Clinical and Experimental Medicine (2025) 25:236
https://doi.org/10.1007/s10238-025-01772-2

RESEARCH

Predicting SARS-CoV-2-specific CD4$^+$ and CD8$^+$ T-cell responses elicited by inactivated vaccines in healthy adults using machine learning models

Jie Ning[1] · Yayi Ren[1] · Zelin Zhang[1] · Xianhuang Zeng[1] · Qinjin Wang[1] · Jia Xie[1] · Yue Xu[1] · Yali Fan[1] · Huilan Li[1] · Aixia Zhai[1] · Bin Li[1] · Chao Wu[1] · Ying Chen[1]

Received: 13 February 2025 / Accepted: 10 June 2025

Abstract

The ongoing evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants highlights the importance of monitoring immune responses to guide vaccination strategies. Although neutralizing antibodies (NAbs) have garnered increasing attention, T-cells are crucial for conferring long-lasting immunity, especially their resilience against viral mutations. However, assessing T-cell responses clinically has been hindered by cost and complexity. In this study, we recruited a cohort of 134 healthy adults, who had been immunized with three doses of the SARS-CoV-2 inactivated vaccine. Cellular immunity elicited by a comprehensive array of overlapping peptides covering the entire sequence of the virus's structural proteins was assessed by intracellular cytokine staining (ICS). Additionally, a dataset including demographic information, routine blood indices, and immune cell indicators comprising 32 variables was collected. Multivariate analysis revealed age and days post-vaccination as key factors influencing the strength of the T-cell response. Importantly, random forest (RF) and classification and regression tree (CART) algorithms were employed, along with 8 easily accessible indicators to formulate predictive models for the SARS-CoV-2-specific CD4$^+$ and CD8$^+$ T-cell responses. Besides, these models demonstrated substantial accuracy ($r > 0.9$) in both the training and testing sets. Our findings offer an efficient and economical methodology for evaluating the T-cell reactions in healthy adults following inactivated SARS-CoV-2 vaccination, which is visualizable and easy to use, providing a novel strategy for assessing cellular immunity after vaccination.

Keywords SARS-CoV-2 · Cellular immunity · Random forest · Classification and regression tree · Machine learning

Abbreviations

COVID-19	Coronavirus disease 2019
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
WHO	World Health Organization
NAbs	Neutralizing antibodies
ICS	Intracellular cytokine staining
ML	Machine learning
RF	Random forest
CART	Classification and regression tree
E	Envelope
M	Membrane
N	Nucleocapsid
S	Spike
T	The total of four structural proteins
WT	Wild-type
WBC	White blood cell
NEU	Neutrophil granulocyte
LYM	Lymphocyte
MONO	Monocyte
EOS	Eosinophil granulocyte
BASO	Basophil granulocyte
RBC	Red blood cell
HGB	Hemoglobin
MCV	Mean corpuscular volume
MCH	Mean corpuscular hemoglobin
HCT	Hematocrit
RDW-CV	Red cell volume distribution width

Chao Wu and Ying Chen jointly supervised this work.

✉ Chao Wu
wuch57@mail.sysu.edu.cn

✉ Ying Chen
cheny883@mail.sysu.edu.cn; wuch57@mail.sysu.edu.cn

[1] Department of Laboratory Medicine, The Eighth Affiliated Hospital of Sun Yat-sen University, Shenzhen, China

Published online: 08 July 2025

RDW-SD	Standard deviation of red blood cell distribution width
PLT	Platelet absolute count
MPV	Mean platelet volume
PCT	Platelet count
PDW	Platelet distribution width
% IncMSE	Increase in Mean Square error
IncNodePurity	Increase in Node Purities
RMSE	Root mean squared error
R2	Coefficient of determination
RCS	Restricted Cubic Spline
GLR	General Linear Regression

Background

Emerging at the close of 2019, the coronavirus disease 2019 (COVID-19), triggered by the zoonotic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has profoundly affected global society [1]. According to the World Health Organization (WHO), by December 2024, the disease had led to over 776 million infections and claimed more than 7 million lives [2]. The onset of COVID-19 spurred an unprecedented worldwide scientific endeavor encompassing all aspects of SARS-CoV-2 biology, resulting in the rapid development of vaccines [3]. Despite the shift of COVID-19 from pandemic to endemic status and a waning sense of urgency, the swift dissemination and mutation of novel SARS-CoV-2 strains continue to bring substantial threats to public health [4, 5].

The severity of COVID-19 is influenced by a complex interaction of factors, including the pathogenicity of different SARS-CoV-2 variants and the strength of the individual's antiviral immune system. The immune response, initiated by vaccination or exposure to the virus, encompasses B cells, $CD4^+$ T-cells, and $CD8^+$ T-cells, all are crucial for virus clearance and the development of long-lasting immunological memory [6, 7]. Studies have highlighted the essential function of $CD8^+$ T-cells in combating the infection during its acute phase. Concurrently, $CD4^+$ T-cells and neutralizing antibodies (NAbs) produced by B cells are acknowledged for their contributions to limiting viral dissemination and facilitating pathogen clearance [8].

Nonetheless, numerous studies indicate that immune reaction intensity diminishes with time. Koerber et al. observed a rapid decrease in the levels of virus-specific NAbs among patients who had recovered from COVID-19. Conversely, polyfunctional T-cells have demonstrated a sustained level of functionality [9, 10]. Furthermore, the protective efficacy of NAbs, particularly those targeting the spike protein's receptor-binding domain (RBD), appears to be diminishing in the face of continuous viral evolution [11, 12]. However, the relative conservation of SARS-CoV-2 T-cell epitopes limits the impact of viral mutations on the specificity of T-cell responses [13, 14]. This intrinsic stability underscores the critical role of cellular immunity in providing sustained protection, particularly against severe illness and mortality [15]. Given these findings, the need for continuous immune surveillance in vaccinated individuals is becoming increasingly evident. The rising incidence of breakthrough infections among vaccinated populations worldwide underscores this necessity. Such surveillance is essential for determining the requirement and optimal timing for booster vaccinations, which are critical for enhancing and prolonging both individual and public immunity.

Currently, various techniques are available for detecting SARS-CoV-2 specific IgA, IgM, and IgG antibodies, such as chemiluminescent immunoassay, enzyme-linked immunosorbent assay, and lateral flow immunoassay [16–18]. Regarding cellular immunity, while the enzyme-linked immunospot assay is capable of quantifying T-cell responses specific to SARS-CoV-2 [19], it lacks the capacity to offer insights into cytokine-producing cell types. Intracellular cytokine staining (ICS) is a more sophisticated technique that can indicate specific cell subpopulations and detect polyfunctional cells. In our previous research, we employed the ICS method to characterize the SARS-CoV-2 specific T-cell responses before vaccination and after the first, second, and third doses of BBIBP-CorV [20]. However, the substantial financial expenditure required to construct antigen peptide pools, along with the considerable time commitment and procedure complexity, have collectively acted as barriers to its broader clinical application [21]. Therefore, despite the extensive use of SARS-CoV-2 serology in identifying individuals with previous infections or active disease and in evaluating vaccine effectiveness [22, 23], routine assessments of $CD4^+$ and $CD8^+$ T-cell responses have not been conducted. Achieving a rapid and direct evaluation of T-cell responses remains a considerable challenge.

Over recent years, the domain of artificial intelligence, particularly with the rise of deep learning and machine learning (ML) algorithms, has garnered significant interest. Characterized by their nonlinearity properties, robustness to errors, and ability to operate in real time, these algorithms are highly suitable for addressing intricate healthcare challenges [24]. The medical field has notably benefited from ML, particularly in the realm of predictive analytics for a spectrum of diseases and facilitating the early identification of chronic health issues, including acute liver failure, cardiovascular diseases, chronic kidney disease, hypertension, and diabetes [25, 26]. According, we sought to expand upon our previous study of individuals who had received three doses of vaccine by recruiting a sufficient number of additional participants. By collecting relevant data from these subjects, we aim to

utilize machine methods to efficiently predict the levels of T-cell responses following SARS-CoV-2 vaccination.

Among the array of ML algorithms, such as random forest (RF), LASSO, classification and regression tree (CART), and XGBoost. RF and CART have demonstrated exceptional precision across a multitude of scientific fields. The RF excels over other methodologies due to its adeptness at handling complex, nonlinearly datasets, its straightforward calibration, resilience against noise, and its capability for rapid parallel processing [27]. Iwendi et al. employed a comprehensive dataset that integrated patient health profiles, demographic details, and geographical information to predict the severity and clinical outcomes of COVID-19 cases using the RF methodology [28]. Besides, CART is recognized for its versatility in both classification and regression tasks, particularly due to its adaptability to a wide range of data types and sensitivity to minor data changes [29]. Zimmerman and colleagues used the CART model to increase the efficacy of COVID-19 screening and detection by differentiating between individuals with confirmed laboratory diagnoses and non-cases [30]. However, there are a few studies on predictive models that measure the specific cellular responses of T-cells post-SARS-CoV-2 vaccination, which is the primary focus of our research.

In our research, we assessed the responses of SARS-CoV-2-specific $CD4^+$ and $CD8^+$ T-cells among adults aged between 18 and 60 years who had been administered three doses of inactivated vaccines. The cellular immunity reaction stimulated by an array of overlapping peptide pools that covered the full sequence of the spike (S), nucleocapsid (N), membrane (M), and envelope (E) proteins, as well as their combined total pool (T), were measured using the Intracellular cytokine staining (ICS) method. Concurrently, a comprehensive dataset comprising 32 variables for the study participants was assembled. With this dataset, we employed the RF and CART algorithms to identify predictive indicators and to develop models that are capable of predicting specific cellular immunity with precision and efficiency. Our research presents a cost-effective and rapid method for assessing the intensity of T-cell responses in individuals vaccinated with inactivated SARS-CoV-2 vaccines. This method does not necessitate advanced laboratory equipment, thereby broadening its use for assessing immune status against SARS-CoV-2 and determining the need for booster vaccinations across various healthcare settings, from hospitals to community clinics. Additionally, these models developed in this study offer innovative methods and insight for evaluating post-vaccination immune responses against multiple infectious diseases.

Methods

Study population and data collection

In this research, we enrolled a cohort of 134 healthy adults aged 18 to 60 years. All participants had no history of SARS-CoV-2 infection, as confirmed by negative results from pharyngeal swab nucleic acid tests. Moreover, none of the participants had received any COVID-19 vaccines other than the inactivated SARS-CoV-2 vaccine BBIBP-CorV. Each individual had completed a three-dose regimen of the BBIBP-CorV between November 2021 and September 2022, with the first and second doses spaced 1 month apart, followed by a 6-month interval before the third dose. A compilation of demographic information, a set of 26 routine blood examination results, along with immune cell markers, are presented in Table 1. The data encompasses a variety of parameters including age, sex, days post-vaccination, antibody binding titers to the RBD of both the wild-type (WT) Wuhan-Hu-1 strain and the Omicron variants. The proportions of T-cells ($CD3^+$, $CD4^+$, $CD8^+$) in the total blood and the $CD4^+/CD8^+$ T-cells ratio were obtained through flow cytometry. The remaining parameters, including white blood cell (WBC), neutrophil granulocyte (NEU and NEU%), monocyte (MONO and MONO%), lymphocyte (LYM and LYM%), basophil granulocyte (BASO and BASO%), eosinophil granulocyte (EOS and EOS%), hemoglobin (HGB), red blood cell (RBC), mean corpuscular hemoglobin (MCH), mean corpuscular volume (MCV), the standard deviation of red blood cell distribution width (RDW-SD), red cell volume distribution width (RDW-CV), hematocrit (HCT), platelet count (PCT), mean platelet volume (MPV), platelet distribution width (PDW), and platelet absolute counts (PLT), were derived from routine blood examination results.

This research was conducted under the Declaration of Helsinki, and approved by the Ethics Review Committee of the Eighth Affiliated Hospital of Sun Yat-sen University, with the reference number 2021–005-01. Besides, written informed consent was obtained from all subjects involved in the study.

Assessment of SARS-CoV-2-specific $CD4^+$ and $CD8^+$ T-cell responses

The evaluation of the T-cell responses was conducted according to the methods previously reported [20]. In summary, peripheral blood mononuclear cells (PBMCs) were separated from venous blood samples by centrifugation, employing Ficoll-Hypaque (TBDscience, CHN), and then re-suspended in RPMI 1640 culture medium (Gibco, USA)

Table 1 Baseline table of the 134 participants

Variables	Total ($n=134$)
Age	36.21 ± 8.15
Sex, *n* (%)	
Female	73 (54.48%)
Male	61 (45.52%)
Days post-vaccination	89.00 (65.00, 145.00)
Serology (AU/ml)	
NAb (WT)	16.80 (4.93, 46.56)
NAb (Omicron)	4.16 (0.00, 12.99)
Immune cell indicators	
$CD3^+$ T-cell%	70.25 (63.28, 75.65)
$CD4^+$ T-cell%	52.42 ± 9.34
$CD8^+$ T-cell%	30.90 (26.57, 36.08)
$CD4^+/CD8^+$ T-cell	1.68 (1.32, 2.10)
Routine blood indicators	
WBC	6.01 (5.25, 7.00)
NEU	3.33 (2.75, 4.03)
NEU%	56.15 (51.23, 59.90)
LYM	2.16 ± 0.52
LYM%	34.95 (30.92, 39.53)
MONO	0.34 (0.28, 0.43)
MONO%	5.82 ± 1.23
EOS	0.12 (0.07, 0.21)
EOS%	2.05 (1.30, 3.10)
BASO	0.03 (0.02, 0.04)
BASO%	0.50 (0.30, 0.60)
RBC	4.94 (4.62, 5.42)
HGB	143.00 (133.25, 159.00)
MCV	90.10 (87.93, 92.27)
MCH	29.50 (28.52, 30.30)
MCHC	327.00 (322.00, 331.00)
HCT	0.44 (0.41, 0.49)
RDW-CV	12.80 (12.43, 13.17)
RDW-SD	40.90 (39.90, 41.80)
PLT	258.50 (237.00, 302.25)
MPV	9.81 ± 0.98
PCT	0.25 (0.23, 0.29)
PDW	16.10 (15.90, 16.40)

NAb (WT) and NAb (Omicron): antibody binding titers to the RBD of the wild-type (WT) Wuhan-Hu-1 strain and Omicron variants; Data are mean ± standard deviation or medians (1st Quartile, 3rd Quartile) and depend on the test for normality

supplemented with 1% penicillin–streptomycin (Gibco) and 10% fetal bovine serum (Gibco). A comprehensive pool of 324 peptides (18-mer), each with a six-amino-acid overlap, were crafted to encompass the entire S, N, M, and E proteins of the SARS-CoV-2 Wuhan-Hu-1 strain, with a purity exceeding 90% (ChinaPeptides, CHN). The PBMCs were distributed into 96-well plates at a density of 1×10^6 cells/well and were exposed to a mixture of S, N, M, E, and T peptide pool for 5 h at a concentration of 1.5 μM, along with GolgiStop (Becton Dickinson, USA). Cells treated with DMSO served as the negative control, while cells stimulated with PMA (50 ng/ml, DARKEWE, CN) and ionomycin (1ug/ml, DARKEWE, CN), were used as the positive control. PBMCs were later harvested, and dead cells were labeled by LIVE/DEAD Zombie Viability (BioLegend). After washing with PBS, cells were collected and incubated with a panel of surface antibodies, including anti-CD3, anti-CD4, and anti-CD8. Following PBS washing, cells were fixed with 4% paraformaldehyde and were permeabilized using permeabilization wash buffer (Becton Dickinson), followed by intracellular staining with an anti-IFN-γ antibody and an isotype control mouse IgG1. All antibodies used in this study are showed in Table S1. The flow cytometry data were captured using the FACSCanto II LSRFortessa (Becton Dickinson), with background signals being subtracted prior to analysis.

NAb assay

To assess SARS-CoV-2 NAb levels, we employed the iFlash-2019-nCoV NAb kit, consistent with our previous methodology [20]. In brief, serum was mixed with RBD-coated microparticles, and then acridinium-ester-labeled ACE2 was added to bind the unoccupied RBD sites. Non-specifically bound materials were removed with a wash buffer, and chemiluminescent signals, indicative of antibody presence, were quantified as relative light units (RLUs) upon introducing a signal buffer. NAb titers were then calculated from RLU values using a calibration curve.

Random forest model to predict and evaluate cellular immunity

The dataset was randomly segregated, allocating 70% for the training set and the remaining 30% designated for the testing set. The RF algorithm was then applied to the training subset to evaluate the influence of diverse indicators on cellular immunity against SARS-CoV-2. The RF model assessed multivariate significance by systematically eliminating predictor variables from individual trees within the forest, subsequently measuring the resultant accuracy changes to determine the importance of these variables. Variables importance was ascertained by quantifying and ranking based on the Increase in Node Purities (IncNodePurity) and percentage of Increase in Mean Square error (% IncMSE), both of which are tied to the loss function, with the optimal loss function selected through the most effective segmentation. To enhance the RF model's stability and accuracy and reduce the risk of overfitting, several strategies have been employed. These steps involved increasing the number of trees (n_estimators) to 500 for enhanced

predictive averaging and variance reduction, limiting the variables considered at each split (max_features) to three to foster tree diversity and minimize correlation, and applying a maximum depth limit (max_depth) to avoid the model from becoming overly complex with the training dataset.

For the predictive model, variables were selected based on the criteria of maximized coefficient of determination (R^2) and minimized root mean squared error (RMSE) values [31]. The model's predictive efficacy was then validated using the test set. Predictive accuracy was evaluated by comparing the predicted values with the actual experimental data, employing scatter plots and Spearman correlation analysis for both the training and testing datasets.

Tenfold cross-validation

To enhance model reliability and assess generalizability, we performed tenfold cross-validation using the caret package in R. The predictor variables were the eight features selected based on variable importance from the initial random forest model. Model performance was evaluated across each fold using root mean square error (RMSE), R-squared (R^2), and mean absolute error (MAE). The optimal number of variables tried at each split (mtry) was selected based on the lowest average RMSE.

Visualization models for assessing T-cell responses through the CART algorithm

The RF methodology is an ensemble learning method that integrates multiple CARTs and combines their predictions to generate a final output. A single CART utilizes the entire feature set to determine the optimal split at each node, employing various thresholds and nodes to assess and calculate the value of the predicted variable. Factor variables can be categorical for classification trees or continuous variables for regression trees [32]. In our study, we performed a regression tree analysis to assess the quantitative interrelations among a range of variables and to predict the extent of cellular immunity across different conditions.

Data analysis

Variables that adhered to a normal distribution were depicted using the mean ± standard deviation, and non-normally distributed continuous variables were described as medians (interquartile ranges, i.e., upper and lower quartiles). Categorical data were presented as counts and percentages (%). Correlations between the two variables were assessed using Spearman's correlation. In addition, General Linear Regression (GLR) was applied to examine linear relationships between the independent and dependent variables, while Restricted Cubic Spline (RCS) analysis was used for nonlinear relationships.

Statistically significant was set at the *P* value threshold of less than 0.05. For the construction of models based on ML, R software, version 4.2.3 (The R Project for Statistical Computing, Vienna, Austria), was employed. The specific R packages employed for these methodologies included the rms package for RCS analysis, the rpart and rpart.plot packages for the CART model, and the random Forest package for the random forest.

Results

Descriptive statistics of the cohort and basic exploratory data analysis

A total of 134 healthy participants, aged between 18 and 60 years old, who had completed the three-dose schedule of inactivated SARS-CoV-2 vaccines and had not been infected with SARS-CoV-2, were enrolled in this study. Table 1 provides an overview of the demographic and baseline characteristics of these participants. The volunteers had an average age of 36 years, and 61 individuals (comprising 45.52% of the study cohort) being men. Serum NAbs against both the Wuhan-Hu-1 and Omicron strains of SARS-CoV-2 were quantified, with median titers being16.80 AU/ml and 4.16 AU/ml, respectively.

In this study, we developed a comprehensive set of 324 18-mer overlapping peptides that encompass the complete sequences of the SARS-CoV-2 four structural proteins. PBMCs were extracted and exposed to the S、N、M、E and T peptide pool, respectively. The specific cellular responses to SARS-CoV-2 were then evaluated employing ICS assays. The gating strategy is detailed in Figure S1A, and the representative fluorescence-activated cell sorting (FACS) plots are presented in Figure S1B, depicting the expression of IFN-γ by $CD4^+$ and $CD8^+$ T-cell (as indicated on the y-axis), following stimulation with the five peptide pools or the negative control. Table 2 illustrates the quantitative results of specific $CD4^+$ and $CD8^+$ T-cell responses to the S, N, M, E, and T pools, highlighting that certain individuals exhibit notably elevated response levels. The T pool, in particular, triggered the highest median and average levels of T-cell responses among the study participants. Additionally, the incidence of individuals with T-cell reactions to the T pool notably surpassed the response rate to any one of the four individual structural proteins, with a positive response rate reaching 85.82% for both $CD4^+$ and $CD8^+$ T-cells, as detailed in Table 3.

Table 2 Quantitative assessment of specific $CD4^+$ and $CD8^+$ T-cell response against SARS-CoV-2 structural proteins

Structural protein	Min	1stQu	Median	Mean	3rdQu	Max
$CD4^+$ T-cell response level (%)						
S	0.000	0.000	0.000	0.072	0.069	1.200
N	0.000	0.000	0.035	0.101	0.115	1.222
M	0.000	0.000	0.000	0.027	0.019	0.570
E	0.000	0.000	0.000	0.028	0.022	0.680
T	0.000	0.000	0.078	0.269	0.220	2.130
$CD8^+$ T-cell response level (%)						
S	0.000	0.000	0.101	0.315	0.444	2.240
N	0.000	0.000	0.035	0.358	0.269	4.855
M	0.000	0.000	0.000	0.068	0.058	0.860
E	0.000	0.000	0.000	0.092	0.026	1.424
T	0.000	0.103	0.372	0.933	1.456	5.510

S: Spike protein peptide pool; N: Nucleocapsid peptide pool; M: Membrane protein peptide pool; E: Envelope protein peptide pool; T: the total of four structural protein peptide pool

Table 3 SARS-CoV-2 structural proteins specific $CD4^+$ and $CD8^+$ T-cell responses rate

Structural protein	$CD4^+$ T-cell response rate *n* (%)	$CD8^+$ T-cell response rate *n* (%)
S, *n* (%)		
No	46 (34.33)	46 (34.33)
Yes	88 (65.67)	88 (65.67)
N, *n* (%)		
No	52 (38.81)	61 (45.52)
Yes	82 (61.19)	73 (54.48)
M, *n* (%)		
No	97 (71.64)	92 (68.66)
Yes	38 (28.36)	42 (31.34)
E, *n* (%)		
No	94 (70.15)	92 (68.66)
Yes	40 (29.85)	42 (31.34)
T, *n* (%)		
No	19 (14.18)	19 (14.18)
Yes	115 (85.82)	115 (85.82)

S: Spike protein peptide pool; N: Nucleocapsid peptide pool; M: Membrane protein peptide pool; E: Envelope protein peptide pool; T: the total of four structural protein peptide pool; data are presented as medians (1st Quartile, 3rd Quartile)

Correlation analysis

To determine the connections between various variables and the T-cell response, we initially conducted a Spearman's correlation analysis, with the findings presented in Table S2 and Table S3. Considering the comprehensive coverage of peptides derived from the four structural proteins by the T pool and the fact that the proportion of individuals manifesting cellular responses to each of the four individual structural proteins is markedly less than the proportion responding to the T pool, the T pool specific T-cell responses were selected as the targets for evaluation and prediction in this research. The findings indicated that T-specific $CD4^+$ and $CD8^+$ T-cell responses were correlated with age, days post-vaccination, and NAb (WT) ($P < 0.05$). Additionally, there was a statistically significant correlation between the levels of NAb (Omicron) and the T-specific $CD8^+$ T-cell responses ($P < 0.001$). Nonetheless, the overall correlation strength was moderate, with a correlation coefficient below 0.5, as depicted in Fig. 1.

This heatmap illustrates the correlation coefficients between the $CD4^+$ and $CD8^+$ T-cell responses and several variables, including age, the number of days post-vaccination, and the levels of NAb (WT) and NAb (Omicron). The intensity of the correlation is indicated by the color spectrum located on the side. NAb, neutralizing antibody; WT, wild type, referring to the SARS-CoV-2 Wuhan-Hu-1 strain.

Linear and nonlinear regression analysis

In our investigation of the linkage between the magnitude of SARS-CoV-2-specific $CD4^+$ and $CD8^+$ T-cell response magnitudes and three highly related indicators—age, days post-vaccination, and NAb—both GLR and RCS analysis were utilized. Our findings indicated that age, after adjusting for days and NAb as covariates, exhibited a statistically significant positive linear correlation with the response levels of both $CD4^+$ and $CD8^+$ T-cell responses ($P < 0.05$, Fig. 2A and B). In contrast, days post-vaccination, when adjusted for age and NAb, demonstrated a statistically significant linear negative relationship with the T-cell response ($P < 0.05$, Fig. 2C and D). According to the univariate RCS analysis, an 'L'-shaped nonlinear relationship between NAb levels and the response levels of both $CD4^+$ and $CD8^+$ T-cells,

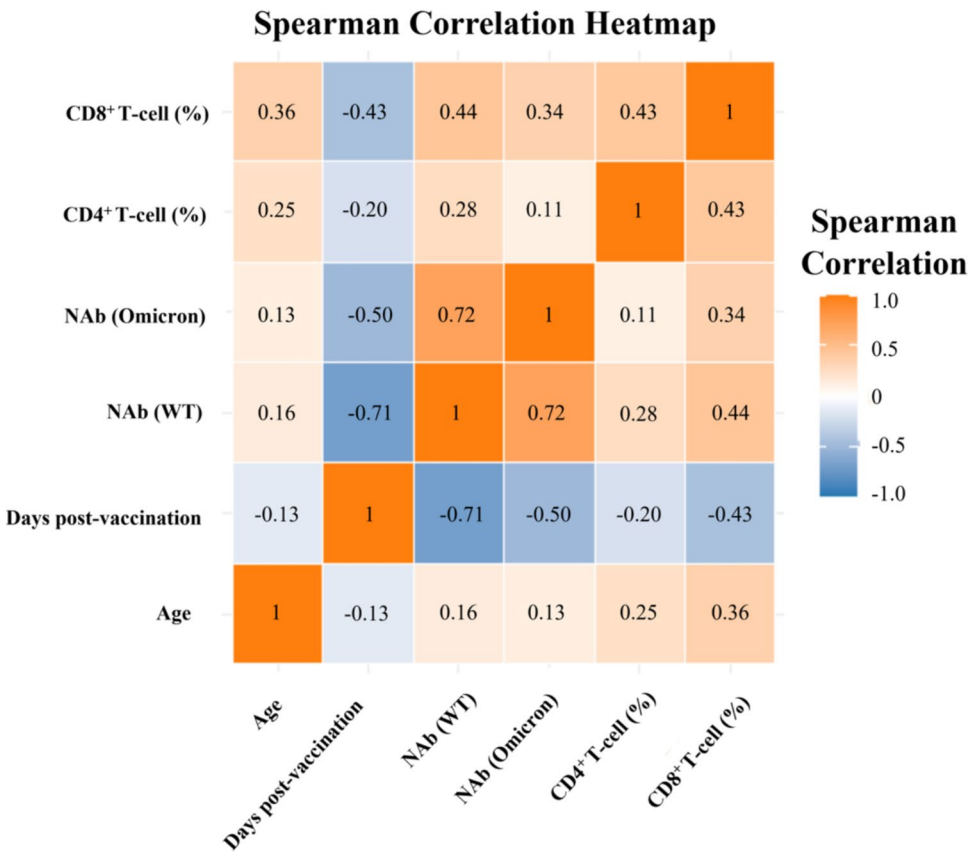

Fig. 1 Heatmap displaying correlations between $CD4^+$ and $CD8^+$ T-cell responses with multiple variables

indicating a threshold effect where T-cell responses plateau beyond a certain NAb titer ($P < 0.05$, Figure S2). However, upon multivariate adjustment for days since vaccination and age, these nonlinear relationships dissipated, suggesting that the initial 'L'-shaped pattern was influenced by the confounding effects of days post-vaccination and age ($P > 0.05$, Fig. 2E and F). These findings indicate that age and days post-vaccination are the predominant factors influencing T-cell response levels, and the impact of NAb was significant only in univariate analysis and not after multivariate adjustments.

(A-B) After adjusting for the number of days post-vaccination and NAb levels as covariates, a linear positive association was identified between age and the responses of $CD4^+$ T-cell (left panel) and $CD8^+$ T-cell (right panel) responses was observed. (C-D) The linear negative correlation between number of days post-vaccination and the T-cell responses, with adjustments for age and NAb as covariates. (E–F) The 'L'-shaped nonlinear relationship between the T-cell responses and the levels of NAb became statistically insignificant ($P > 0.05$) after applying a multivariate adjustment for age and days post-vaccination. The y-axis (β) represents the partial effect size (regression coefficient) of age on log-transformed $CD4^+$/$CD8^+$ T-cell response, estimated from a generalized additive model adjusted for days post-vaccination and neutralizing antibody (NAb) levels.

The prediction model established by RF algorithms

The RF model encompassed a comprehensive set of 32 variables as detailed in Table 1, to identify the key determinants of T-cell response magnitudes. Importance analysis was conducted to screen these variables, with the outcomes ranked by the % IncMSE to spotlight the most impactful variables on T-cell response levels. The prediction model was then constructed, comprising the top 8 predictors, selected based on their higher R^2 values and lower RMSE values. Alongside the strongly correlated variables- age, days post-vaccination, and NAb (WT)-the proportions of $CD3^+$, $CD4^+$, and $CD8^+$ T-cells were found to be significantly predictive of the $CD4^+$ T-cell response (Fig. 3A). Conversely, for the $CD8^+$ T-cell response, MONO, MONO%, MCV, NAb (Omicron), and NEU emerged as pivotal predictors (Fig. 3B). The importance metrics are detailed in Tables S4 and S5.

The impact of the 32 candidate variables listed in Table 1 on the levels of SARS-CoV-2-specific T-cell response was assessed. The top 8 determinants for $CD4^+$ T-cell (left panel) and $CD8^+$ T-cell responses (right panel) are displayed in descending order relative to their contribution based on the % IncMSE. MONO, monocyte; NEU, neutrophil granulocyte; MCV, mean corpuscular volume; % IncMSE, percentage of Increase in Mean Squared Error.

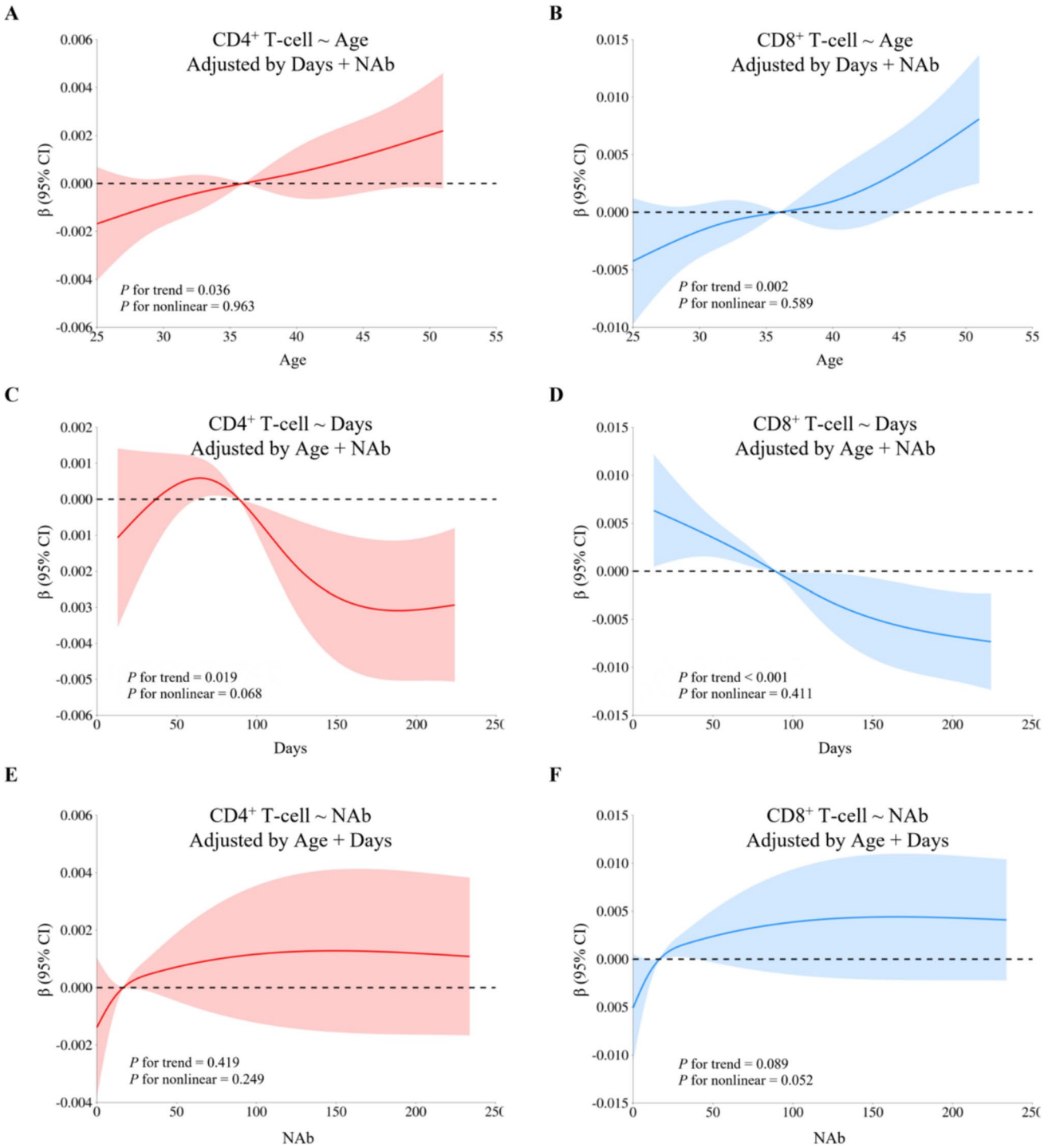

Fig. 2 Analysis of variables' impact on SARS-CoV-2 specific T-cell response via multivariate linear and nonlinear regression

The dataset was divided, with 70% ($n = 94$) allocated to the training subset and the remaining 30% ($n = 40$) designated for the testing set. Subsequently, the predictive models were subjected to verification against the test set to assess their accuracy. An analysis comparing the alignment of the predicted values with the actual data from both the training and testing datasets revealed a high degree of correlation. Specifically, the predictive model for the $CD4^+$ T-cell response demonstrated r values above 0.95 for both sets (Fig. 4A and B). Similarly, the model for the $CD8^+$ T-cell response achieved an r value greater than 0.94 (Fig. 4C and D). These findings suggest that the random forest model

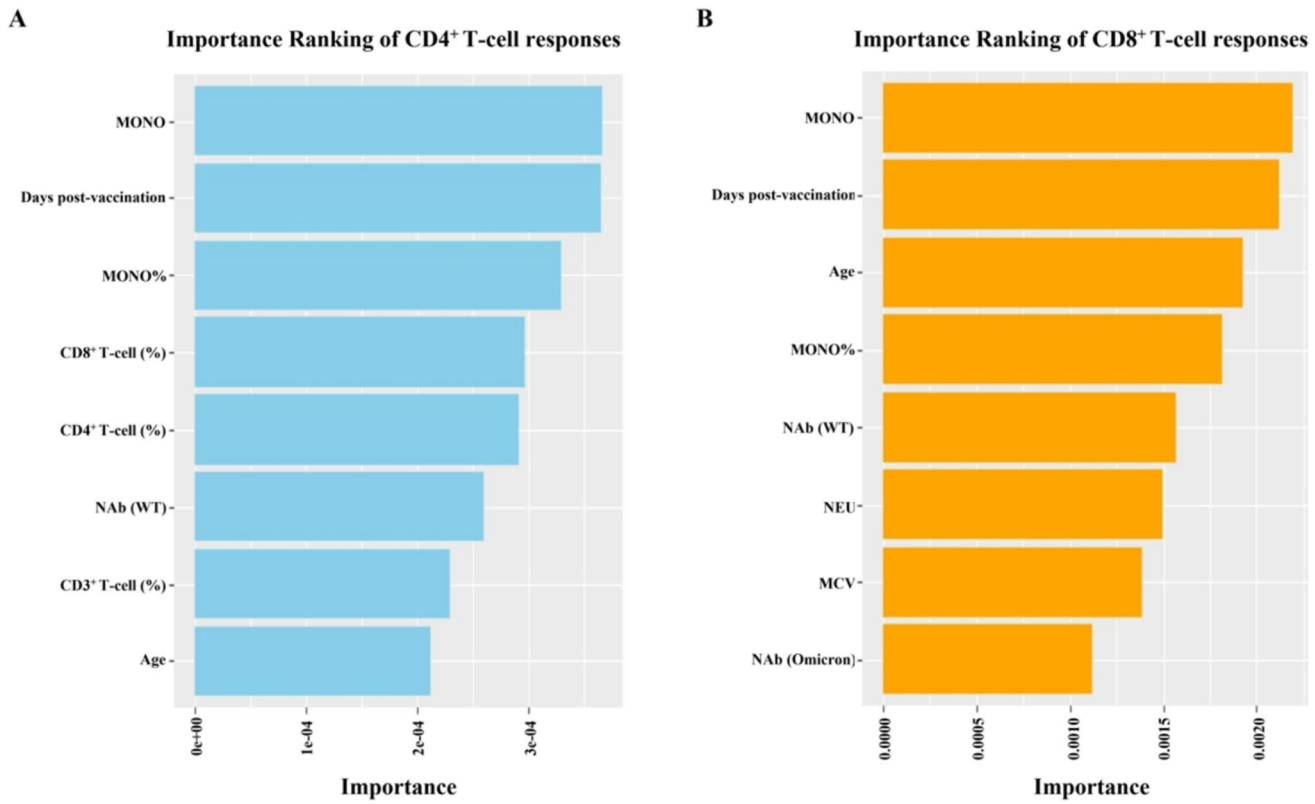

Fig. 3 Ranking of the 8 most influential variables determined by the random forest (RF) model

offers a high level of precision in predicting the responses of SARS-CoV-2-specific T-cells in healthy adults aged 18–60 years, who have received three doses of inactivated vaccines.

(A-B) The scatter plot shows the alignment between the predicted values for the $CD4^+$ T-cell response and the actual values from both the training and testing datasets. (C-D) The scatter plot illustrates the verification of the alignment between the predicted values for the $CD8^+$ T-cell response and their actual values from the training and testing datasets.

In the tenfold cross-validation, the random forest model showed stable predictive performance for $CD8^+$ T-cell responses (T8), with R^2 ranging from 0.315 to 0.675 and MAE from 0.0044 to 0.0084, indicating good model fit. In comparison, the $CD4^+$ T-cell response (T4) model exhibited greater variability, with R^2 ranging from 0.0065 to 0.796. However, its overall RMSE and MAE remained low, suggesting that despite inconsistent explanatory power, the model retains reasonable predictive value (Table S6 and S7).

Quantitative predictive T-cell responses using the CART model

CART models were built for the accurate prediction and evaluation of T-cell responses specific to SARS-CoV-2. Variables including days post-vaccination, age, the proportion of $CD3^+$ and $CD4^+$ T-cells, and LYM% were screened by the CART model for predicting the specific value of the $CD4^+$ T-cell response, identified to range from 0.0723% to 1.3% (Fig. 5A). In parallel, the specific value of the $CD8^+$ T-cell response was determined using days post-vaccination, MCV, MPV, and HGB, with values ranging from 0.17% to 3.8% (Fig. 5B). For example, considering a 50-year-old individual received the SARS-CoV-2 vaccine 100 days prior, and the proportion of $CD3^+$ T-cells is 50% of the total T-cell count, the model estimates that the SARS-CoV-2-specific $CD4^+$ T-cell response rate is likely to be around 0.34%.

(A) Variables including days post-vaccination, age, the percentages of $CD3^+$ T-cell and $CD4^+$ T-cell, and LYM were screened by the CART model to predict the $CD4^+$ T-cell response quantitatively. (B) For the prediction of $CD8^+$ T-cell responses, the CART model incorporated variables including days post-vaccination, MCV, MPV, and HGB. LYM, lymphocyte; MPV, mean platelet volume; HGB, hemoglobin; MCV, mean corpuscular volume.

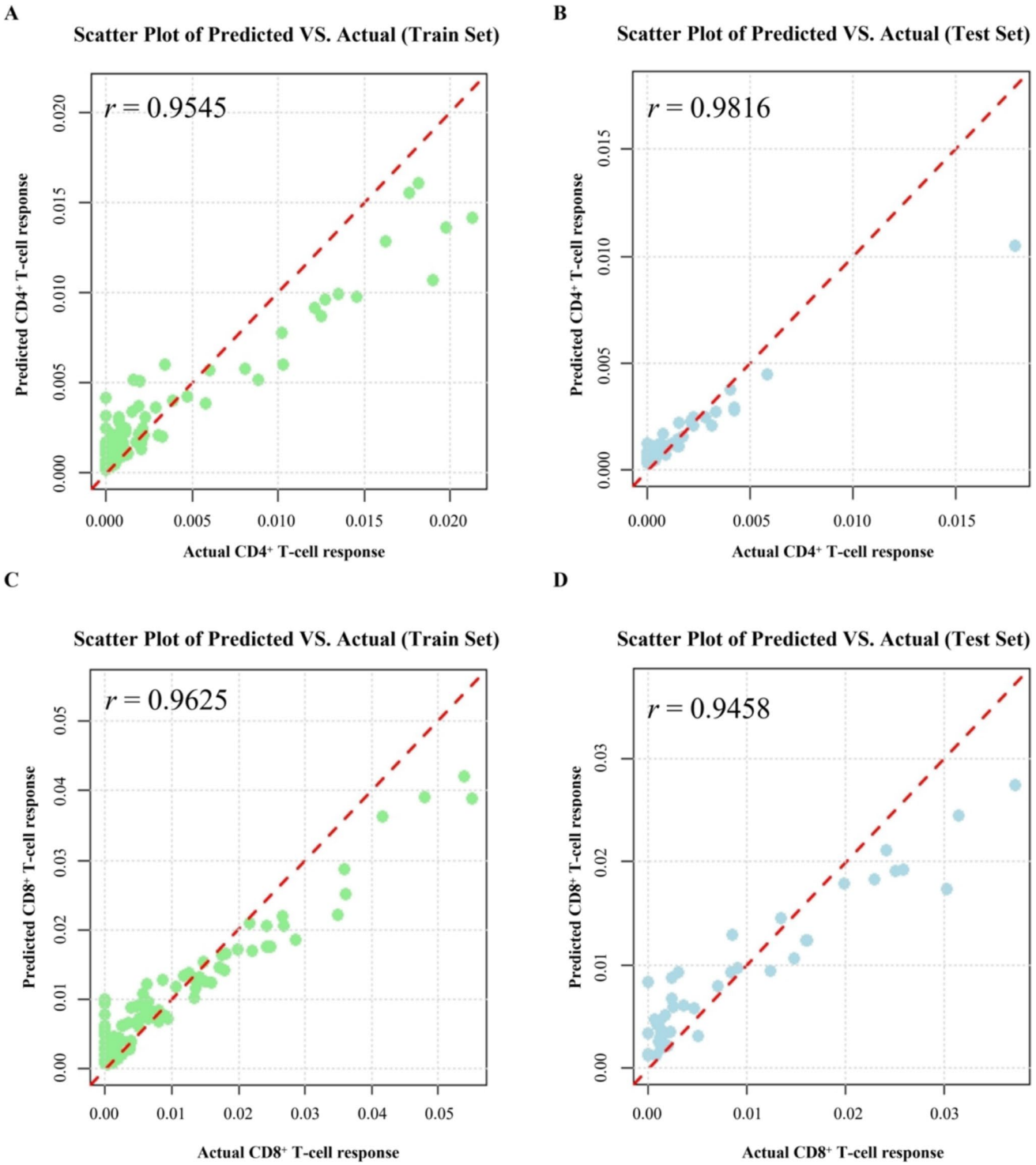

Fig. 4 RF model for predicting the responses of SARS-CoV-2-specific $CD4^+$ and $CD8^+$ T-cells

Discussion

For this study, we enrolled a group of 134 healthy donors aged 18–60 years who had received three doses of the inactivated SARS-CoV-2 vaccine. Then, the T-cell responses specific to the S, N, M, E, and T pools were detected using the ICS method. As anticipated, the result showed a higher proportion of participants displayed T-cell responses to the T pool compared to the individual structural proteins, owing to the T pool's comprehensive coverage of peptides from all

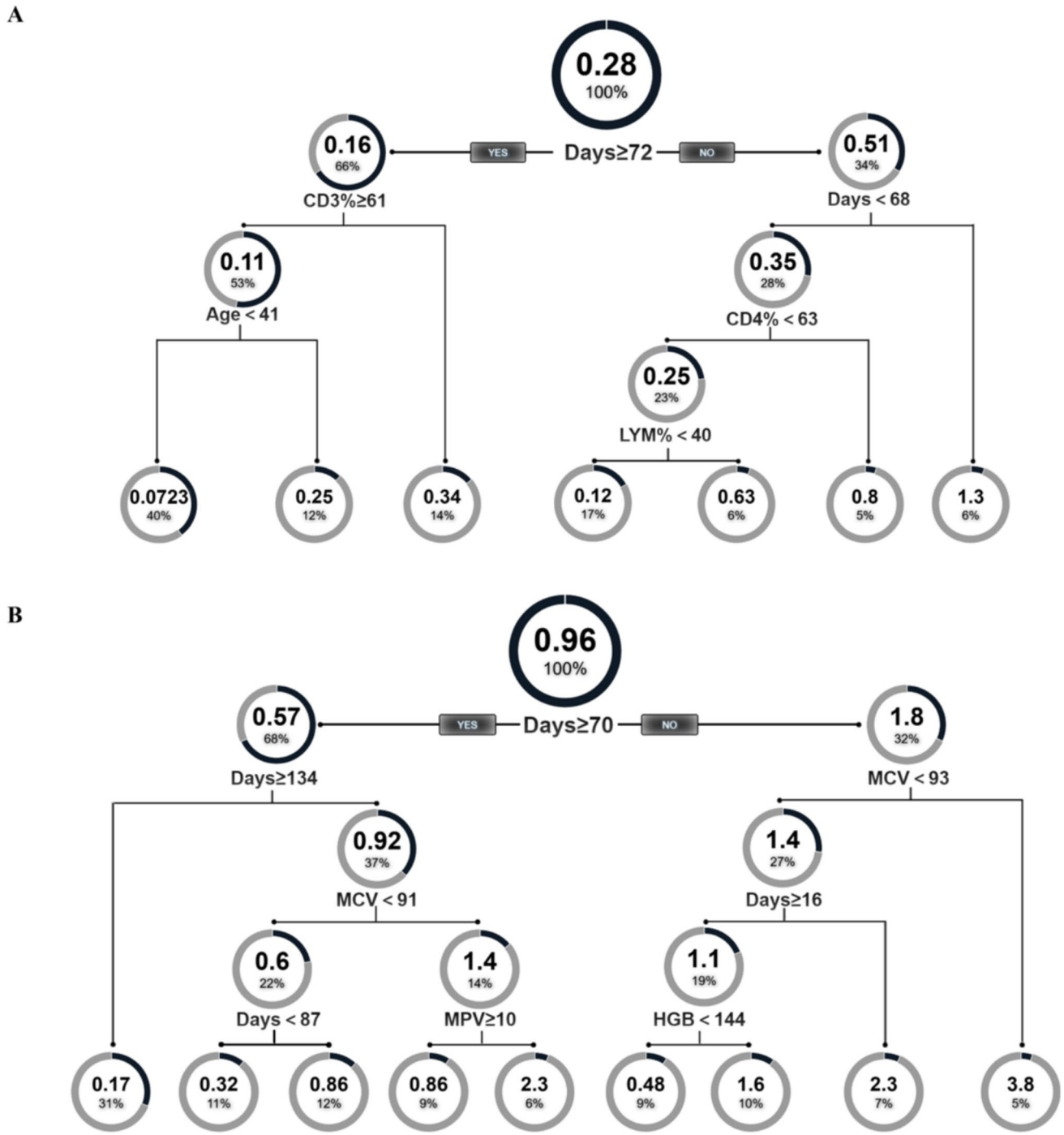

Fig. 5 Classification and regression tree (CART) model for quantitative prediction of the SARS-CoV-2-specific T-cells responses

four structural proteins. Consequently, the T pool-specific T-cell response offers a more comprehensive representation of the cellular immune response against SARS-CoV-2. Thus, our study concentrates on the analysis and predictive modeling for the response of T pool-specific T-cells.

Vaccination elicits both cellular and humoral immune responses, which collectively mediate immune protection. While inactivated vaccines primarily confer protection by inducing NAbs that block viral entry into host cells, their titers tend to wane significantly over time and exhibit reduced efficacy against variants bearing mutations in the spike protein. Nevertheless, previous studies have confirmed that BBIBP-CorV can also induce significant SARS-CoV-2-specific T-cell responses [33]. These T-cell responses,

which target conserved viral epitopes, can endure for several years, thereby offering long-term protection against viral infections and variants [34]. Therefore, the magnitude of T-cell responses to SARS-CoV-2 within a population may serve as an indicator for resistance to SARS-CoV-2, and the potential need for booster immunization.

The findings revealed associations between responses of SARS-CoV-2-specific $CD4^+$ and $CD8^+$ T-cells and factors including age, days post-vaccination, and NAb (WT). Additionally, a correlation was identified between the levels of NAb (Omicron) and $CD8^+$ T-cell responses. Albeit, the overall correlation was relatively weak. Notably, age demonstrated a statistically significant positive linear correlation with SARS-CoV-2-specific cellular immunity ($P < 0.05$), which might seem counterintuitive considering the increased disease severity and mortality with older age in COVID-19 patients [35, 36]. Besides, Dietz et al. reported that adaptive immune responses, particularly spike-specific responses, decline with age [37]. However, Dietz and colleagues assessed post-vaccination immunity across groups aged > 65, 65–74, and > 75 years. In contrast, our study focused on healthy adults within the 18 to 60 age range, with an average age of 36 years (1st Quartile:30; 3rd Quartile: 42), indicating that the impact of aging within this cohort is quite minimal. What's more, a meta-analysis of 31 lung single-cell RNA-sequencing uncovered cell type-specific connections between age and the expression levels of receptors and proteases crucial for SARS-CoV-2 entry, including TMPRRSS2, ACE2, and CTSL [38]. This provides insight into the increased symptom severity in older COVID-19 patients. Additionally, cross-reactive T-cell immunity against SARS-CoV-2 has been observed in individuals with no prior exposure to the virus, which is thought to be due to the partial homology of T-cell epitopes between SARS-CoV-2 and common cold coronaviruses like HKU1, OC43, 229E, and NL63 [39]. Given the recurrent nature of common cold coronaviruses, elderly individuals may have more frequent encounters with these viruses, which could account for the stronger cellular immune responses post-vaccination observed in the older participants in our study.

Days post-vaccination maintained a linear negative impact on T-cell responses, even after the analysis was adjusted for age and NAb. This finding aligns with numerous studies indicating that SARS-CoV-2-specific T-cell responses decrease as time progresses [40]. Additionally, the nonlinear, 'L'-shaped increase in specific T-cell responses with NAb levels was observed, potentially attributable to the combined effect of $CD4^+$ and $CD8^+$ T-cells in neutralizing antibody production [41, 42]. However, after adjusting for age and days post-vaccination, the correlation between specific T-cell responses and NAb became statistically insignificant. This suggests that the initial 'L'-shaped pattern observed is likely influenced by the confounding effects of time since vaccination and the age of participants. Notably, while the $CD4^+$ T-cell response showed an apparent stabilization after approximately day 100, this nonlinear pattern was not statistically significant (P for nonlinearity = 0.068). This observation may reflect limited sample density at longer follow-up intervals or individual heterogeneity. In summary, the responses of T-cells targeting SARS-CoV-2 were predominantly influenced by age and duration following vaccination, with the impact of NAb being significant in univariate analysis but not after multivariate adjustments.

AI-driven technologies have proven instrumental in refining clinical decision-making processes, accelerating the evolution of pharmaceuticals, enhancing diagnostic procedures for a spectrum of diseases, and strengthening health surveillance systems [43, 44]. ML has adeptly been deployed to craft early alert systems for monitoring emerging SARS-CoV-2 strains [45], pinpoint possible neutralizing agents [46], and predict epitopes for B cells and T-cells that could be targeted in vaccine development [47]. However, the application of ML in predicting the level of cellular immunity post-vaccination remains limited. Such predictive capabilities are crucial for assessing the necessity for booster vaccinations, which are pivotal for reinforcing and maintaining individual and public immunity.

In this study, we collected an array of datasets encompassing epidemiological information, routine blood indices, and immune cell parameters, totaling 32 distinct variables. Through the assessment of the % IncMSE, variables with the most substantial influence were identified. Besides age, days post-vaccination, and NAb, MONO was also identified as a key predictor for the T-cell response, possibly because of their ability to produce mediators that influence T-cell polarization [48]. The proportions of $CD3^+$, $CD4^+$, and $CD8^+$ T-cells were essential for predicting the responses of $CD4^+$ T-cells, potentially owing to their foundational role in driving T-cell reactions. However, the underlying reasons for the importance of the MCV and NEU in predicting $CD8^+$ T-cell responses warrant further investigation. Subsequently, variables were chosen for the predictive model based on their higher R^2 value and lower RMSE values. The model's validation against both training and testing datasets confirmed its predictive efficacy, with a high degree of alignment observed between predicted and actual values. This substantiates the random forest model's high precision in predicting T-cell responses.

In our initial analysis, the model demonstrated a high Pearson's correlation coefficient on the training dataset derived from a 70/30 split. While this result suggests excellent predictive accuracy, it likely reflects overfitting, as the model was evaluated on data it had already seen during training. To obtain a more robust estimate of model performance and generalizability, we applied tenfold cross-validation across the entire dataset. This approach yielded

a lower but more realistic correlation, indicating moderate predictive power when applied to unseen data. The difference between the two results underscores the importance of cross-validation, especially when working with small datasets, to prevent overestimation of model accuracy and ensure more reliable conclusions.

To address the limitations of generalized linear models in capturing complex, nonlinear relationships, we further employed a CART model. This model provides an interpretable, nonlinear approach to quantitatively predict T-cell responses. Utilizing the CART algorithm, we developed visualization models to evaluate the response levels of $CD4^+$ and $CD8^+$ T-cell responses. Notably, the day of the peak in the linear relationship graph for $CD4^+$ T-cells with the days post-vaccination corresponds precisely with the first branching condition in the CART decision tree. This congruence between the two analytical methods underscores the dependability of our data analysis and further supports the validity of our findings.

However, this study also has several limitations. Although the 32 variables included in this study were mostly easily obtained from clinical sources, the number of variables was limited, and the participant's sample size was modest. Moreover, the absence of data from participants with breakthrough infections, coupled with the fact that the maximum duration of vaccination considered for our volunteers was 297 days, and participants who received types other than inactivated ones were not included, may restrict the predictive model's generalizability to those vaccinated for over 297 days, recipients of alternative vaccine types, or individuals with breakthrough infections. Furthermore, our study was conducted within a population cohort from Shenzhen, China. Given the substantial heterogeneity in HLA genotype distribution across different geographical regions, caution should be exercised when extrapolating these findings to a global context [49, 50].

While our model demonstrates strong predictive performance through internal validation, external validation remains an important goal. However, our predictive framework is based on a multi-dimensional feature set—including not only age and time post-vaccination, but also detailed immunological markers (e.g., $CD3^+$, $CD4^+$, and $CD8^+$ T-cell percentages), neutralizing antibody levels (WT and Omicron), and routine clinical laboratory indicators (e.g., MONO, NEU, and MCV). To date, few external datasets provide this full set of predictors alongside matched T cell response outcomes. Nevertheless, future work will aim to identify partial external cohorts to validate model components and further test generalizability.

Despite these limitations, we proposed an economical and efficient model capable of evaluating the levels of SARS-CoV-2-specific T-cell responses following vaccination. This approach requires no sophisticated laboratory facilities, thereby expanding its applicability for evaluating SARS-CoV-2 immune status and assessing the need for booster vaccination across a range of healthcare settings, from large hospitals to local community clinics. While the direct applicability of our model is limited by the specificity of our cohort, it provides a foundation for future research. By incorporating more diverse populations and additional data sources, future models can address the global circulation of SARS-CoV-2 and its variants, enhancing the generalizability and applicability of immune response prediction. Moreover, these methodologies are well-suited for multicenter studies and may prove instrumental in predicting adaptive immunity after vaccination with a variety of vaccines targeting various pathogens. Additionally, the framework of our model, which currently focuses on predicting T-cell responses-a clinically challenging measurement-could be adapted to forecast antibody dynamics using similar input features. This adaption could potentially enable low-cost monitoring of humoral immunity in rescore-limited settings.

Conclusions

In this research, multivariate analysis revealed that age and the number of days post-vaccination were the most significant factors affecting the responses of SARS-CoV-2-specific T-cells. Utilizing the RF and CART algorithms, we have crafted predictive models that employ readily available data, including age, days post-vaccination, NAb, and routine blood parameters, to predict the levels of T-cell responses with remarkable precision and efficiency. This visualizable and easy-to-use methodology also has the potential for application in the evaluation of other vaccines and offers clinical utility.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10238-025-01772-2.

Acknowledgements We would like to thank all of the volunteers in the study.

Author Contributions Jie Ning and Ying Chen designed this study and drafted the manuscript. Ying Chen also played a key role in analyzing the data. The recruitment of participants for the study was managed by Jie Ning, Yayi Ren, Zelin Zhang, Xianhuang Zeng, Qinjin Wang, Jia Xie, Yue Xu, Yali Fan, and Huilan Li. Aixia Zhai, Bin Li, and Chao Wu refined the article for intellectual content and accuracy. The final review and editing of the manuscript, obtaining funding for the project, and overall supervision of the study were accomplished by Chao Wu, Jie Ning, and Ying Chen.

Funding This project was supported by the China postdoctoral Science Foundation under Grant Number 2024M763770, the Postdoctoral Fellowship Program of CPSF under Grant Number GZC20233232, the National Natural Science Foundation of China (72204276), the Shenzhen Science and Technology Innovation Program

(JCYJ20210324115204012), and the Futian Healthcare Research Project (FTWS2023023).

Data availability The dataset generated from this study is publicly available in the Harvard Data verse repository. It is titled "Predicting SARS-CoV-2-specific CD4 + and CD8 + T-cell responses elicited by inactivated vaccines in healthy adults using machine learning models" and can be accessed at https://doi.org/10.7910/DVN/1GZHO4.

Declarations

Conflicts of interest The authors declare no competing interests.

Ethical approval This research was approved by the Ethics Review Committee of the Eighth Affiliated Hospital of Sun Yat-sen University, with the reference number 2021–005-01, and it was carried out under the principles of the 1975 Declaration of Helsinki.

Informed consent Written informed consent was obtained directly from each participant involved in the study.

References

1. Steiner S, Kratzel A, Barut GT, Lang RM, Aguiar Moreira E, Thomann L, et al. SARS-CoV-2 biology and host interactions. Nat Rev Microbiol. 2024;22(4):206–25. https://doi.org/10.1038/s41579-023-01003-z.
2. WHO. WHO COVID-19 dashboard [updated 2025.3.14. Available from: https://data.who.int/dashboards/covid19/deaths?n=c.
3. Zhu C, Pang S, Liu J, Duan Q. Current progress, challenges and prospects in the development of COVID-19 vaccines. Drugs. 2024;84(4):403–23. https://doi.org/10.1007/s40265-024-02013-8.
4. Zhang L, Kempf A, Nehlmeier I, Cossmann A, Richter A, Bdeir N, et al. SARS-CoV-2 BA.2.86 enters lung cells and evades neutralizing antibodies with high efficiency. Cell. 2024;187:596.
5. Callaway E. Beyond Omicron: what's next for COVID's viral evolution. Nature. 2021;600(7888):204–7. https://doi.org/10.1038/d41586-021-03619-8.
6. Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. Cell. 2021;184(4):861–80. https://doi.org/10.1016/j.cell.2021.01.007.
7. Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. Science. 2021. https://doi.org/10.1126/science.abf4063.
8. DeWolf S, Laracy JC, Perales MA, Kamboj M, van den Brink MRM, Vardhana S. SARS-CoV-2 in immunocompromised individuals. Immunity. 2022;55(10):1779–98. https://doi.org/10.1016/j.immuni.2022.09.006.
9. Koerber N, Priller A, Yazici S, Bauer T, Cheng CC, Mijočević H, et al. Dynamics of spike-and nucleocapsid specific immunity during long-term follow-up and vaccination of SARS-CoV-2 convalescents. Nat Commun. 2022;13(1):153. https://doi.org/10.1038/s41467-021-27649-y.
10. Jung JH, Rha MS, Sa M, Choi HK, Jeon JH, Seok H, et al. SARS-CoV-2-specific T cell memory is sustained in COVID-19 convalescent patients for 10 months with successful development of stem cell-like memory T cells. Nat Commun. 2021;12(1):4043. https://doi.org/10.1038/s41467-021-24377-1.
11. Garcia-Beltran WF, Lam EC, St Denis K, Nitido AD, Garcia ZH, Hauser BM, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell. 2021. https://doi.org/10.1016/j.cell.2021.03.013.
12. Uriu K, Kimura I, Shirakawa K, Takaori-Kondo A, Nakada TA, Kaneda A, et al. Neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine serum. N Engl J Med. 2021;385(25):2397–9. https://doi.org/10.1056/NEJMc2114706.
13. Meyer S, Blaas I, Bollineni RC, Delic-Sarac M, Tran TT, Knetter C, et al. Prevalent and immunodominant CD8 T cell epitopes are conserved in SARS-CoV-2 variants. Cell Rep. 2023;42(1):111995. https://doi.org/10.1016/j.celrep.2023.111995.
14. Keeton R, Tincho MB, Ngomti A, Baguma R, Benede N, Suzuki A, et al. T cell responses to SARS-CoV-2 spike cross-recognize Omicron. Nature. 2022;603(7901):488–92. https://doi.org/10.1038/s41586-022-04460-3.
15. Yang M, Meng Y, Hao W, Zhang J, Liu J, Wu L, et al. A prognostic model for SARS-CoV-2 breakthrough infection: analyzing a prospective cellular immunity cohort. Int Immunopharmacol. 2024;131:111829. https://doi.org/10.1016/j.intimp.2024.111829.
16. Ong DSY, Fragkou PC, Schweitzer VA, Chemaly RF, Moschopoulos CD, Skevaki C. How to interpret and use COVID-19 serology and immunology tests. Clin Microbiol Infect. 2021;27(7):981–6. https://doi.org/10.1016/j.cmi.2021.05.001.
17. Wang C, Yang X, Gu B, Liu H, Zhou Z, Shi L, et al. Sensitive and simultaneous detection of SARS-CoV-2-specific IgM/IgG using lateral flow immunoassay based on dual-mode quantum dot nanobeads. Anal Chem. 2020;92(23):15542–9. https://doi.org/10.1021/acs.analchem.0c03484.
18. Roda A, Cavalera S, Di Nardo F, Calabria D, Rosati S, Simoni P, et al. Dual lateral flow optical/chemiluminescence immunosensors for the rapid detection of salivary and serum IgA in patients with COVID-19 disease. Biosens Bioelectron. 2021;172:112765. https://doi.org/10.1016/j.bios.2020.112765.
19. Slota M, Lim JB, Dang Y, Disis ML. ELISpot for measuring human immune responses to vaccines. Expert Rev Vaccines. 2011;10(3):299–306. https://doi.org/10.1586/erv.10.169.
20. Ning J, Wang Q, Chen Y, He T, Zhang F, Chen X, et al. Immunodominant SARS-CoV-2-specific CD4(+) and CD8(+) T-cell responses elicited by inactivated vaccines in healthy adults. J Med Virol. 2023;95(4):e28743. https://doi.org/10.1002/jmv.28743.
21. Freer G, Rindi L. Intracellular cytokine detection by fluorescence-activated flow cytometry: basic principles and recent advances. Methods. 2013;61(1):30–8. https://doi.org/10.1016/j.ymeth.2013.03.035.
22. Fung CYJ, Scott M, Lerner-Ellis J, Taher J. Applications of SARS-CoV-2 serological testing: impact of test performance, sample matrices, and patient characteristics. Crit Rev Clin Lab Sci. 2024;61(1):70–88. https://doi.org/10.1080/10408363.2023.2254390.
23. Kahn R, Schrag SJ, Verani JR, Lipsitch M. Identifying and alleviating bias due to differential depletion of susceptible people in

postmarketing evaluations of COVID-19 vaccines. Am J Epidemiol. 2022;191(5):800–11. https://doi.org/10.1093/aje/kwac015.
24. Chang CH, Lin CH, Lane HY. Machine learning and novel biomarkers for the diagnosis of Alzheimer's disease. Int J Mol Sci. 2021. https://doi.org/10.3390/ijms22052761.
25. Dritsas E, Trigka M. Supervised machine learning models to identify early-stage symptoms of SARS-CoV-2. Sensors. 2022. https://doi.org/10.3390/s23010040.
26. Rehman MU, Naseem S, Butt AUR, Mahmood T, Khan AR, Khan I, et al. Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment. Sci Rep. 2025;15(1):13361. https://doi.org/10.1038/s41598-025-96437-1.
27. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci. 2017;9:329. https://doi.org/10.3389/fnagi.2017.00329.
28. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. Front Public Health. 2020;8:357. https://doi.org/10.3389/fpubh.2020.00357.
29. Danieli MG, Paladini A, Longhi E, Tonacci A, Gangemi S. A machine learning analysis to evaluate the outcome measures in inflammatory myopathies. Autoimmun Rev. 2023;22(7):103353. https://doi.org/10.1016/j.autrev.2023.103353.
30. Zimmerman RK, Nowalk MP, Bear T, Taber R, Clarke KS, Sax TM, et al. Proposed clinical indicators for efficient screening and testing for COVID-19 infection using classification and regression trees (CART) analysis. Hum Vaccin Immunother. 2021;17(4):1109–12. https://doi.org/10.1080/21645515.2020.1822135.
31. Dicu T, Cucoş A, Botoş M, Burghele B, Florică Ş, Baciu C, et al. Exploring statistical and machine learning techniques to identify factors influencing indoor radon concentration. Sci Total Environ. 2023;905:167024. https://doi.org/10.1016/j.scitotenv.2023.167024.
32. Chen Y, Chen X, Liang Z, Fan S, Gao X, Jia H, et al. Epidemiology and prediction of multidrug-resistant bacteria based on hospital level. J Glob Antimicrob Resist. 2022;29:155–62. https://doi.org/10.1016/j.jgar.2022.03.003.
33. Vályi-Nagy I, Matula Z, Gönczi M, Tasnády S, Bekő G, Réti M, et al. Comparison of antibody and T cell responses elicited by BBIBP-CorV (Sinopharm) and BNT162b2 (Pfizer-BioNTech) vaccines against SARS-CoV-2 in healthy adult humans. Geroscience. 2021;43(5):2321–31. https://doi.org/10.1007/s11357-021-00471-6.
34. Nesamari R, Omondi MA, Baguma R, Höft MA, Ngomti A, Nkayi AA, et al. Post-pandemic memory T cell response to SARS-CoV-2 is durable, broadly targeted, and cross-reactive to the hypermutated BA.2.86 variant. Cell Host Microbe. 2024. https://doi.org/10.1016/j.chom.2023.12.003.
35. Perez-Saez J, Lauer SA, Kaiser L, Regard S, Delaporte E, Guessous I, et al. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. Lancet Infect Dis. 2021;21(4):e69–70. https://doi.org/10.1016/s1473-3099(20)30584-3.
36. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395(10229):1054–62. https://doi.org/10.1016/s0140-6736(20)30566-3.
37. Dietz LL, Juhl AK, Søgaard OS, Reekie J, Nielsen H, Johansen IS, et al. Impact of age and comorbidities on SARS-CoV-2 vaccine-induced T cell immunity. Commun Med. 2023;3(1): 58. https://doi.org/10.1038/s43856-023-00277-x.
38. Muus C, Luecken MD, Eraslan G, Sikkema L, Waghray A, Heimberg G, et al. Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. Nat Med. 2021;27(3):546–59. https://doi.org/10.1038/s41591-020-01227-z.
39. Sette A, Crotty S. Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. Nat Rev Immunol. 2020;20(8):457–8. https://doi.org/10.1038/s41577-020-0389-z.
40. Moss P. The T cell immune response against SARS-CoV-2. Nat Immunol. 2022;23(2):186–93. https://doi.org/10.1038/s41590-021-01122-w.
41. Law H, Venturi V, Kelleher A, Munier CML. Tfh cells in health and immunity: potential targets for systems biology approaches to vaccination. Int J Mol Sci. 2020. https://doi.org/10.3390/ijms21228524.
42. Elzein SM, Zimmerer JM, Han JL, Ringwald BA, Bumgardner GL. CXCR5(+)CD8(+) T cells: a review of their antibody regulatory functions and clinical correlations. J Immunol. 2021;206(12):2775–83. https://doi.org/10.4049/jimmunol.2100082.
43. Réda C, Kaufmann E, Delahaye-Duriez A. Machine learning applications in drug development. Comput Struct Biotechnol J. 2020;18:241–52. https://doi.org/10.1016/j.csbj.2019.12.006.
44. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. npj Digit Med. 2019;2:69. https://doi.org/10.1038/s41746-019-0148-3.
45. Telenti A, Hodcroft EB, Robertson DL. The evolution and biology of SARS-CoV-2 variants. Cold Spring Harb Perspect Med. 2022. https://doi.org/10.1101/cshperspect.a041390.
46. Elkashlan M, Ahmad RM, Hajar M, Al Jasmi F, Corchado JM, Nasarudin NA, et al. A review of SARS-CoV-2 drug repurposing: databases and machine learning models. Front Pharmacol. 2023;14:1182465. https://doi.org/10.3389/fphar.2023.1182465.
47. Bukhari SNH, Jain A, Haq E, Mehbodniya A, Webber J. Machine learning techniques for the prediction of B-cell and T-cell epitopes as potential vaccine targets with a specific focus on SARS-CoV-2 pathogen: a review. Pathogens. 2022. https://doi.org/10.3390/pathogens11020146.
48. Schreiber HA, Loschko J, Karssemeijer RA, Escolano A, Meredith MM, Mucida D, et al. Intestinal monocytes and macrophages are required for T cell polarization in response to Citrobacter rodentium. J Exp Med. 2013;210(10):2025–39. https://doi.org/10.1084/jem.20130903.
49. Liu J, Quan ZR, Zhu TH, Zhong YP, Jiang RH, Yang BN, et al. Allele and haplotype frequencies of 17 HLA-related loci in Shenzhen Chinese population by next-generation sequencing. Hla. 2025;105(4):e70148. https://doi.org/10.1111/tan.70148.
50. Liu S, Li Y, Song T, Zhang J, Zhang P, Luo H, et al. The pathogen adaptation of HLA alleles and the correlation with Autoimmune Diseases in the Han Chinese. Genomics Proteomics Bioinform. 2025. https://doi.org/10.1093/gpbjnl/qzaf038.