higher in the high Nab group compared with the low Nab group (87.27 vs. 40.82, 404.90 vs. 277.42, and 836.24 vs. 431.73, respectively), which is consistent with our observation in grouping with IgG level.

## 3.2 | The temporal change of IgG and neutralizing activities during 18-month follow-up

We first analyzed the temporal changes of IgG levels to reflect long-term immune responses following SARS-CoV-2 infection in our large-scaled prospective cohort. As shown in Table 2 and Figure 3, IgG levels peaked at 3 months and remained stable in 12 months postdischarge, followed by a significant decline in 18 months postdischarge. We next asked whether IgG levels were impacted by various clinical and demographic factors, including disease severity, age, and sex. During the follow-up, we observed higher

**TABLE 1** Demographic and clinical characteristics of COVID-19 patients, grouped by IgG level.

| | All patients (n = 450) | IgG low (n = 150) | IgG moderate (n = 150) | IgG high (n = 150) | p Value |
|---|---|---|---|---|---|
| *Patient characteristics* | | | | | |
| Age, years | 60.0 (±0.6) | 58.0 (±1.1) | 60.0 (±1.0) | 61.8 (±1.0) | 0.027 |
| Female/male | 239/211 | 79/71 | 79/71 | 81/69 | 0.965 |
| *Baseline symptoms* | | | | | |
| Admission severity | / | / | / | / | 0.002 |
| Moderate | 32 (7.1%) | 15 (10.0%) | 12 (8.0%) | 5 (3.3%) | / |
| Severe | 362 (80.4%) | 125 (83.3%) | 119 (79.3%) | 118 (78.6%) | / |
| Critical | 56 (12.4%) | 10 (6.7%) | 19 (12.7%) | 27 (18.0%) | / |
| Nucleic acid positivity, days | 28.2 (±0.5) | 26.5 (±0.9) | 29.1 (±1.0) | 28.9 (±1.0) | 0.103 |
| To ICU | 6/405 (1.48%) | 0/137 (0%) | 1/134 (0.75%) | 5/134 (3.73%) | 0.027 |
| SpO2, % | 95.2 (±0.3) | 96.1 (±0.4) | 94.9 (±0.4) | 94.5 (±0.5) | 0.027 |
| HR, bpm | 89.8 (±0.8) | 89.1 (±1.3) | 92.1 (±1.3) | 88.4 (±1.3) | 0.103 |
| SBP, mmHg | 131.4 (±0.9) | 135.1 (±1.5) | 128.2 (±1.4) | 130.9 (±1.5) | 0.004 |
| DBP, mmHg | 80.4 (±0.5) | 81.7 (±1.0) | 80 (±0.9) | 79.6 (±1.0) | 0.243 |
| Peak body temperature, °C | 38.36 (±0.05) | 38.30 (±0.08) | 38.36 (±0.09) | 38.40 (±0.08) | 0.719 |
| *Medical history* | | | | | |
| Liver disease | 43/448 (10%) | 15/150 (10%) | 12/149 (8%) | 16/149 (11%) | 0.72 |
| Kidney disease | 15/449 (3%) | 8/150 (5%) | 3/150 (2%) | 4/149 (3%) | 0.238 |
| Endocrine disease | 100/449 (3%) | 26/150 (17%) | 37/150 (25%) | 37/149 (25%) | 0.206 |
| Cardiovascular disease | 192/449 (43%) | 63/150 (42%) | 61/150 (41%) | 68/149 (46%) | 0.669 |
| Coronary heart disease | 35/450 (8%) | 15/150 (10%) | 9/150 (6%) | 11/150 (7%) | 0.422 |
| Hypertension | 193/450 (43%) | 67/150 (45%) | 58/150 (39%) | 68/150 (45%) | 0.44 |
| Diabetes | 95/433 (22%) | 26/145 (18%) | 35/145 (24%) | 34/143 (24%) | 0.36 |
| Respiratory disease | 54/450 (12%) | 15/150 (10%) | 22/150 (15%) | 17/150 (11%) | 0.442 |
| Hepatitis B | 16/432 (4%) | 6/147 (4%) | 5/142 (4%) | 5/143 (3%) | 0.957 |
| History of smoking | 78/450 (17%) | 21/150 (14%) | 29/150 (19%) | 28/150 (19%) | 0.415 |
| *Baseline investigations* | | | | | |
| White blood count, $10^9$/L | 6.04 (±0.11) | 5.69 (±0.16) | 5.96 (±0.20) | 6.48 (±0.22) | 0.015 |
| Neutrophil count, $10^9$/L | 4.27 (±0.11) | 3.85 (±0.15) | 4.23 (±0.20) | 4.74 (±0.21) | 0.004 |
| Lymphocyte count, $10^9$/L | 1.19 (±0.03) | 1.28 (±0.05) | 1.17 (±0.05) | 1.11 (±0.04) | 0.027 |
| C-reactive protein, mg/L | 42.0 (±2.5) | 29.0 (±3.4) | 47.2 (±4.9) | 49.4 (±4.5) | 0.002 |
| Hemoglobin Count, g/L | 126.2 (±0.7) | 126.1 (±1.3) | 125.9 (±1.1) | 126.6 (±1.2) | 0.923 |
| Platelet count, $10^9$/L | 240 (±4) | 238 (±7) | 233 (±8) | 249 (±8) | 0.34 |
| ALT, U/L | 33.8 (±1.6) | 31.1 (±2.9) | 32.9 (±2.5) | 37.5 (±2.9) | 0.235 |
| AST, U/L | 33.0 (±1.3) | 30.3 (±2.6) | 33.6 (±2.2) | 35.2 (±1.8) | 0.282 |
| Lactate dehydrogenase, U/L | 297 (±6) | 265 (±10) | 303 (±9) | 322 (±10) | <0.001 |
| BUN, mmol/L | 4.87 (±0.12) | 4.65 (±0.21) | 4.89 (±0.19) | 5.08 (±0.23) | 0.359 |
| Creatinine, umol/L | 73.3 (±1.7) | 73.2 (±2.6) | 71.1 (±1.7) | 75.4 (±4.2) | 0.602 |
| IL-1B, pg/mL | 6.22 (±0.36) | 6.37 (±0.64) | 5.38 (±0.17) | 6.90 (±0.84) | 0.215 |

(Continues)

**TABLE 1** (Continued)

| | All patients (n = 450) | IgG low (n = 150) | IgG moderate (n = 150) | IgG high (n = 150) | p Value |
|---|---|---|---|---|---|
| IL-2R, u/mL | 653 (±23) | 594 (±35) | 641 (±31) | 715 (±48) | 0.092 |
| IL-6, pg/mL | 19.6 (±2.3) | 12.4 (±2.1) | 16.1 (±2.5) | 29.4 (±5.8) | 0.006 |
| IL-8, pg/mL | 19.1 (±1.6) | 15.4 (±1.4) | 17.9 (±2.2) | 23.5 (±3.9) | 0.119 |
| IL-10, pg/mL | 6.52 (±0.22) | 6.39 (±0.36) | 6.52 (±0.43) | 6.63 (±0.33) | 0.900 |
| TNF-a, pg/mL | 10.81 (±1.84) | 8.28 (±0.33) | 13.54 (±5.36) | 10.39 (±0.79) | 0.512 |

*Note*: Data are expressed as mean (±standard error of mean) or number (percent).

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BUN, blood urea nitrogen.

**TABLE 2** IgGs of COVID-19 patients grouped by admission severity, age, and sex.

| | V1 IgG mean (±SEM) | V2 IgG mean (±SEM) | V3 IgG mean (±SEM) | V4 IgG mean (±SEM) |
|---|---|---|---|---|
| Total | 99.85 (±2.98) | 80.45 (±2.98) | 90.00 (±3.68) | 77.78 (±6.59) |
| Admission severity | | | | |
| Moderate | 75.34 (±7.84) | 67.33 (±11.07) | 73.30 (±9.38) | 39.51 (±35.22) |
| Severe | 97.86 (±3.30) | 78.75 (±3.32) | 88.77 (±4.18) | 79.58 (±7.59) |
| Critical | 126.74 (±9.04) | 96.31 (±8.18) | 105.52 (±10.26) | 75.91 (±14.58) |
| Age | | | | |
| <45 | 82.65 (±8.15) | 70.02 (±8.03) | 83.17 (±11.38) | 78.54 (±17.02) |
| 45–65 | 97.87 (±4.26) | 79.68 (±4.12) | 89.43 (±5.38) | 59.02 (±7.61) |
| >65 | 107.28 (±4.74) | 84.51 (±4.99) | 92.40 (±5.64) | 88.56 (±9.75) |
| Sex | | | | |
| Male | 100.80 (±4.48) | 82.89 (±4.45) | 86.98 (±5.46) | 77.67 (±9.22) |
| Female | 99.02 (±3.99) | 78.32 (±4.02) | 92.72 (±4.98) | 77.87 (±9.37) |

Abbreviations: SEM, standard error of mean; V1, 3-month follow-up; V2, 6-month follow-up; V3, 12-month follow-up; V4, 18-month follow-up.

IgG levels in critical patients compared with moderate and severe cases (Figure 3A,D). Meanwhile, patients aged >65 years had higher IgG levels over time (Figure 3B). However, there was no difference in IgG levels between male and female patients (Figure 3C). Another critical issue of concern is the persistence and decay of antibodies. Here, by analyzing the trends of IgG levels with discharged days in this prospective cohort study, we identified that the decay rate of critical patients was more than six times that of moderate patients (decay rate: −0.079 vs. −0.013) and more than two times that of severe patients (decay rate: −0.079 vs. −0.039) (Figure 3E), despite the higher baseline IgG levels.

Next, we measured the neutralizing activities of 45 representative patients in V1–V3. As shown in Figure 4A, the neutralizing activities declined through V1–V3. Also, a higher neutralizing titer was observed in critical patients compared to moderate and severe patients (Figure 4B). Importantly, we illustrated a stable correlation between IgG and neutralizing activities (Figure 4C–E), implying the possibility of predicting antibody-neutralizing activities using IgG levels (Figure 4F).

### 3.3 | Performance of our LTAP model in predicting long-term COVID-19 IgG levels

To distinguish convalescent COVID-19 patients with different IgG antibody levels, we employed a novel diagnostic system assembled from a pretrained lung lesion segmentation model and a transformer-based diagnosis analysis model using deep learning. The diagnosis classification took the quantitative CT features generated by segmentation networks and longitudinal clinical data as input. Our LTAP model was able to differentiate low IgG antibody level convalescents from the other two classes (moderate level and high level) with 0.827 accuracy, 0.826 precision, 0.827 recall, 0.826 F1-score, and AUC of 0.905 (95% CI: 0.884–0.925) on 3 months postdischarge IgG level prediction. The LTAP model demonstrated an accuracy of 0.696 and an AUC score of 0.793 (95% CI: 0.760–0.824) on 6 months IgG level prediction. At 12 and 18 months postdischarge, the LTAP achieved AUC scores of 0.737 (95% CI: 0.702–0.772) and 0.730 (95% CI: 0.651–0.805) for IgG level prediction, respectively
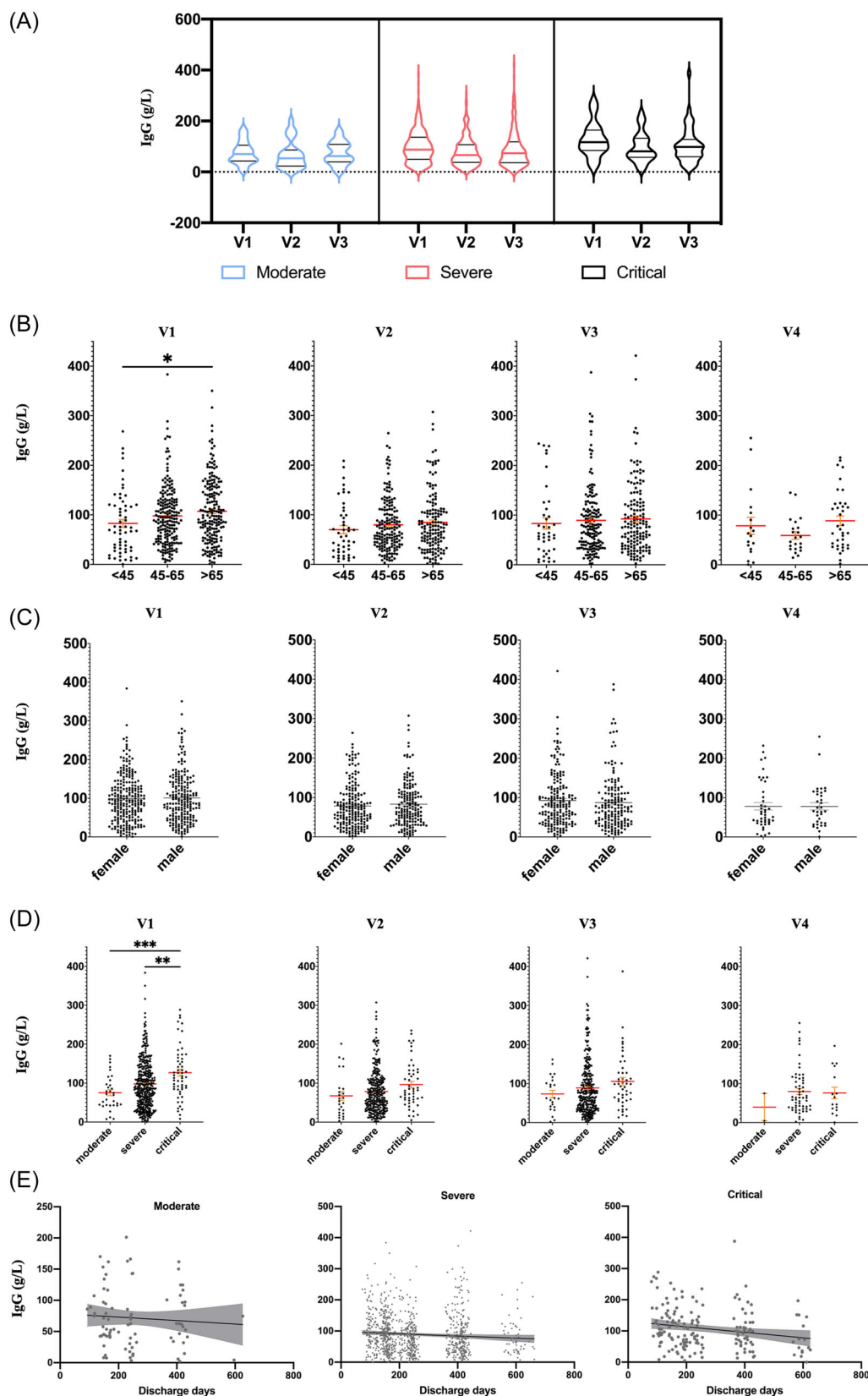
**FIGURE 3** The temporal change and decay rate of IgG levels in COVID-19 patients. (A) The temporal change in IgG levels of COVID-19 patients grouped by disease severity. (B–D) IgG levels of COVID-19 patients grouped by age, sex, and disease severity in V1-V4. (E) Longitudinal dynamics of IgGs for moderate, severe, and critical patients. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.
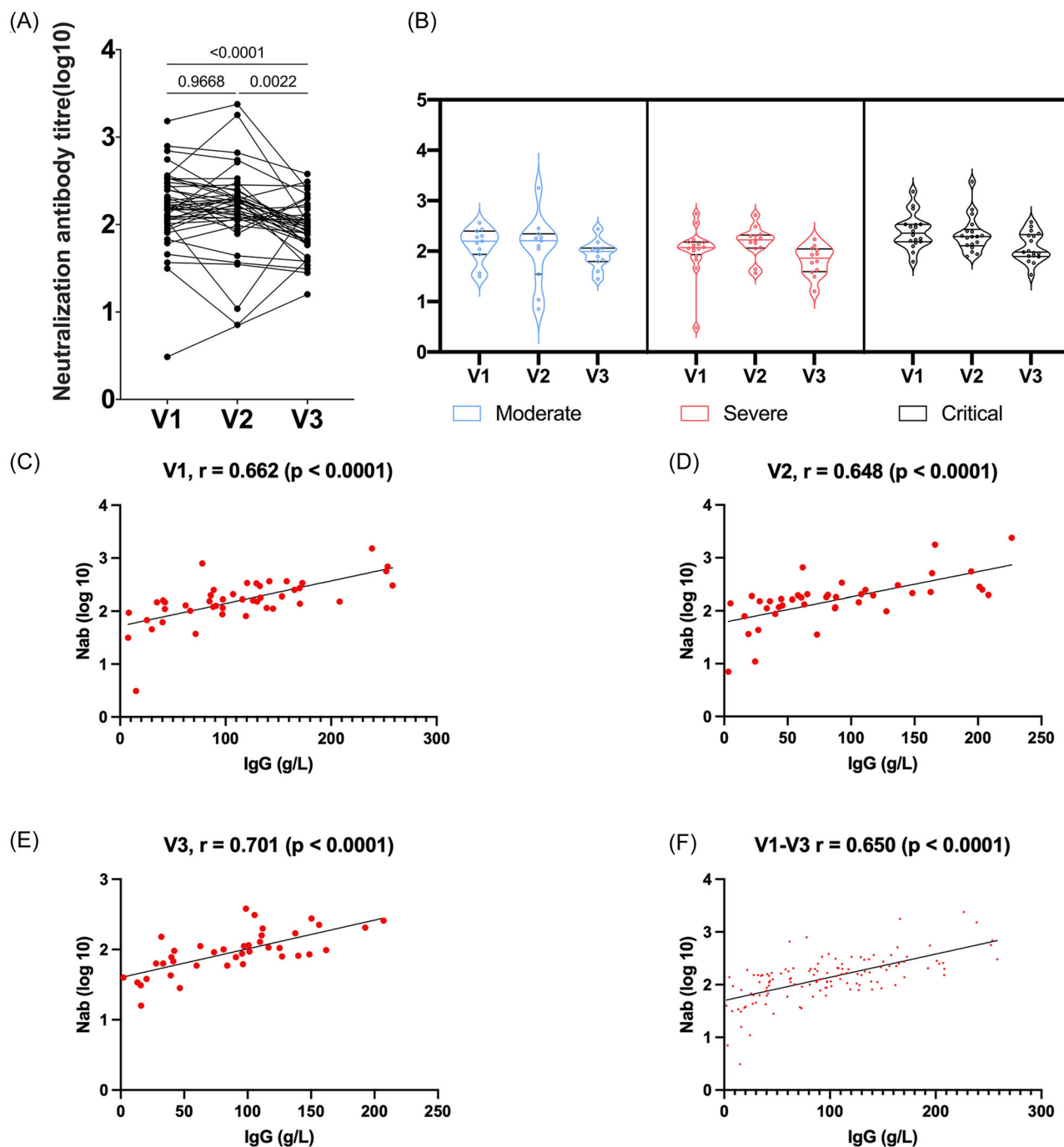
**FIGURE 4** The temporal change of Nab titers and their association with IgG. (A) The temporal change of Nab titers across V1-V3. (B) Nab titers of COVID-19 patients grouped by disease severity in V1-V3. (C–F) Polynomial curves fitted to Nab in V1-V3.

(Table 3). As shown in Supporting Information: Figure S8, AUCs of the high and low IgG levels are consistently higher than that of the moderate IgG levels at each follow-up time point (V1: 0.928/ 0.937 vs. 0.836, V2: 0.781/0.818 vs. 0.682, V3: 0.698/0.754 vs. 0.639, V4: 0.732/0.640 vs. 0.593). The possible reason is that compared to patients with particularly high or low IgG levels, patients with moderate IgG levels generally share characteristics that are not as discriminative, which can lead to lower performance in predicting moderate IgG levels.

## 3.4 | Ablation study on embeddings, CT feature, and attention module

To thoroughly evaluate the effectiveness of individual components of our LTAP model, we performed extensive ablation study experiments. In this subsection, we use the raw 3 months postdischarge data of the IgG data set without CT features and transformer architecture as the baseline. Compared to the baseline with position learnable embedding, adding quantitative CT features obtains a

**TABLE 3** Performances of our LTAP model from 3-, 6-, 12-, and 18-month follow-ups.

| | 3 months | 6 months | 12 months | 18 months |
|---|---|---|---|---|
| AUC | **0.905 [0.884–0.925]** | 0.793 [0.760–0.824] | 0.737 [0.702–0.772] | 0.730 (0.651–0.805) |
| Accuracy | **0.827 [0.798–0.856]** | 0.696 [0.657–0.738] | 0.626 [0.583–0.669] | 0.649 (0.558–0.740) |
| Precision | **0.826 [0.797–0.857]** | 0.689 [0.647–0.734] | 0.618 [0.576–0.665] | 0.629 (0.530–0.739) |
| Recall | **0.827 [0.798–0.856]** | 0.696 [0.657–0.738] | 0.626 [0.583–0.669] | 0.649 (0.558–0.740) |
| F1-score | **0.826 [0.797–0.856]** | 0.690 [0.646–0.732] | 0.613 [0.567–0.658] | 0.622 (0.520–0.714) |

*Note*: 95% confidence intervals are included in brackets. The best average results are shown in bold.

relative improvement of 1.1% in AUC (0.599 vs. 0.592). With data imputation, the improvement brought by CT features is more obvious, boosting the AUC to 0.895 with an increment of 4.6% (0.895 vs. 0.856). Then we introduce interval embedding to replace position embedding for sampling time information coding, which contributes to the 5.7% (0.633 vs. 0.599) performance gains of relative with the raw data set and 0.9% (0.903 vs. 0.895) relative with the imputed data set in AUC. In addition, when we included the attention module, we obtained 2.7% (0.650 vs. 0.633) and 0.2% (0.905 vs. 0.903) uplifts in AUC for baseline and imputed data, respectively. The performance improvements from the attention module in terms of accuracy, precision, recall, and F1-score are more distinctive. Moreover, by incorporating the imputed data into the raw data set, our framework consistently improves the performance over baseline in terms of AUC, accuracy, precision, recall, and F1-score. As listed in Supporting Information: Table S3, the proposed data imputation method can significantly boost all model settings, including position learnable embedding (POS), POS + CT, Interval Embedding (IE) + CT, IE + Attention module (Att) + CT, yielding increments of 44.6% (0.856 vs. 0.592), 49.4% (0.895 vs. 0.599), 42.7% (0.903 vs. 0.633), 39.2% (0.905 vs. 0.650) in AUC.

## 3.5 | Prognostic factors of our LTAP model for the IgG level prediction

To interpret the impact and relative importance of various clinical features on the long-term IgG levels of COVID-19 convalescents, we utilized the SHAP[37] and GradCAM[38] explainers. Martinez et al.[26] employed SHAP to identify that PDL1 and MPO were the primary biomarkers contributing to the prediction of severity levels in sepsis and septic shock patients within the ANN derived from Boruta and Lasso Regression, respectively. Similar to the above research, we trained the LTAP model to predict low-level IgG convalescence from the 3-month follow-up data set, reaching an AUC of 0.948, with an accuracy of 0.927, a precision of 0.927, a recall of 0.927, and an F1 score of 0.927. With this model, we computed the top 20 prognostic factors. As shown in Figure 5, coagulation disorders, immune response, and immune status were closely related to IgG production in both interpretability analysis methods. Among them, fibrinogen degradation products (FDP) and fibrinogen showed the highest

weight, suggesting that dysregulated coagulation, such as hypercoagulable state, was associated closely with antibody production. In addition, the percentages of T cells, B cells and natural killer (NK) cells, older adults, males, underlying diseases (hypertension, endocrine diseases), and immune-suppression state (glucocorticoid and immunoglobulin therapy) also served as important indicators in this predictive model, respectively.

## 3.6 | Comparison of our LTAP model against other methods for the SARS-CoV-2 Nab and IgG levels prediction

To validate the effectiveness of our method, we compared it with other typical prediction models on the Nab titer and IgG datasets. Due to the relatively small number of Nab detection, we applied top 15 salient factors identified by the IgG level prediction model to the Nab level prediction and achieved an AUC of 0.828. This lightweight model makes full use of the general knowledge learned from the IgG level prediction model, enhances the practicability of the LTAP model, and reduces the burden of data collection. We compared the LTAP model against MLP and long short-term memory (LSTM) on 3-month to 12-month follow-ups of the Nab titer data set with and without quantitative CT features. As shown in Supporting Information: Table S4, the LTAP model is superior to MLP and LSTM using top 15 salient factors and quantitative CT features, yielding improvements of 8.9% (0.828 vs. 0.76) and 6.6% (0.828 vs. 0.777) in AUC. Using only top 15 salient factors, improvements are more distinctive, reaching 12.6% (0.807 vs. 0.717) and 12.2% (0.807 vs. 0.719). When replacing all clinical data with top 15 salient factors, the performance of our LTAP model maintained the previous level (0.828 vs. 0.824). Furthermore, using only top 15 salient factors, our LTAP model achieved an AUC of 0.807, which is practical in clinical diagnosis.

As shown in Supporting Information: Table S5, on the 3 months postdischarge data set of the IgG data set, the proposed transformer-based method is also superior to other machine learning- based methods, including logistic regression (LR), MLP, LSTM, yielding improvements of 11.0% (0.905 vs. 0.815), 7.4% (0.905 vs. 0.843), 4.7% (0.905 vs. 0.864) in AUC. Meanwhile, data imputation is also effective for other methods, bringing performance improvements of
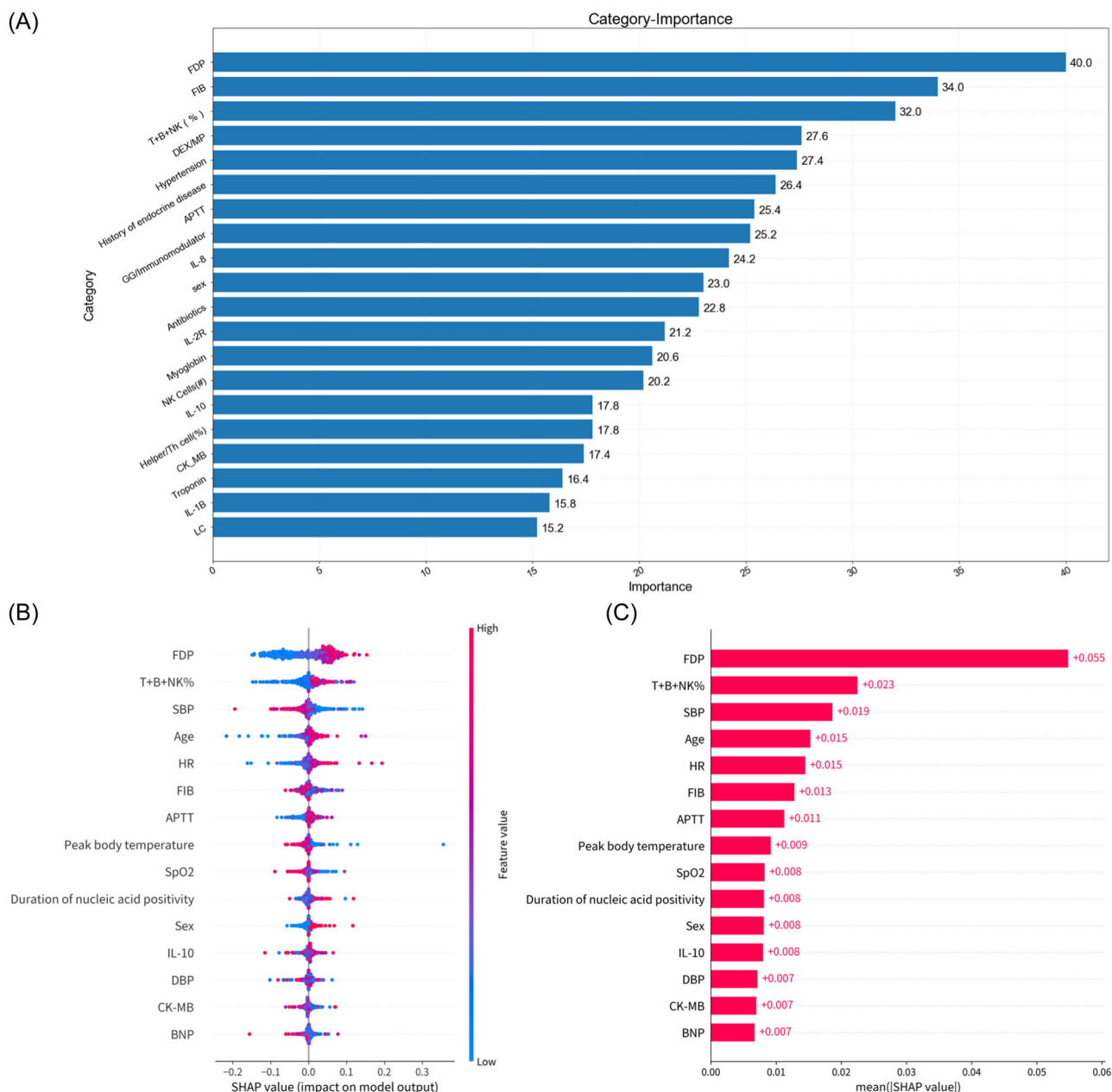
(A)



(B)



(C)



**FIGURE 5** Salient factors and clinical prognostic analysis for the long-term IgG level prediction. (A) The top 20 features for prediction of the long-term IgG level. (B) The relative contributions of clinical features for long-term IgG level prediction. Features on the right of the explanation bar pushed the high IgG level probability higher, and features on the left pushed the low IgG level probability higher. (C) The mean relative contribution of each clinical feature to predict long-term IgG level. APTT, activated partial thromboplastin time; BNP, B-natriuretic peptide; CK-MB, creatine kinase MB isoenzyme; DBP, diastolic blood pressure; FDP, fibrinogen degradation products; FIB, fibrinogen; HR, heart rate; IL-10, interleukin 10; SBP, systolic blood pressure.

55.8% (0.815 vs. 0.523), 37.7% (0.843 vs. 0.612), 39.6% (0.864 vs. 0.619) in AUC for LR, MLP, LSTM, respectively.

## 4 | DISCUSSION

Humoral immunity, especially the production of the Nab, is essential for providing a protective effect. Therefore, long-term immune responses are crucial for determining the risk of reinfection in previously infected individuals and the vaccination strategies. However, due to the limited clinical data and public health resources in the early stage of public health emergencies, large-scale detection of long-term immune responses in patients' postinitial infection cannot be achieved. Existing medical-aided diagnostic systems mainly focus on the recognition, diagnosis, and treatment of patients with COVID-19. However, few works explore the prognosis of convalescent patients. Therefore, the development of machine learning algorithms is required to make it possible to predict long-term

immunity based on the results of smaller clinical cohorts. With these algorithms, it is possible to promptly predict patients' long-term immunity according to incomplete hospitalization data and provide appropriate vaccination strategies individually, saving a lot of public health resources.

Recent reports have indicated that low-strength evidence points to a connection between older age and increased disease severity with humoral immunity.[9] However, the absence of a large-scale longitudinal study has limited our understanding. In this study, by analyzing the levels of IgG and Nab titers, we found that the IgG level declined from 3 to 6 months postdischarge, and remained stable from 6 to 12 months postdischarge, and declined by 18 months postdischarge. This trend was partially in accordance with the previous reporting that neutralizing activity remained stable from 6 to 12 months.[10] Another study reported that the levels of IgG declined faster in patients with a high level of peak titer,[39] which might be associated with the basic capacity of B cells in severe patients. Further research is needed to understand the underlying mechanisms.

We also identified a stable and strong correlation between the levels of IgG and Nab that could have implications for predicting serum-neutralizing activity using IgG levels. Additionally, we found that disease severity and older age were associated with increased levels of IgG. This observation suggested that more immune cells and enhanced immune memory were formed in critical patients following the battle against COVID-19. Cervia et al.[40] also reported that asymptomatic infection only involved mucosal antibody responses while severe infection caused high serum IgA and IgG production following SARS-CoV-2 infection. Moreover, an elevated number of S-RBD-specific and S-RBD-nonspecific activated memory B cells was observed at about 2 months following infection,[41] suggesting enhanced immune memory in critical patients. Taken together, these data suggested that patients with more severe infections and diseases would have established more broadly effective immune memory.

The correlation of the laboratory parameters identified by our system and the clinical and biochemical evidence of the immunodynamics highlighted the immune mechanism interlink involving different antibodies level of COVID-19. With this model, we identified coagulation disorders, immune response, and immune status that were closely related to IgG production. FDP served as the most important indicator in this model, suggesting the close relationship between coagulation disorders and IgG production. This observation could be explained by the increased coagulation disorders in patients with higher disease severity,[42] although the mechanism of how FDP in regulates B cell function has not been explored yet. At the same time, the percentages of T cells, B cells, and NK cells account for a relatively high weight in distinguishing antibody levels, implying their contribution to antibody production. Specific subtypes of CD4+ T cells, such as Th and Tfh cells, could stimulate B cell differentiation and antibody production by secreting cytokines and chemokines.[43] On the other hand, antibodies could mediate antibody-dependent cytotoxicity (ADCC) through NK cells,

exacerbating disease severity.[44] Older adults, males, underlying diseases (hypertension, endocrine diseases), and immune-suppression states (glucocorticoid and immunoglobulin therapy) served as high-risk factors, which may be caused by the increased risk of coagulation disorders[4,45] and dysregulated immune response[46,47] in patients under such immune status. Overall, the highest weighting of these metrics in this newly established novel LTAP model supported the observation of elevated IgG levels in patients with higher disease severity.

Despite the observation of Nab levels, we provide evidence that it is feasible to continuously predict long-term Nab and IgG levels from 3 to 18 months postdischarge using longitudinal hospitalization data. These nonequidistant sampling data with clinical and quantitative CT features contained the inevitable missing values, which is common in the real-world medical system,[48] especially during the chaotic period at the beginning of the outbreak. Also, we found that when introducing interval embedding to our system, a singular model can capture the continuous longitudinal relationships, enabling the predictions of Nab levels at arbitrary postdischarge timings. To our knowledge, few studies predicted the prospective long-term Nab levels using longitudinal multimodal data. Our framework brings various performance advantages that can be realized by new medical-aided diagnostic systems with possibly highly nonintuitive designs enabled by deep learning, contrasting with traditional machine learning and statistical methods through intuition-driven design choices and analytical modeling.

Missing data is a thorny problem that we cannot avoid in medical-assisted diagnosis.[49] Improper handling of missing data can lead to undesired effects such as introducing data bias and noise, and reducing efficiency.[50] A common practice for handling the missing data is discarding any sample that has a missing value or filling the missing values with a default number. However, this procedure simplifies the data analysis process at the cost of amplifying the existing bias in the data set. In this work, we introduce a novel deep learning-based data imputation framework to address the sparse or missing data problem, inspired by collaborative filtering.[51,52] Collaborative filtering is a technique commonly used in recommender systems[53] to predict the preference of items based on a user's previous behaviors or the interests of similar users. Given a set of patients and their corresponding features, we can also model them in a collaborative filtering[31] manner. The patients and features in the clinical data can be seen as the users and items, respectively. Our data imputation framework can learn the latent relationships between patients with similar feature patterns to impute the missing values. Moreover, our missing data imputation framework is a general method that can be applied to many other scenarios. Such an AI system can significantly reduce the uncertainty of individual immunity development, help the medical system establish a regular epidemic prevention and control mechanism, and improve the response speed of the government.

In terms of our model's extendability, our work establishes a unified framework to handle different medically aided diagnostic tasks on longitudinal data with an average missing data rate of 8.5%. The basic principles of the deep learning-enabled intelligent

diagnostic system discussed above can also be applied to dynamically and continuously improve diagnostic performance. For example, classification networks could autonomously decide which features to use and which time intervals to emphasize, as longitudinal data naturally has different significance on features over time for different model objectives. This adaptability could be governed by a supervised learning framework with a concrete task goal or an unsupervised framework for general discovery and anomaly detection. The identified prognostic factors in turn inspire pathogenesis research, so as to involve more appropriate features and representations.

Our work still has some limitations. First, not all participants in the study were subjected to Nab detection, considering the limited economic cost and efficiency. Second, the number of participants decreased in the later follow-up period due to loss of follow-up and other issues (vaccination, chronic disease, etc.). Consequently, only 77 patients were recorded at 18 months postdischarge. Also, we only included single-center data. It would be more accurate and broadly applicable if multicenter data were available for our model. Our long-term antibody prediction model, if connected in a widely distributed, cost-effective, and privacy-protected manner, as a part of a federated learning system, will have the advantage of collaboratively learning online from evidence-based medical outcomes.

# 5 | CONCLUSIONS

In conclusion, utilizing data from a large prospective follow-up cohort of up to 18 months, we not only described the temporal change between IgG and Nab levels but also proposed a deep learning-based model to continuously predict long-term Nab and IgG levels with real-world clinical and laboratory data during hospitalization, providing diagnostic information for health agencies to personalize vaccine allocation. Our model trained on nonequidistant sampling data with a high missing rate achieves state-of-the-art results compared with other advanced models. Furthermore, we have identified prognostic factors to inspire pathogenesis study and iteratively optimize network architecture and feature selection strategy. As a general framework, this model could also be used in various medical fields such as environmental sensing, biomedical diagnostics, global health, and security/defense, and promote the development of related fields.

## AUTHOR CONTRIBUTIONS

Qin Ning, Xiaojing Wang, and Xiang Bai designed the project. Weiming Yan and Tao Chen provided key interactive information and discussion. Cong Fang and Xiaojing Wang helped in designing experiments. Tingting Liu collected clinical data. Yuying Chen performed the experiments with the assistance of Zhize Yuan and Yuxin Niu. Cong Fang and Zhiyong Dou trained and developed the deep learning model with the assistance of Xiaojing Wang and Yajie Chen. Peng Wang and Xiaoping Luo participated in the data discussion and provided suggestions. Wenzhen Zhu and Weiwei Chen provided essential materials. Cong Fang, Xiaojing Wang, and Zhiyong Dou wrote the manuscript with the assistance of Xitang Li, Wenhui Wu, and Fengning Luo. Qin Ning, Xiang Bai, Tao Chen, and Weiming Yan helped revise the manuscript.

## ACKNOWLEDGMENTS

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

Data supporting the present study are available from the corresponding author upon reasonable request. Python code for neural net construction, training, and prediction will be found at the public GitHub repository.

## ETHICS STATEMENT

The research protocol and written informed consent were approved by the Medical Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (2020S242).

## ORCID

*Qin Ning* http://orcid.org/0000-0002-2027-9593

## REFERENCES

1. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-733.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, China. *Lancet*. 2020; 395(10223):497-506.
3. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273.
4. Chen T, Wu D, Chen H, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ*. 2020;1091:m1091.
5. Chakraborty I, Maity P. Covid-19 outbreak: migration, effects on society, global environment and prevention. *Sci Total Environ*. 2020;728:138882.
6. Richards M, Anderson M, Carter P, Ebert BL, Mossialos E. The impact of the covid-19 pandemic on cancer care. *Nature Cancer*. 2020;1(6):565-567.
7. Organization, W.H. Coronavirus disease (COVID-19) pandemic. 2019. Accessed December 5, 2022. https://www.who.int/emergencies/diseases/novel-coronavirus-2019
8. Liu T, Wu D, Yan W, et al. Twelve-month systemic consequences of coronavirus disease 2019 (covid-19) in patients discharged from hospital: a prospective cohort study in wuhan, China. *Clin Infect Dis*. 2022;74(11):1953-1965.

9. Arkhipova-Jenkins I, Helfand M, Armstrong C, et al. Antibody response after sars-cov-2 infection and implications for immunity: a rapid living review. *Ann Intern Med*. 2021;174(6):811-821.

10. Wang Z, Muecksch F, Schaefer-Babajew D, et al. Naturally enhanced neutralizing breadth against sars-cov-2 one year after infection. *Nature*. 2021;595(7867):426-431.

11. Chen G, Wu D, Guo W, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J Clin Invest*. 2020;130(5):2620-2629.

12. Garcia-Beltran WF, Lam EC, Astudillo MG, et al. Covid-19-neutralizing antibodies predict disease severity and survival. *Cell*. 2021;184(2):476-488.

13. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial intelligence (ai) applications for covid-19 pandemic. *Diabetes Metab Syndr*. 2020;14(4):337-339.

14. Jamshidi M, Lalbakhsh A, Talla J, et al. Artificial intelligence and covid-19: deep learning approaches for diagnosis and treatment. *IEEE Access*. 2020;8:109581-109595.

15. Zhang K, Liu X, Shen J, et al. Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. *Cell*. 2020;181(6):1423-1433.

16. Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets. *Nat Commun*. 2020;11(1):4080.

17. Yao Q, Xiao L, Liu P, Zhou SK. Label-free segmentation of covid-19 lesions in lung ct. *IEEE Trans Med Imaging*. 2021;40(10):2808-2819.

18. Wang G, Liu X, Li C, et al. A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images. *IEEE Trans Med Imaging*. 2020;39(8):2653-2663.

19. Feng Z, Yu Q, Yao S, et al. Early prediction of disease progression in covid-19 pneumonia patients with chest ct and clinical characteristics. *Nat Commun*. 2020;11(1):4968.

20. Fang C, Bai S, Chen Q, et al. Deep learning for predicting covid-19 malignant progression. *Med Image Anal*. 2021;72:102096.

21. Zhou Y, Wang F, Tang J, Nussinov R, Cheng F. Artificial intelligence in covid-19 drug repurposing. *Lancet Digit Health*. 2020;2(12): e667-e676.

22. Santos SS, Torres M, Galeano D, Sánchez MM, Cernuzzi L, Paccanaro A. Machine learning and network medicine approaches for drug repositioning for covid-19. *Patterns*. 2022;3(1):100396.

23. Chia WN, Zhu F, Ong SWX, et al. Dynamics of sars-cov-2 neutralising antibody responses and duration of immunity: a longitudinal study. *Lancet Microbe*. 2021;2(6):e240-e249.

24. Guo J, Li L, Wu Q, et al. Detection and predictors of anti-sars-cov-2 antibody levels in covid-19 patients at 8 months after symptom onset. *Future Virol*. 2021;16(12):795-804.

25. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020;20:310.

26. Martinez GS, Ostadgavahi AT, Al-Rafat AM, et al. Model-interpreted outcomes of artificial neural networks classifying immune biomarkers associated with severe infections in ICU. *Front Immunol*. 2023;14:1137850.

27. Nie J, Li Q, Wu J, et al. Establishment and validation of a pseudovirus neutralization assay for sars-cov-2. *Emerg Microbes Infect*. 2020;9(1): 680-686.

28. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv*. 2015;9351:234-241.

29. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30:4765-4774.

30. Dosovitskiy A, Beyer L, Kolesnikov A, et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. International Conference on Learning Representations; 2021.

31. He X, Liao L, Zhang H, Nie L, Hu X, Chua T-S. *Neural Collaborative Filtering*. International Conference on World Wide Web; 2017: 173-182.

32. Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32:8024-8035.

33. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(56):1929-1958.

34. Kingma DP, Ba J. *Adam: A Method for Stochastic Optimization*. International Conference on Learning Representations; 2015.

35. Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Series B Stat Methodol*. 1974;36(2): 111-133.

36. Khoury DS, Cromer D, Reynaldi A, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic sars-cov-2 infection. *Nature Med*. 2021;27(7):1205-1211.

37. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30: 4765-4774.

38. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual Explanations From Deep Networks Via Gradient-Based Localization. *Int J Comput Vis*; 2020:128(2):336-359.

39. Pradenas E, Trinité B, Urrea V, et al. Stable neutralizing antibody levels 6 months after mild and severe covid-19 episodes. *Med*. 2021;2(3):313-320.

40. Cervia C, Nilsson J, Zurbuchen Y, et al. Systemic and mucosal antibody responses specific to sars-cov-2 during mild versus severe covid-19. *J Allergy Clin Immunol*. 2021;147(2):545-557.

41. Ogega CO, Skinner NE, Blair PW, et al. Durable sars-cov-2 b cell immunity after mild or severe disease. *J Clin Invest*. 2021;131(7): 145516.

42. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemostasis*. 2020;18(4):844-847.

43. Zhu J, Paul WE. Cd4 t cells: fates, functions, and faults. *Blood*. 2008;112(5):1557-1569.

44. Vietzen H, Danklmaier V, Zoufaly A, Puchhammer-Stöckl E. High-affinity fcyriiia genetic variants and potent nk cell-mediated antibody-dependent cellular cytotoxicity (adcc) responses contributing to severe covid-19. *Genet Med*. 2022; 24(7):1449-1458.

45. Klok FA, Kruip MJHA, van der Meer NJM, et al. Incidence of thrombotic complications in critically ill icu patients with covid-19. *Thromb Res*. 2020;191:145-147.

46. Brodin P. Immune determinants of covid-19 disease presentation and severity. *Nature Med*. 2021;27(1):28-33.

47. Siemieniuk RA, Bartoszko JJ, Zeraatkar D, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ*. 2020;370:m2980.

48. Janssen KJM, Donders ART, Harrell FE, et al. Missing covariate data in medical research: to impute is better than to ignore. *JCE*. 2010;63(7):721-727.

49. Purwar A, Singh SK. Hybrid prediction model with missing value imputation for medical data. *Expert Syst Appl*. 2015;42(13): 5621-5631.

50. Barnard J, Meng X-L. Applications of multiple imputation in medical studies: from aids to nhanes. *Stat Methods Med Res*. 1999;8(1): 17-36.

51. Sarwar B, Karypis G, Konstan J, Reidl J. *Item-Based Collaborative Filtering Recommendation Algorithms*. International Conference on World Wide Web; 2001:285-295.

52. Su X, Khoshgoftaar TM. A survey of collaborative filtering techniques. *Adv Artif Intell*. 2009;2009:1-19.

53. Zhang S, Yao L, Sun A, Tay Y. Deep learning based recommender system: a survey and new perspectives. *ACM Comput Surv.* 2020;52(1):1-38.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.