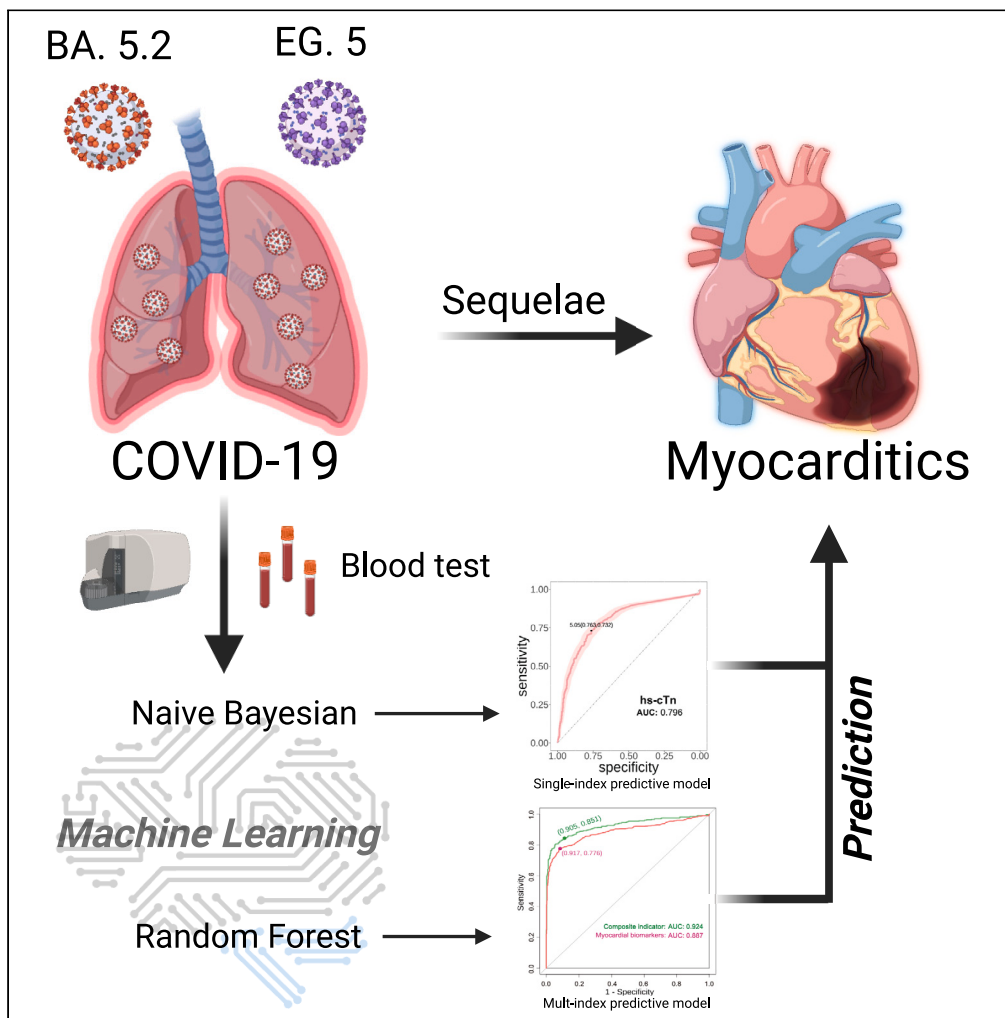


Article

Decoding acute myocarditis in patients with COVID-19: Early detection through machine learning and hematological indices



Haiyang Li,
Zhangkai J. Cheng,
Xing Fu, ..., Zhiman
Liang, Fei Wang,
Baoqing Sun

haiyang.li@mrc-bsu.cam.ac.uk (H.L.)
wangfei0163@163.com (F.W.)
sunbaoqing@vip.163.com (B.S.)

Highlights

Younger people face higher myocarditis risk

Omicron EG.5 variant has higher myocarditis risk than BA.5.2

Multi-index model (AUC 0.924) improves early myocarditis detection

COVID-19 shows unique markers; low ALB/GLB in severe cases helps stratify



Article

Decoding acute myocarditis in patients with COVID-19: Early detection through machine learning and hematological indices

Haiyang Li,^{1,4,5,*} Zhangkai J. Cheng,^{1,5} Xing Fu,³ Mingtao Liu,¹ Peng Liu,² Wenhan Cao,¹ Zhiman Liang,¹ Fei Wang,^{2,*} and Baoqing Sun^{1,6,*}

SUMMARY

During the persistent COVID-19 pandemic, the swift progression of acute myocarditis has emerged as a profound concern due to its augmented mortality, underscoring the urgency of prompt diagnosis. This study analyzed blood samples from 5,230 COVID-19 individuals, identifying key blood and myocardial markers that illuminate the relationship between COVID-19 severity and myocarditis. A predictive model, applying Bayesian and random forest methodologies, was constructed for myocarditis' early identification, unveiling a balanced gender distribution in myocarditis cases contrary to a male predominance in COVID-19 occurrences. Particularly, older men exhibited heightened vulnerability to severe COVID-19 strains. The analysis revealed myocarditis was notably prevalent in younger demographics, and two sub-variants COVID-19 progression paths were identified, characterized by symptom intensity and specific blood indicators. The enhanced myocardial marker model displayed remarkable diagnostic accuracy, advocating its valuable application in future myocarditis detection and treatment strategies amidst the COVID-19 crisis.

INTRODUCTION

The coronavirus disease 2019 (COVID-19) has swept almost the world's population. It is a pandemic infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2).^{1,2} As the disease progresses, it can lead to a range of symptoms and complications.³ It will also produce different degrees of sequela after the patient recovers.^{4,5} Among them, there is now increasing evidence that acute viral myocarditis is a highly prevalent and fatal sequela of COVID-19,⁶ especially in young adult and adolescent males. In the coming years, we may also be affected by COVID-19 in the long term. Therefore, predicting myocarditis in patients is particularly important for the health of people and even adolescents.

Acute myocarditis is a recognized complication of acute viral infection⁷ which can have serious consequences, including heart failure, arrhythmias, dilated cardiomyopathy, sudden cardiac death, and long-term complications. Relevant studies on prognostic indicators of myocarditis include cardiac troponin-I and homocysteine.⁸ Prompt diagnosis and treatment are important to minimize the harm caused by this condition. Hendrickson et al.⁹ conducted a follow-up investigation on the athletes infected with COVID-19 and found that nearly 50% of the athletes had cardiac function defects in the prognosis. However, other studies have suggested that the incidence of myocarditis may be higher in certain subgroups of patients. Modin et al.¹⁰ points out that 15% had evidence of myocardial injury, which can sometimes lead to myocarditis. It is important to note that the available data on myocarditis in patients with COVID-19 is still evolving, and more research is needed to understand the risks and outcomes associated with this complication fully. Additionally, factors such as age, sex, and underlying medical conditions may play a role in the incidence of myocarditis in patients with COVID-19.¹¹

Significantly, acute myocarditis and COVID-19 have been shown to affect the blood routine index, including white blood cell count, red blood cell count, and platelet count. Y. Liu, Y. Yang et al.¹² found that patients with COVID-19 have lower lymphocyte and platelet levels than healthy controls. The findings suggest that these blood cell count abnormalities may be associated with the severity of the disease and could potentially serve as prognostic indicators. Terpos et al.¹³ analyzed the hematological parameters of patients with COVID-19 and found that

¹Department of Clinical Laboratory, National Clinical Research Center of Respiratory Disease, Guangzhou Institute of Respiratory Health, First Affiliated Hospital of Guangzhou Medical University, Guangzhou Medical University, Guangzhou 510120, China

²Department of Clinical Pharmacy, Dazhou Central Hospital, Dazhou 635000, China

³Group of Theoretical Biology, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

⁴MRC Biostatistics Unit, University of Cambridge, Cambridge CB2 0SR, UK

⁵These authors contributed equally

⁶Lead contact

*Correspondence: haiyang.li@mrc-bsu.cam.ac.uk (H.L.), wangfei0163@163.com (F.W.), sunbaoqing@vip.163.com (B.S.)
<https://doi.org/10.1016/j.isci.2023.108524>



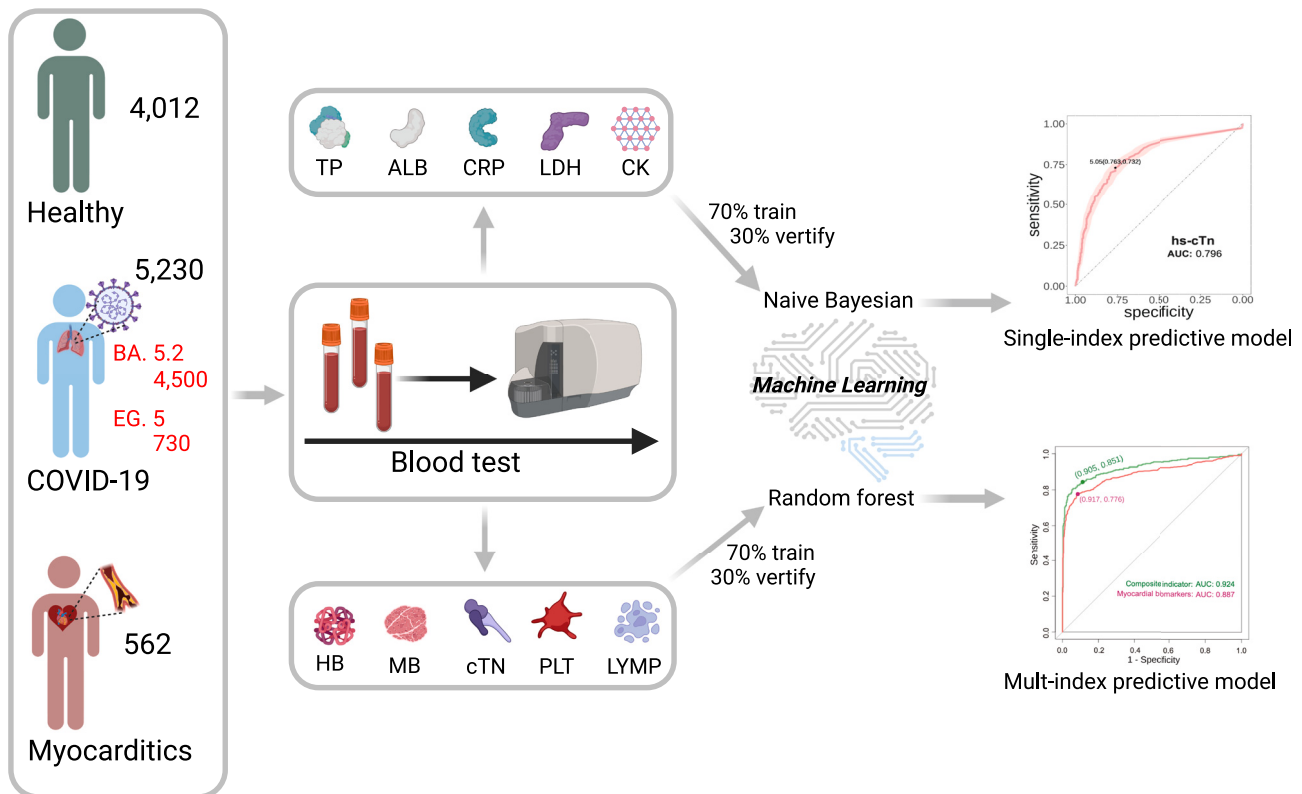


Figure 1. The scheme of analysis and prediction of indexes of COVID-19 and myocarditis

The blood samples were collected from a diverse group of individuals, including 5,230 patients diagnosed with COVID-19 (4500 BA.5.2 & 730 EG.5), 562 individuals suffering from myocarditis, and approximately 4,012 healthy individuals. These samples were then analyzed to identify various biochemical indicators and biomarkers. The data derived from this analysis was subsequently input into a statistical model. This model was designed to systematically predict both the diagnosis and prognosis of myocarditis, providing a potentially powerful tool in the fight against this condition.

they had higher levels of inflammatory markers, such as C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), as well as lower levels of hemoglobin and red blood cells compared to healthy controls. They also observed that patients with severe COVID-19 had higher CRP and ESR levels and lower hemoglobin and red blood cell counts than those with mild cases. The findings suggest that these hematological abnormalities may be associated with the severity of COVID-19 infection and could potentially serve as prognostic indicators. These changes in blood parameters may be related to the inflammatory response to the virus and can help to identify patients with severe disease.¹⁴ At present, much work has been devoted to finding relevant markers and patient indicators to study COVID-19 and predict the Sequelae of COVID-19.¹⁵ Ullah et al.¹⁶ studied the predictive correlation between the lymphocyte-to-neutrophil ratio (LNR) in patients with COVID-19 and found that high LNR had a higher probability of predicting in-hospital mortality. Tamaki et al.¹⁷ investigated the indicators of 1026 patients with heart failure and conducted follow-ups. These results verify the combined role of LNR and lymphocyte-to-platelet ratio (LPR) in evaluating the prognosis of patients with acute heart failure. Along with albumin¹⁸ peripheral lymphocyte count^{19–21} lymphocyte to neutrophil ratio²² their research proved changes in blood routine indicators are commonly observed in patients with COVID-19. It may provide important diagnostic and prognostic information. However, it is essential to note that these changes may vary among patients and should be interpreted in the context of the patient's overall clinical picture. Further research is needed to understand the mechanisms underlying these changes fully and to develop more effective treatments for COVID-19.

Here, we collected blood samples from 5,230 patients with COVID-19, 4,012 healthy individuals, and 562 patients with myocarditis recovered from COVID-19. We measured patients' blood indicators, biochemical indicators and related markers and then further compared the differences between patients with COVID-19 and myocarditis. We trained the classification predictive model according to the relevant data to explore how patient indicators change to develop myocarditis. Figure 1 shows the experimental and control groups' routine biochemical and specific indicators. Then, a naive Bayesian model and random forest (RF) machine learning algorithm were applied to the training of index data of sample blood, to construct receiver operating characteristic (ROC) spectrum of subjects, and to calculate the area under the curve (AUC) to distinguish the condition of regular and patients with myocarditis. Further correlation analysis was conducted between COVID-19 and the indicators of patients with myocarditis to predict the patients with COVID-19 at risk of myocarditis, thereby improving the survival rate of patients. At the same time, we performed a thorough analysis of the correlation and difference of indicators, as well as the distribution of ratios of severe indicators.

Table 1. The sample information and blood indicator of healthy population, patients with COVID-19, and myocarditis

	Healthy	COVID-19 (BA.5.2)	Myocarditis	p value (COVID-19 VS Myoca.)
Male	1976 (49.3%)	2810 (62.4%)	275 (48.9%)	
Female	2036 (50.7%)	1690 (37.6%)	275 (51.1%)	
Age	35 (27,49)	70 (56,81)	35 (22,56)	<0.001
TP	76 (72,79)	65 (59,74)	65 (60,71)	0.5
ALB	44 (43,46)	36 (32,42)	37 (32,40)	0.7
hs-CRP	1 (1,2)	19 (5,53)	5 (1,27)	<0.001
CRP	0.3 (0.1,0.7)	0.3 (0.1,1.8)	2.3 (0.6,8.1)	<0.001
LDH	166 (145,190)	256 (200,360)	217 (168,329)	<0.001
CK	76 (78,145)	73 (41,130)	90 (55,161)	<0.001
HB	142 (129,153)	113 (91,131)	120 (99,133)	<0.001
MB	19 (15,22)	36 (20,92)	33 (8,51)	<0.001
hs-cTn	2 (2,3)	8 (3,21)	27 (3,78)	<0.001
PLT	241 (206,291)	218 (154,286)	230 (165,298)	0.008
LYMP%	33 (28,39)	13 (6,25)	19 (9,29)	<0.001

Interquartile range and t-test, where a statistically significant difference is p value shown in table.

RESULTS

The demographic characteristics of all samples

In this study, the gender ratio of males to females in both the patients with myocarditis and the healthy population is approximately balanced at 1:1. The average age of the healthy population is 38 years, with a standard deviation of 20 years. The age range for patients with myocarditis is 39 ± 22 years. Both groups share the same median age of 35 years within the IQR. In contrast, among patients with COVID-19, the male-to-female ratio is 2810:1690 (Figure 2A). The average age for this group is higher, at 65 years, with a standard deviation of 20 years. Moreover, the median age within the IQR is also higher, at 70 years. Females account for 37.6% of patients with COVID-19, a proportion notably lower than that of males. This contrasts with the myocarditis and healthy groups, where the gender ratio is essentially balanced. Table 1 presents the detailed demographic characteristic information. It is worth noting that the age distribution among the healthy population is more uniform across age ranges. In contrast, the age distribution among patients with COVID-19 is more concentrated in the older age group (60–80 years). Conversely, patients with myocarditis are primarily concentrated in the younger age group (20–40 years), as shown in Figure 2B.

The levels of various blood indicators exhibited significant variations between groups, as detailed in Table 1, showing the main indicators are the results of the quartile data, presented in the form of the median (IQR: 25%, 75%). The p value represents the Kruskal-Wallis test between patients with COVID-19 and myocarditis. The comprehensive indicators for each group demonstrate extreme statistical significance, with only TP and ALB not showing statistical significance. Furthermore, each index's corresponding values display the indicators' average value and fluctuation range (interquartile range) in different groups. Among all indicators, the indicator level of healthy people is significantly different from that of patients with COVID-19 and myocarditis. However, it can be challenging to fully comprehend the relationships between data through numerical values in Table 1 alone. Therefore, in the following analysis, we plan to conduct a more in-depth analysis of various indicators using different statistical methods.

Statistical analysis of blood protein indicators

As a prospective biomarker, the albumin/globulin (ALB/GLB) ratio (AGR) could be instrumental in risk categorization and management of patients with COVID-19. Prompt identification of patients exhibiting a lower AGR could enable early potentially mitigating or managing severe disease progression. Additionally, tracking the AGR throughout the disease's duration could offer valuable insights into the patient's response to treatment and the path of recovery. The AGR essentially mirrors the balance between the body's inflammatory and anti-inflammatory responses. A diminished AGR signifies an increased the production of globulins, specifically immunoglobulins and acute-phase proteins, in reaction to infection and inflammation. This amplified immune response can trigger severe inflammation and tissue damage in patients with COVID-19, leading to poorer clinical outcomes. Furthermore, ALB, a negative acute-phase protein, decreases its concentration during infection and inflammation. A decreased ALB level could be indicative of malnutrition, which can exacerbate outcomes in patients with COVID-19. Therefore, the AGR can serve as an indicator of both nutritional status and disease severity.

The AGR with a cutoff value of 1.2 has been identified as a highly sensitive combination marker.²³ This is visually represented in Figure 3A–3C, where the red line denotes an AGR of 1.2. The blue line, fitted using loess regression, provides a fundamental representation of the two-dimensional data distribution. Notably, the scatter points for the healthy population predominantly lie above the AGR = 1.2 line, while a more significant proportion of scatter points for patients with COVID-19 and myocarditis are found below this line, this suggests a clear

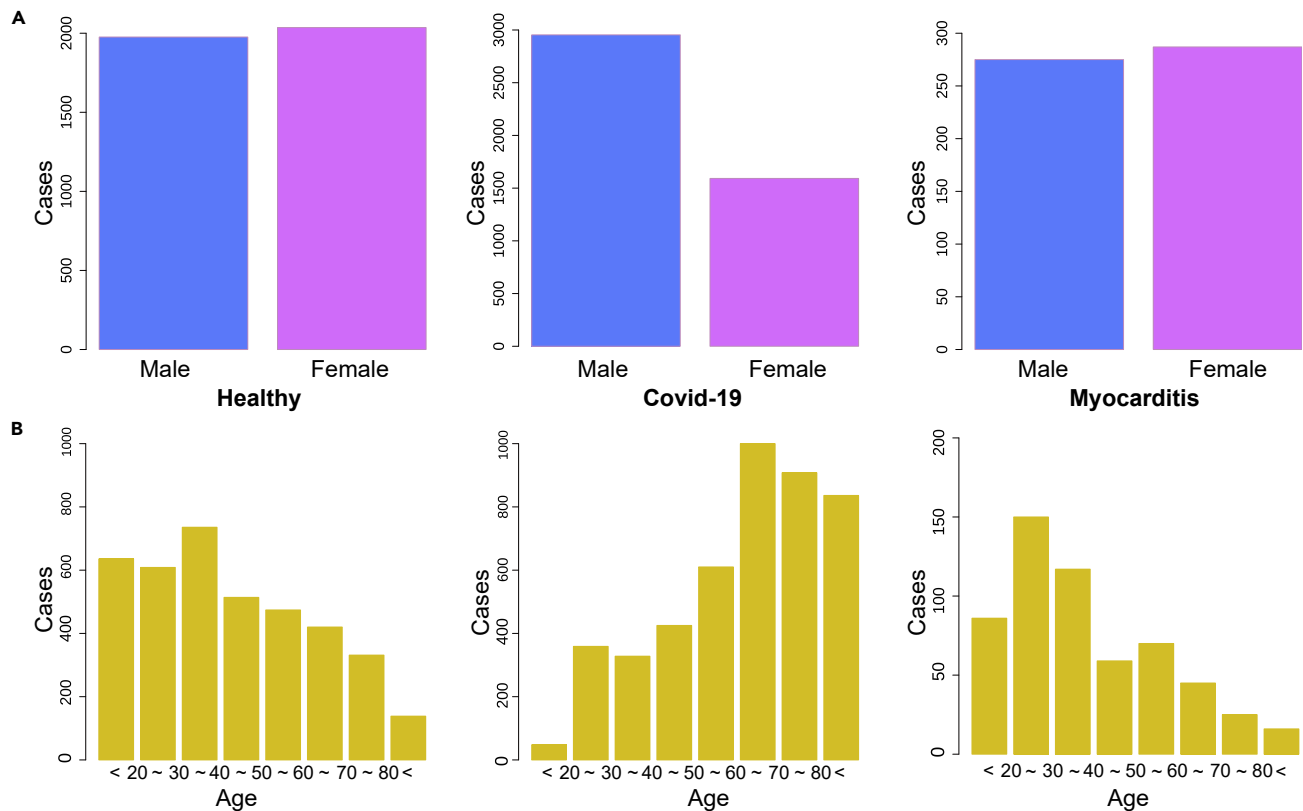


Figure 2. Gender and age distribution of the study sample

(A) The sexual distribution of the healthy population, patients with COVID-19 (BA.5.2), and myocarditis. It displayed the number of men and women in each participating population.

(B) The age distribution of healthy population, patients with COVID-19 (BA.5.2), and myocarditis. Statistics the number of people in each age group.

differentiation in the AGR between healthy individuals and those with COVID-19 or myocarditis, which could serve as a valuable tool for early detection and intervention. These scatter points are presented as a heatmap, where the red area indicates concentrated points of the indicator and the purple area represents scattered points. The proportion of asymptomatic youth was higher, and the results showed higher TP indicators (Figure S1). This visual representation provides an intuitive way to understand the distribution of these indicators, which could be helpful in both clinical and research settings. The peripheral bar chart complements this by visually representing the concentration of scatter points in the relative area. This allows for a straightforward comparison of the density of data points across different ranges of AGR. Interestingly, the hotspot center, which represents the area with the highest concentration of data points, aligns with the peak distribution of the bar chart and the boxplot located in the red area. This consistency across different visual representations reinforces the reliability of these findings.

Upon analyzing the results depicted in Figure 3D, it was observed that the ALB levels in patients with COVID-19 and myocarditis were notably lower compared to those in the healthy population. Interestingly, the ALB levels in patients with COVID-19 and myocarditis were virtually identical, with no statistical significance between them. Concurrently, it was noted that the trends in TP and ALB levels were in line with these findings. The globulin level appeared to be unaffected by the disease, with all three groups maintaining a consistent level. Subsequently, upon comparing indicators between male and female samples, we found no such statistically significant was observed in ALB, GLB, and TP indicators among patients with myocarditis of both genders. Similarly, TP levels showed no significant variation between healthy men and women. Notably, the gender-based difference in protein levels was most pronounced in patients with COVID-19.

The risk factors in COVID-19 and myocarditis

The lymphocyte-to-neutrophil ratio (LNR) has been identified as an independent risk factor for mortality in hospitalized patients with COVID-19.²⁴ This ratio, calculated from routine blood tests, measures systemic inflammation and immune response. Several studies have found that patients with severe COVID-19 often have a lower LNR compared to those with mild disease. A lower LNR may have been associated with more severe disease. In some patients with severe COVID-19, there can be a change in the ratio of lymphocytes to neutrophils, often with an increase in neutrophils and a decrease in lymphocytes. This is often referred to as the lymphocyte-to-neutrophil ratio (LNR). This is likely due to the increased neutrophil count, a response to infection and inflammation, and decreased lymphocyte count, which could result from

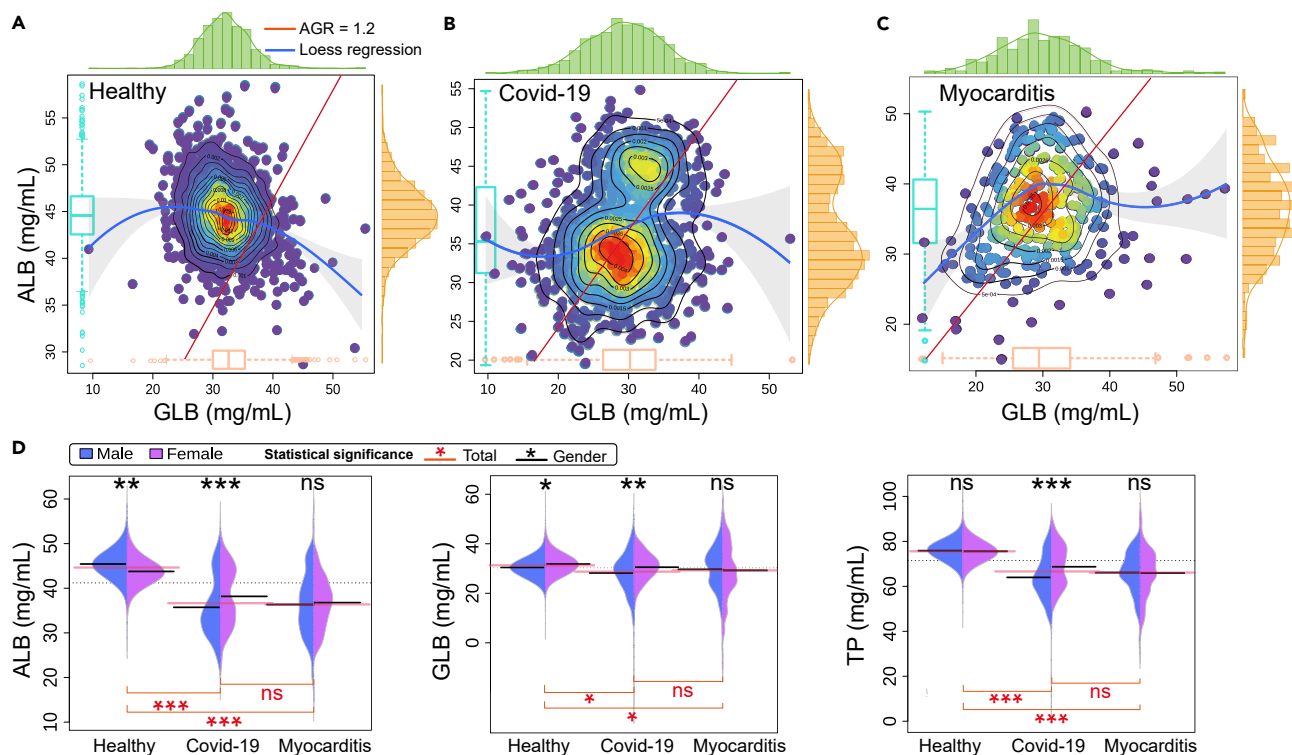


Figure 3. The distribution of ALB and GLB with statistically significant analysis

(A) The distribution of ALB/GLB ratios in a healthy population. The red line represents the threshold with an AGR of 1.2. The boxplot indicates the mean \pm SD, and the blue line depicts the loess regression across all data points, with the surrounding shaded area representing the confidence interval.
 (B) The ALB/GLB distribution of COVID-19. The red line is the dividing line of AGR = 1.2. The boxplot indicates the mean \pm SD, and the blue line is the loess regression at all points, the shadow is the confidence interval.
 (C) The ALB/GLB distribution of patients with myocarditis. The red line is the dividing line of AGR = 1.2. The boxplot indicates the mean \pm SD, and the blue line is the loess regression at all points, the shadow is the confidence interval.
 (D) The ALB, GLB, TP distribution and statistical significance of healthy population, patients with COVID-19, and myocarditis. Interquartile range and t-test, where a statistically significant difference is denoted by *** ($p < 0.001$), ** ($p < 0.01$), and * ($p < 0.05$), and non-significance ($p > 0.05$) is denoted by "ns".

viral-induced lymphocyte apoptosis or redistribution. In the scatter analysis of neutrophils and lymphocytes, the sample distribution showed a good linear relationship and a linear model fitted the relevant data. The results were shown in Figures 4A–4C, and the healthy population showed a good fit. The R2 of patients with COVID-19 and myocarditis decreased successively, which was mainly due to the disease-causing the spread of scatter distribution. After linear fitting, we obtained the slope of the fitting equation, with health being -0.83 , COVID-19 being -0.82 , and myocarditis being -0.74 . The slope of the corresponding law is an increasing relationship. L/N shows a rising disease trend and is more prevalent in patients with more severe myocarditis. Through the scattered density heat analysis, it can be found that the healthy people are mainly concentrated near the midpoint of the ratio. In contrast, the density hot spots of COVID-19 and myocarditis are concentrated in the low lymphocyte area, myocarditis has two density concentration areas, and the bar chart also shows two peaks.

The results of single indicator analysis showed indicators such as NEUTP (Figure 4D), PLT (Figure 4E), and LYMPHP (Figure 4F) between men and women. Only the LYMPHP of patients with COVID-19 showed statistical significance, while the results of other indicators did not show statistical significance. In the three groups, the three indicators of patients with COVID-19 and myocarditis basically did not have statistical significance. In comparison, the healthy people and these two groups had extremely high statistical significance. Early studies suggested that the LPR could be a valuable tool in predicting the severity of COVID-19. Patients with severe forms of the disease appeared to have a lower LPR compared to those with milder forms.²⁵ These ratios, often referred to as the LPR, can serve as potential biomarkers for disease prognosis and treatment response. Figures 4G–4I, In the context of COVID-19, (Figure 4H), a decrease in lymphocyte count (lymphopenia) and an increase in platelet count (thrombocytosis) have been frequently observed. This imbalance, reflected in a low LPR, could indicate a severe inflammatory response and hypercoagulable state. However, while the LNR is a promising tool, it is not without limitations. It can be influenced by various factors, including age, sex, and comorbidities, and it is not specific to COVID-19. Therefore, it should be used in conjunction with other clinical parameters and biomarkers to assess the risk of severe disease and mortality in patients with COVID-19.

Our analysis highlights the concurrence between the number of lymphocytes and the LYMPHP distribution (Figure S2A). Moreover, the hemoglobin (HB) index in healthy individuals is markedly higher than that observed in patients suffering from COVID-19 and myocarditis. This disparity offers crucial insight into the pathophysiology of these conditions and their impact on normal body functioning. Equally

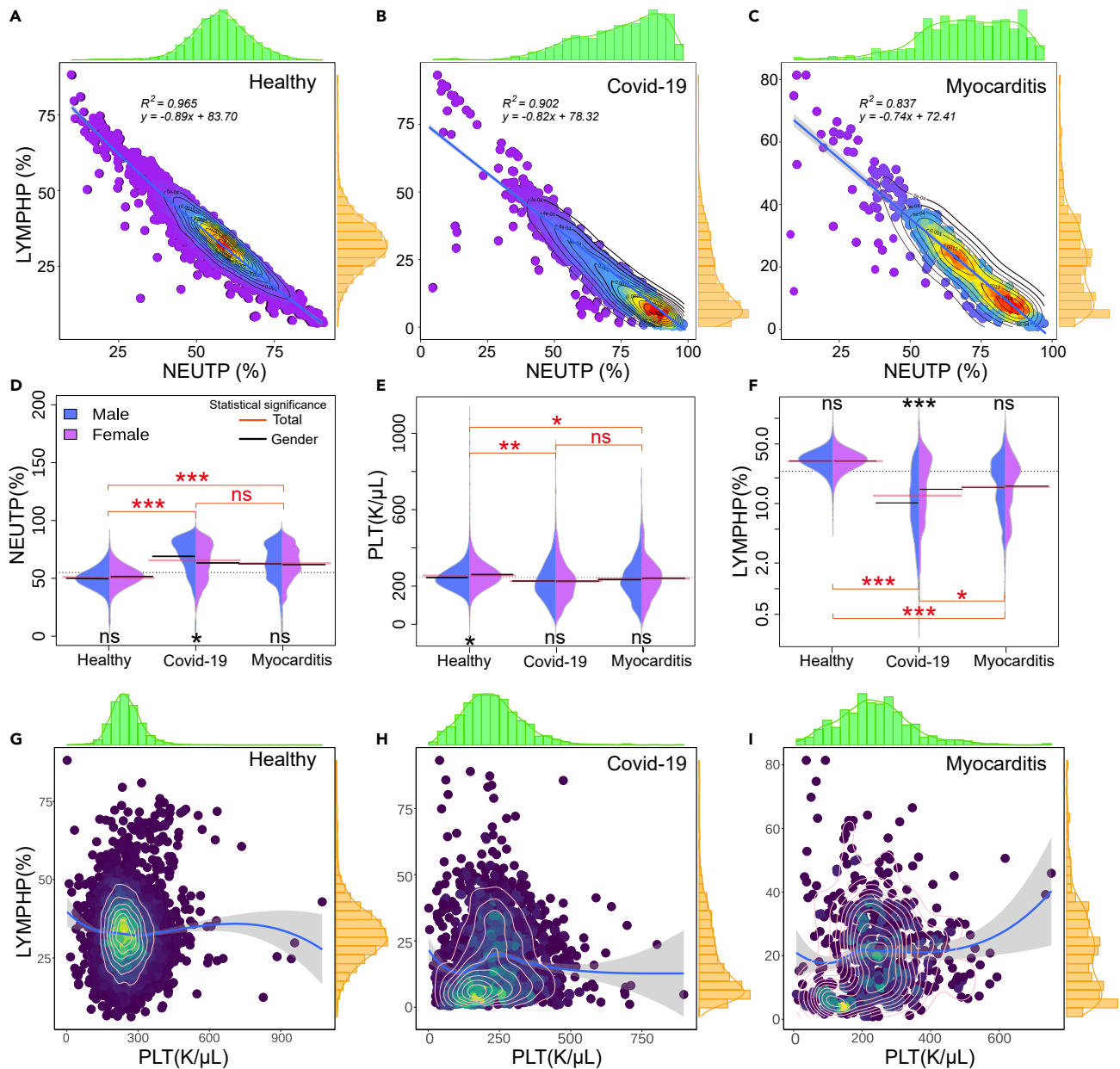


Figure 4. The blood index distribution of LNR, LPR, and individual indicators

(A) The blood index distribution of LNR in the healthy population.

(B) The blood index distribution of LNR in the COVID-19.

(C) The blood index distribution of LNR in the patients with myocarditis. Linear regression.

(D) The distribution and Statistical significance analysis of the healthy population.

(E) The distribution and Statistical significance analysis of the COVID-19.

(F) The distribution and Statistical significance analysis of the patients with myocarditis.

(G) The blood index distribution of LPR in the healthy population.

(H) The blood index distribution of LPR in the COVID-19.

(I) The blood index distribution of LPR in the patients with myocarditis. The blue line is the loess regression at all points, the shadow is the confidence interval. Interquartile range and t-test, where a statistically significant difference is denoted by *** ($p < 0.001$), ** ($p < 0.01$), and * ($p < 0.05$), and non-significance ($p > 0.05$) is denoted by "ns".

significant is the elevated inflammatory marker C-reactive protein (CRP) observed in patients with COVID-19, surpassing levels detected in patients with myocarditis. This considerable change indicates the body's intense inflammatory response in response to COVID-19 infection. Such an inflammatory surge could potentially exacerbate disease severity, which warrants further investigation. The statistical significance

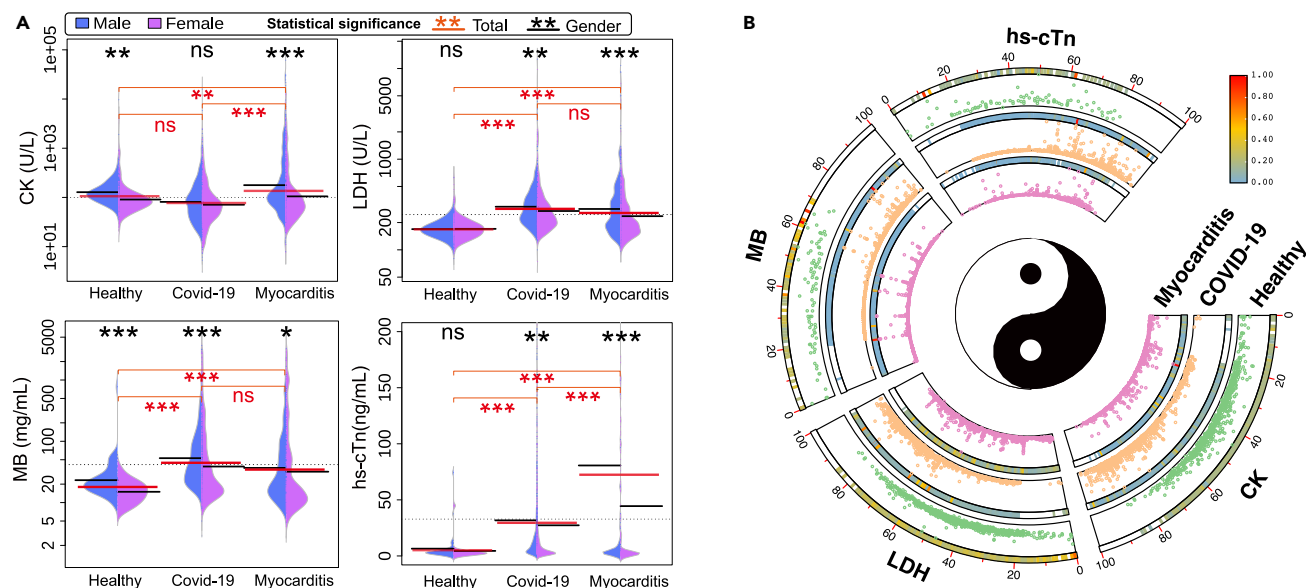


Figure 5. The distribution and statistical significance analysis results of cardiac markers (CK, LDH, MB, hs-cTn) in the healthy population, COVID-19, and myocarditis patient's groups

(A) Gender statistical significance distribution. Interquartile range and t-test, where a statistically significant difference is denoted by *** ($p < 0.001$), ** ($p < 0.01$), and * ($p < 0.05$), and non-significance ($p > 0.05$) is denoted by "ns".

(B) The age distribution of relevant myocardial indicators in each group (normalized).

analysis (Figure S2B) indicates that barring CRP levels in patients with COVID-19 and myocarditis, all other examined indicators exhibited statistically significant differences between the groups. This distinction in indicators could serve as a pivotal tool in distinguishing between COVID-19 and myocarditis, assisting in timely and accurate diagnosis and, thus, appropriate treatment.

Furthermore, correlation analyses among indicators in different groups (Figure S3) reveal a higher incidence of negative correlation amongst indicators in patients with COVID-19 and myocarditis, which could suggest a complex interplay of factors in these diseases. This relationship necessitates a more comprehensive exploration. Moreover, the correlation scatter analysis (Figure S5) shows that the indicators of patients with myocarditis largely overlap with the range of patients with COVID-19. This overlap might underscore the standard pathophysiological processes in both conditions and may have implications for shared therapeutic approaches. These findings illuminate significant patterns and correlations among critical indicators in patients with COVID-19 and myocarditis. They hold the potential to facilitate a better understanding of these diseases, aid in the development of diagnostic tools, and inform effective therapeutic strategies.

Statistical analysis of myocardial indicators

Research into myocardial biomarkers such as CK, LDH, MB, and hs-cTn has underscored their potential importance in predicting myocarditis. The distribution of myocardial-related indicators (Figure 5). We compared the myocardial markers of healthy individuals, patients with COVID-19, and patients with myocarditis. The results showed the distribution of indicators in distinct populations and differences in indicators between male and female samples. It was found that the levels of CK, LDH, MB, and hs-cTn (Figure 5A) in patients with myocarditis were significantly elevated compared to the healthy population, all displaying a highly significant statistical difference. Moreover, for patients with COVID-19 and myocarditis, the LDH and MB indicators were essentially consistent, demonstrating no statistical significance. We also visualized the distribution of age and myocardial-related markers at scattered points for each group using Figure 5B. The heat maps and scatterplots were presented for each group. The narrowed heat maps represent datasets from different age groups. After normalization, the scattered points allow for the comparison of healthy individuals, patients with COVID-19, and patients with myocarditis. This comprehensive analysis of multiple biomarkers, along with the visualization of their distribution across different age groups and populations, provides valuable insights for diagnosing and managing myocarditis.

Performance of a predictive model for myocarditis

Our data analysis highlighted marked differences in myocarditis relative to healthy individuals and patients with COVID-19. Accordingly, we are trained in a predictive model that employs relevant markers to predict the risk of myocarditis in patients with greater precision. The performance of single index prediction was tested and optimized using a naive Bayes model. We constructed and trained a model using a random forest approach for compound multi-index prediction. Predictive models for myocarditis have been receiving significant attention in contemporary research. These models aim to identify high-risk individuals using a variety of parameters such as clinical symptoms, laboratory test results, and blood analysis. The efficacy of a myocarditis predictive model is appraised based on its sensitivity, specificity, positive

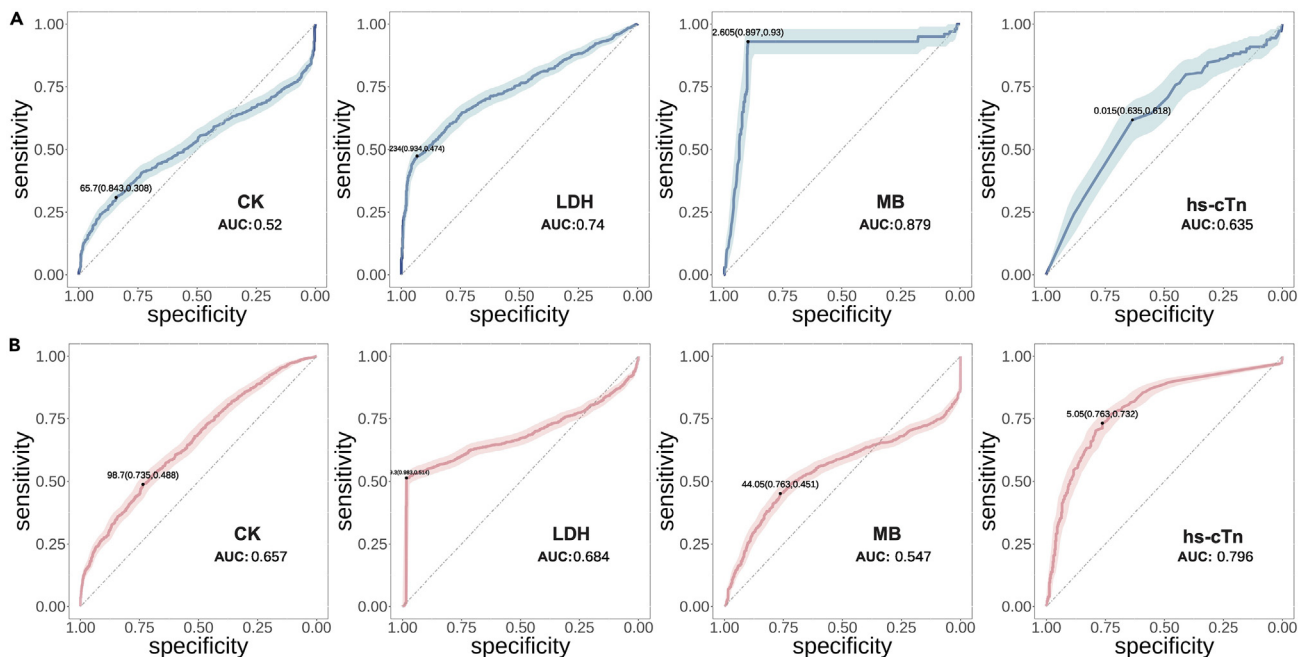


Figure 6. The performance of test results for a single indicator of the predictive model

(A) Prediction model results of myocardial marker indicators in healthy individuals and patients with myocarditis by naive Bayes classifier.

(B) Prediction model results of myocardial markers in patients with COVID-19 and myocarditis by naive Bayes classifier. (Light color is 95% confidence intervals, 25%–75% percentiles, median).

predictive value, negative predictive value, and area under the curve (AUC) of the receiver operating characteristic (ROC) curve. The results shown in Figure 6A reflect the performance of myocardial markers in healthy individuals versus those with myocarditis, light color is 95% confidence intervals, 25%–75% percentiles of model predictive performance. Both MB and LDH display commendable performance, with MB having an AUC of 0.887, indicative of high specificity and sensitivity. We concurrently examined the predictive performance of myocardial markers in patients with COVID-19 and myocarditis. As depicted in Figure 6B, the traditional marker, hs-cTn, demonstrates relatively high predictive performance, with an AUC of 0.796, a sensitivity of 0.76, and a specificity of 0.73. Additionally, the CK and LDH markers exhibit predictive solid performance.

Recognizing the difficulty in accurately predicting a complex disease using a single indicator, we focused particularly on the occurrence of secondary myocarditis following the recovery of patients with COVID-19. To achieve this objective, we analyzed myocardial markers in patients with both COVID-19 and myocarditis. We then employed a random forest machine learning model to perform multi-indicator predictions using a combination of four key myocardial markers. The results, shown by the red line in Figure 7 indicate an AUC of the ROC curve predicted by myocardial markers of 0.887, with a sensitivity of 0.776 and a specificity of 0.917, Youden's index is 0.693, suggesting good overall performance. For a more comprehensive prediction, we used ALB, HB, hs-CRP, LYMPHP, NEUTP, PLT (Figure S4), and four cardiac markers to train the multi-index predictive model. Under the comprehensive effect of multiple indicators, the performance results of this approach, depicted by the blue line in Figure 7 exceeded those of the cardiac markers alone, with an AUC of 0.924, superior sensitivity of 0.851, and a specificity of 0.905 with Youden's index is 0.775. It shows that the prediction performance of the model is greatly improved, but also has high sensitivity and specificity.

DISCUSSION

In our study, we statistically analyzed the correlation and index ratio (AGR, LNR, LPR) of each group of indicators by comparing the data of relevant patients with COVID-19 (Omicron subvariant BA. 5.2, EG.5 in supplemental information) and myocarditis with the healthy population, determined the sample information and index distribution of each group, and conducted single and comprehensive index prediction of myocarditis risk of patients with COVID-19 prognosis by the training Bayesian classification model and the random forest model. It is proven that the prediction models are feasible and have high accuracy. The demographic characteristics of these samples indicate a higher average and median age among patients with COVID-19, along with a male predominance (Figure 2A), this could suggest that older males may be at a higher risk of severe COVID-19. Clark et al.²⁶ also came to the same conclusion in their research of the population at increased risk of severe COVID-19 due to underlying health conditions. Conversely, the concentration of patients with myocarditis in the younger age group (Figure 2B) could indicate that this disease may affect younger individuals to a greater extent.²⁷ These insights could be valuable for informing public health strategies and clinical management of these conditions.

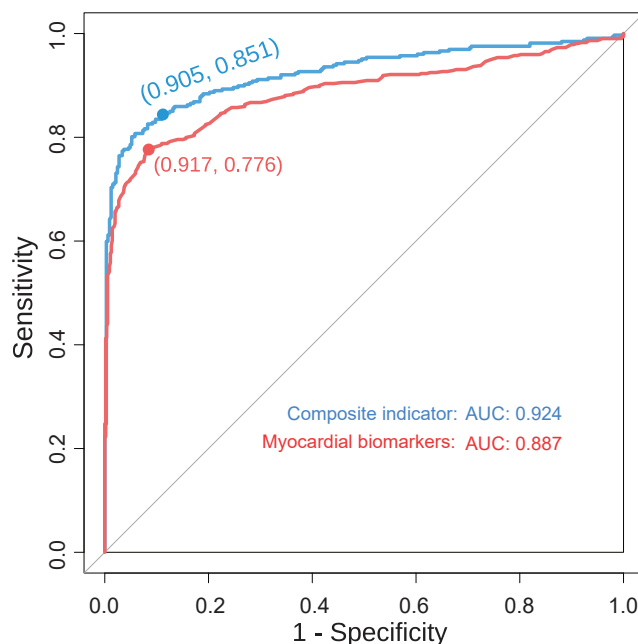


Figure 7. The performance of random forest multi-index hybrid forecasting model

The red curve showed the ROC curve of the comprehensive predictive performance of four cardiac markers, and the blue curve showed the ROC curve of ALB, HB, hs-CRP, LYMPHP, NEUTP, PLT, and the multi-indicator predictive performance of four cardiac markers by random forest.

SARS-CoV-2 may potentially target organ endothelial cells, leading to endothelial dysfunction and resulting in complications such as myocardial injury and thrombosis.^{28,29} Endothelial dysfunction emerges as a crucial element in understanding COVID-19's broad impacts. Gustafson et al.³⁰ and Fox et al.³¹ also reported on the impact of the SARS-CoV-2 on cardiac abnormalities. The severity of the condition can be assessed by examining blood-related markers such as C-reactive protein, white blood cells and other biomarkers. In their findings, the indicators for patients with myocardial injury were notably elevated, which is in line with our results shown in Figures 4F and S2. Interestingly, a unique feature of patients with COVID-19, as opposed to healthy individuals and patients with myocarditis, is the presence of two distinct indicator distribution communities. Some asymptomatic patients displayed indicators consistent with the healthy population (Figure 3A), while patients with severe COVID-19 symptoms exhibited lower ALB and globulin levels (Figure 3B). The significance of these findings lies in their potential to enhance our understanding of the disease progression and patient stratification in COVID-19.³² Among the different sub-variants, patients with EG.5 have fewer instances of low ALB, with a distribution similar to healthy individuals. This suggests that the symptoms in EG.5 patients are milder than those in BA.5.2 patients (Figure S6D). K. Li et al.³³ found that patients with acute myocarditis had lower levels of ALB and higher levels of GLB compared to healthy controls. Besides, the ALB/GLB ratio was significantly lower in patients with acute myocarditis, suggesting that this ratio may be a valuable biomarker for diagnosing and monitoring myocarditis. Jin et al.³⁴ found that the ALB/GLB ratio was significantly lower in patients with COVID-19-associated myocarditis compared to those without myocarditis. The results (Figures 3B and 3C) also clearly showed that the AGR of patients with COVID-19 and myocarditis were significantly lower than that of healthy people, which are consistent with previous research results.^{23,34,35} Furthermore, the heatmap and bar chart (Figures 3A–3C) provide a visual and intuitive way to understand the distribution of these indicators, which could be helpful in both clinical and research settings. The clear demarcation in the AGR between healthy individuals and those with COVID-19 or myocarditis could serve as a valuable tool for early detection and intervention.³⁶ Overall, these findings underscore the potential utility of the AGR as a diagnostic and prognostic tool in managing COVID-19 and myocarditis. Besides, the dual indicator distribution communities observed among patients with COVID-19 highlight the variable impact of the disease. This could have implications for personalized treatment strategies, where asymptomatic and severely symptomatic patients might require different therapeutic approaches. The findings from Figure 3D have significant consequences for our understanding of the impact of COVID-19 and myocarditis on ALB and TP levels. The marked decrease in ALB levels in patients with these conditions, compared to the healthy population, suggests that these diseases may disrupt the body's protein balance.³⁷ This could potentially lead to a variety of health complications, as ALB plays a crucial role in maintaining fluid balance, transporting substances, and providing the body with the proteins it needs.³⁸ The gender differences observed in the healthy population and their absence in patients with myocarditis could suggest that these diseases may affect men and women differently, which was also proved by Fairweather et al.³⁹ This could affect how these conditions are diagnosed and treated in different genders. The pronounced difference in protein levels between men and women due to COVID-19 is particularly striking and warrants further investigation.

Currently, LNR and LPR, as new predictive markers, have been used in many studies to predict the risk of myocarditis after recovery from COVID-19.^{40–43} Bozkurt et al.⁴⁴ studied that within three months after infection, the admission rate of myocarditis in their

center is more than ten times that of the vaccination period. Therefore, myocarditis can basically be regarded as the main sequela of COVID-19. The LNR, calculated from standard blood tests, measures systemic inflammation and immune response. It has been identified as an independent risk factor for mortality in hospitalized patients with COVID-19.^{40,41} In severe COVID-19 cases, patients often have a lower LNR due to an increase in neutrophil count, a response to infection and inflammation, and a decrease in lymphocyte count, potentially a result of viral-induced lymphocyte apoptosis or redistribution.⁴⁵ The scattering analysis of neutrophils and lymphocytes showed a strong linear relationship in the healthy population (Figures 4A–4C). By comparing the two sub-variants of SARS-CoV-2, the LNR distribution of BA.5.2 and EG.5 is basically consistent, and the slope value of fitting is also similar (Figure S6E). However, this correlation weakened in patients with COVID-19 and myocarditis, possibly due to disease-related scatter distribution.⁴⁰ The slopes of the fitted linear models for health, COVID-19, and myocarditis were 0.83, 0.82, and 0.74, respectively, indicating a decreasing trend in the LNR with disease severity. It is important to note that the density of hot spots for patients with COVID-19 and myocarditis was concentrated in areas with low lymphocyte ratios, suggesting an increase in disease severity. Our study also looked at individual indicators, such as NEUTP, PLT, and LYMPHP, in men and women. It found statistical significance only in the LYMPHP of patients with COVID-19. The indicators did not show significant differences between patients with COVID-19 and myocarditis, but there was a significant difference when these groups were compared with healthy individuals. The LPR, another ratio calculated from routine blood tests, provides insight into a person's immune and coagulation status. It has been explored as a predictive marker in various diseases.⁴⁶ Early studies suggest that the LPR could be a valuable tool in predicting the severity of COVID-19, with severe cases often showing a lower LPR due to a decrease in lymphocyte count (lymphopenia) and an increase in platelet count (thrombocytosis) in Figures 4G–4I. This imbalance could be indicative of a severe inflammatory response and hypercoagulable state.

Our research has demonstrated that CK levels tend to be elevated in patients with myocarditis (Figure 5A). Increased CK levels indicate myocardial damage, as CK is an enzyme found predominantly in the heart muscle.⁴⁷ However, it is essential to note that elevated CK levels can also be observed in other cardiac conditions.⁴⁸ Despite this limitation, CK remains a valuable biomarker in the overall assessment of myocarditis. Our study has indicated that LDH levels are elevated in myocarditis cases, reflecting tissue damage and inflammation. However, LDH is not specific to myocarditis and can also be elevated in other conditions. Besides, MB is a cardiac-specific isoform of CK and is released into the bloodstream when there is damage to the heart muscle. Research has shown that elevated levels of CK-MB are associated with myocarditis. However, it should be noted that CK-MB can also be elevated in other cardiac conditions and is not exclusive to myocarditis. Our results demonstrated that hs-cTn levels are elevated in myocarditis cases, suggesting ongoing myocardial injury, its high sensitivity enables the early detection of myocarditis and helps assess disease severity.⁴⁹ However, we can clearly observe that EG.5 patients, despite having milder symptoms, have higher levels of hs-cTn (Figure S6F), this suggests that the mutated virus might have reduced pathogenicity in the respiratory system, but enhanced pathogenicity in terms of epidermal barrier disruptions and myocardial damage. Continued research in this area will contribute to refining and optimizing myocarditis diagnosis and treatment strategies. The burgeoning field of machine learning (ML) has ushered in revolutionary capabilities in clinical medicine. Their vast potential in diagnostic and prognostic contexts is gaining traction, particularly in cancer diseases,^{50,51} and the challenging domain of COVID-19.⁵² However, the actual deployment of AI and ML in clinical decision-making is not without its challenges. The precision and reliability of these tools are contingent upon the comprehensiveness and quality of the data they are trained on. As observed in infectious disease research, many ML models are constructed on data from high-income countries, possibly skewing their efficacy in diverse socioeconomic settings.⁵⁰ Furthermore, while algorithmic performance metrics such as sensitivity and specificity are impressive, the real-world applicability and interpretability of these systems remain paramount concerns. Diving into our research, after comparing the acute myocarditis and COVID-19 groups, it is evident that the ROC of MB and LDH (Figure 6A) differ substantially between the two groups. This variance is further corroborated in the statistical significance analysis showcased in Figure 5A. Notably, the levels of MB and LDH in the COVID-19 and myocarditis groups do not present with statistical significance, leading to the stark divergence in their respective ROC curves.

Our composite multi-index prediction model (Figure 7), which amalgamates a medley of indicators including ALB, HB, hs-CRP, LYMPHP, NEUTP, PLT (as delineated in Figure S4), and a quartet of cardiac markers, elevates the predictive acumen for myocarditis risk in patients. This watershed moment in myocarditis predictive modeling harbors the promise of pinpointing high-risk individuals at an early juncture, thereby catalyzing timely interventions and ameliorating patient outcomes. Moreover, this model holds significant potential in surveilling patients with COVID-19 for emergent secondary myocarditis, equipping clinicians with a tool to preclude further health complications. In essence, our endeavors provide a deeper insight into the nexus between myocarditis and COVID-19, fortifying our arsenal to navigate and neutralize this intricate and capricious condition.

In conclusion, older males may be more susceptible to severe COVID-19, while myocarditis affects younger individuals more. Two distinct communities of indicator distribution were found among patients with COVID-19, with asymptomatic patients exhibiting markers similar to a healthy population. In contrast, those with severe symptoms had lower ALB and GLB levels. These findings inform disease progression understanding and patient stratification. Our research identified the LNR and LPR as potentially useful predictive markers for post-COVID-19 myocarditis risk, which caused the ratio to drop. However, a pronounced increase in myocarditis cases post-COVID-19 infection was noted, underlining its significance as the sequelae of COVID-19. Certain markers (CK, LDH, MB, and hs-cTn) were found to be elevated in patients with myocarditis, indicating myocardial damage. Notably, the composite multi-index prediction model demonstrated significantly improved predictive power for myocarditis risk in patients. This study provides vital insights into myocarditis and its correlation with COVID-19, potentially facilitating earlier identification of high-risk individuals paving the way for timely intervention and improved patient outcomes.

Limitations of the study

Our study focused on specific COVID-19 subvariants (Omicron subvariant BA. 5.2, EG.5), which may not represent the broader population. Second, this study primarily relied on blood markers, excluding other potential risk factors. Additionally, the sample size, although substantial, may not encompass the full spectrum of COVID-19 and myocarditis cases. External validation is necessary to assess the generalizability of the predictive models. Lastly, the study does not consider evolving COVID-19 variants or long-term effects. These limitations underscore the need for further research for comprehensive insights.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Prediction model development and validation
 - Data visualization
 - Ethics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.108524>.

ACKNOWLEDGMENTS

Thanks to Dazhou Central Hospital to supply patients' data. Meanwhile, thanks to Sun Yat-sen University's support for our research provided the instrumentation. This study was supported by Zhong Nanshan Medical Foundation of Guangdong Province (ZNSXS-2021005), Guangdong Basic and Applied Basic Research Foundation (2022A1515110555, 2023A1515010932), Guangdong-Hong Kong-Macao Joint Laboratory of Respiratory Infectious Disease (GHMJLRID-Z-202102), Dazhou City Innovation and Entrepreneurship Talent and 'Seedling' Project (23CYRC0007), and the Emergency Key Project of Guangzhou Laboratory (EKPG21-30-2).

AUTHOR CONTRIBUTIONS

Haiyang Li and Zhangkai J. Cheng led this research, analyzed data, and wrote the article. Haiyang Li guided research methods and proofread research theories. Zhangkai J. Cheng visualized figures, revised the article and supported some funding. Xing Fu sorted out the data and visualized part of the data. Mingtao Liu typeset the result drawings and article. Peng Liu, Wenhan Cao and Zhiman Liang collected and screened participants and healthy control data. Fei Wang and Baoqing Sun supervised and validated the article and supported the funding. All authors contributed to the article and agreed to submit it.

DECLARATION OF INTERESTS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: May 31, 2023

Revised: November 14, 2023

Accepted: November 20, 2023

Published: November 23, 2023

REFERENCES

1. Lu, H., Stratton, C.W., and Tang, Y.-W. (2020). Outbreak of pneumonia of unknown etiology in wuhan, china: The mystery and the miracle. *J. Med. Virol.* 92, 401–402.
2. Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., and Hsueh, P.R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob. Agents* 55, 105924.
3. Iqbal, A., Iqbal, K., Arshad Ali, S., Azim, D., Farid, E., Baig, M.D., Bin Arif, T., and Raza, M. (2021). The covid-19 sequelae: A cross-sectional evaluation of post-recovery symptoms and the need for rehabilitation of covid-19 survivors. *Cureus* 13, e13080.
4. Sanyaolu, A., Marinkovic, A., Prakash, S., Zhao, A., Balendra, V., Haider, N., Jain, I., Simic, T., and Okorie, C. (2022). Post-acute sequelae in covid-19 survivors: an overview. *SN Compr. Clin. Med.* 4, 91.

5. Huang, C., Huang, L., Wang, Y., Li, X., Ren, L., Gu, X., Kang, L., Guo, L., Liu, M., Zhou, X., et al. (2021). 6-month consequences of covid-19 in patients discharged from hospital: a cohort study. *Lancet* 397, 220–232.
6. Long, B., Brady, W.J., Koefman, A., and Gottlieb, M. (2020). Cardiovascular complications in covid-19. *Am. J. Emerg. Med.* 38, 1504–1507.
7. Bansal, M. (2020). Cardiovascular disease and covid-19. *Diabetes Metab. Syndr.* 14, 247–250.
8. Ponti, G., Ruini, C., and Tomasi, A. (2020). Homocysteine as a potential predictor of cardiovascular risk in patients with covid-19. *Med. Hypotheses* 143, 109859.
9. Hendrickson, B.S., Stephens, R.E., Chang, J.V., Amburn, J.M., Pierotti, L.L., Johnson, J.L., Hyden, J.C., Johnson, J.N., and Philip, R.R. (2021). Cardiovascular evaluation after covid-19 in 137 collegiate athletes: Results of an algorithm-guided screening. *Circulation* 143, 1926–1928.
10. Modin, D., Claggett, B., Sindet-Pedersen, C., Lassen, M.C.H., Skaarup, K.G., Jensen, J.U.S., Fralick, M., Schou, M., Lamberts, M., Gerds, T., et al. (2020). Acute covid-19 and the incidence of ischemic stroke and acute myocardial infarction. *Circulation* 142, 2080–2082.
11. Madjid, M., Safavi-Naeini, P., Solomon, S.D., and Vardeny, O. (2020). Potential effects of coronaviruses on the cardiovascular system: A review. *JAMA Cardiol.* 5, 831–840.
12. Liu, Y., Yang, Y., Zhang, C., Huang, F., Wang, F., Yuan, J., Wang, Z., Li, J., Li, J., Feng, C., et al. (2020). Clinical and biochemical indexes from 2019-ncov infected patients linked to viral loads and lung injury. *Sci. China Life Sci.* 63, 364–374.
13. Terpos, E., Ntanas-Stathopoulos, I., Elalamy, I., Kastritis, E., Sergentanis, T.N., Politou, M., Psaltopoulou, T., Gerotziakas, G., and Dimopoulos, M.A. (2020). Hematological findings and complications of covid-19. *Am. J. Hematol.* 95, 834–847.
14. Zhang, L., Yan, X., Fan, Q., Liu, H., Liu, X., Liu, Z., and Zhang, Z. (2020). D-dimer levels on admission to predict in-hospital mortality in patients with covid-19. *J. Thromb. Haemost.* 18, 1324–1329.
15. Inciardi, R.M., Lupi, L., Zacccone, G., Italia, L., Raffo, M., Tomasoni, D., Cani, D.S., Cerini, M., Farina, D., Gavazzi, E., et al. (2020). Cardiac involvement in a patient with coronavirus disease 2019 (covid-19). *JAMA Cardiol.* 5, 819–824.
16. Ullah, W., Basyal, B., Tariq, S., Almas, T., Saeed, R., Roomi, S., Haq, S., Madara, J., Boigon, M., Haas, D.C., and Fischman, D.L. (2020). Lymphocyte-to-c-reactive protein ratio: A novel predictor of adverse outcomes in covid-19. *J. Clin. Med. Res.* 12, 415–422.
17. Tamaki, S., Nagai, Y., Shutta, R., Masuda, D., Yamashita, S., Seo, M., Yamada, T., Nakagawa, A., Yasumura, Y., Nakagawa, Y., et al. (2023). Combination of neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios as a novel predictor of cardiac death in patients with acute decompensated heart failure with preserved left ventricular ejection fraction: a multicenter study. *J. Am. Heart Assoc.* 12, e026326.
18. Rabbani, G., and Ahn, S.N. (2021). Roles of human serum albumin in prediction, diagnoses and treatment of covid-19. *Int. J. Biol. Macromol.* 193, 948–955.
19. Tan, L., Wang, Q., Zhang, D., Ding, J., Huang, Q., Tang, Y.-Q., Wang, Q., and Miao, H. (2020). Lymphopenia predicts disease severity of covid-19: A descriptive and predictive study. *Signal Transduct. Target. Ther.* 5, 33.
20. Huang, D., Tang, Z., Qiu, X., Liu, X., Guo, Z., Yang, B., Guo, Q., Xiong, H., and Wang, S. (2020). Clinical features of severe patients infected with 2019 novel coronavirus: A systematic review and meta-analysis. *Ann. Transl. Med.* 8, 1664.
21. Yamasaki, Y., Ooka, S., Tsuchida, T., Nakamura, Y., Hagiwara, Y., Naitou, Y., Ishibashi, Y., Ikeda, H., Sakurada, T., Handa, H., et al. (2020). The peripheral lymphocyte count as a predictor of severe covid-19 and the effect of treatment with ciclesonide. *Virus Res.* 290, 198089.
22. Xu, X., Han, M., Li, T., Sun, W., Wang, D., Fu, B., Zhou, Y., Zheng, X., Yang, Y., Li, X., et al. (2020). EGective treatment of severe covid-19 patients with tocilizumab. *Proc. Natl. Acad. Sci. USA* 117, 10970–10975.
23. Choe, H., Kobayashi, N., Abe, K., Hieda, Y., Tezuka, T., and Inaba, Y. (2023). Evaluation of serum albumin and globulin in combination with c-reactive protein improves serum diagnostic accuracy for low-grade periprosthetic joint infection. *J. Arthroplasty* 38, 555–561.
24. Liu, Y., Du, X., Chen, J., Jin, Y., Peng, L., Wang, H.H.X., Luo, M., Chen, L., and Zhao, Y. (2020). Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with covid-19. *J. Infect.* 81, e6–e12.
25. Levenfus, I., Ullmann, E., Petrowski, K., Rose, J., Huber, L.C., Stüssi-Helbling, M., and Schuurmans, M.M. (2022). The aifell score as a predictor of coronavirus disease 2019 (covid-19) severity and progression in hospitalized patients. *Diagnostics* 12, 604.
26. Clark, A., Jit, M., Warren-Gash, C., Guthrie, B., Wang, H.H.X., Mercer, S.W., Sanderson, C., McKee, M., Troeger, C., Ong, K.L., et al. (2020). Global, regional, and national estimates of the population at increased risk of severe covid-19 due to underlying health conditions in 2020: a modelling study. *Lancet. Glob. Health* 8, e1003–e1017.
27. Snape, M.D., and Viner, R.M. (2020). Covid-19 in children and young people. *Science* 370, 286–288.
28. Sardu, C., Gambardella, J., Morelli, M.B., Wang, X., Marfella, R., and Santulli, G. (2020). Hypertension, thrombosis, kidney failure, and diabetes: is covid-19 an endothelial disease? a comprehensive evaluation of clinical and basic evidence. *J. Clin. Med.* 9, 1417.
29. Vrints, C.J.M., Krychtiuk, K.A., Van Craenenbroeck, E.M., Segers, V.F., Price, S., and Heidebuchel, H. (2021). Endothelialitis plays a central role in the pathophysiology of severe covid-19 and its cardiovascular complications. *Acta Cardiol.* 76, 109–124.
30. Gustafson, D., Raju, S., Wu, R., Ching, C., Veitch, S., Rathnakumar, K., Boudreau, E., Howe, K.L., and Fish, J.E. (2020). Overcoming barriers: the endothelium as a linchpin of coronavirus disease 2019 pathogenesis? *Arterioscler. Thromb. Vasc. Biol.* 40, 1818–1829.
31. Fox, S.E., Lameira, F.S., Rinker, E.B., and Vander Heide, R.S. (2020). Cardiac endothelialitis and multisystem inflammatory syndrome after covid-19. *Ann. Intern. Med.* 173, 1025–1027.
32. Gao, Z., Xu, Y., Sun, C., Wang, X., Guo, Y., Qiu, S., and Ma, K. (2021). A systematic review of asymptomatic infections with covid-19. *J. Microbiol. Immunol. Infect.* 54, 12–16.
33. Li, K., Fu, W., Bo, Y., and Zhu, Y. (2018). Effect of albumin-globulin score and albumin to globulin ratio on survival in patients with heart failure: a retrospective cohort study in China. *BMJ Open* 8, e022960.
34. Jin, M., Lu, Z., Zhang, X., Wang, Y., Wang, J., Cai, Y., Tian, K., Xiong, Z., Zhong, Q., Ran, X., et al. (2021). Clinical characteristics and risk factors of fatal patients with covid-19: A retrospective cohort study in wuhan, china. *BMC Infect. Dis.* 21, 1–13.
35. Huang, J., Gao, J., Zhu, W., Feng, R., Liu, Q., Chen, X., Huang, J., Yang, Z., Lin, X., Zhang, Z., and Lin, Y. (2021). Indicators and prediction models for the severity of covid-19. *Int. J. Clin. Pract.* 75, e14571.
36. Bhattacharyya, P.J., Attri, P.K., and Farooqui, W. (2020). Takotsubo cardiomyopathy in early term pregnancy: A rare cardiac complication of sars-cov-2 infection. *BMJ Case Rep.* 13, e239104.
37. Belinskaia, D.A., Voronina, P.A., Batalova, A.A., and Goncharov, N.V. (2020). Serum albumin. *Encyclopedia* 1, 65–75.
38. Quinlan, G.J., Martin, G.S., and Evans, T.W. (2005). Albumin: Biochemical properties and therapeutic potential. *Hepatology* 41, 1211–1219.
39. Fairweather, D., Beetler, D.J., Musigk, N., Heidecker, B., Lyle, M.A., Cooper, L.T., Jr., and Bruno, K.A. (2023). Sex and gender differences in myocarditis and dilated cardiomyopathy: An update. *Front. Cardiovasc. Med.* 10, 1129348.
40. Ortega-Rojas, S., Salazar-Talla, L., Romero-Cerdán, A., Soto-Becerra, P., Diaz-Vélez, C., Urrunaga-Pastor, D., and Maguiña, J.L. (2022). The neutrophil-to-lymphocyte ratio and the platelet-to-lymphocyte ratio as predictors of mortality in older adults hospitalized with covid-19 in peru. *Dis. Markers* 2022, 2497202.
41. Seyit, M., Avci, E., Nar, R., Senol, H., Yilmaz, A., Ozen, M., Oskay, A., and Aybek, H. (2021). Neutrophil to lymphocyte ratio, lymphocyte to monocyte ratio and platelet to lymphocyte ratio to predict the severity of covid-19. *Am. J. Emerg. Med.* 40, 110–114.
42. Ahmed, M., El Amrousy, D., Hodeib, H., and Elnemr, S. (2023). Neutrophil-to-lymphocyte ratio as a predictive and prognostic marker in children with dilated cardiomyopathy. *Cardiol. Young* 33, 2493–2497.
43. Mirna, M., Schmutzler, L., Topf, A., Hoppe, U.C., and Lichtenauer, M. (2021). Neutrophil-to-lymphocyte ratio and monocyte-to-lymphocyte ratio predict length of hospital stay in myocarditis. *Sci. Rep.* 11, 18101–18109.
44. Bozkurt, B., Kamat, I., and Hotez, P.J. (2021). Myocarditis with covid-19 mRNA vaccines. *Circulation* 144, 471–484.
45. Li, X., Liu, C., Mao, Z., Xiao, M., Wang, L., Qi, S., and Zhou, F. (2020). Predictive values of neutrophil-to-lymphocyte ratio on disease severity and mortality in covid-19 patients: a systematic review and meta-analysis. *Crit. Care* 24, 1–10.
46. Hamad, D.A., Aly, M.M., Abdelhameid, M.A., Ahmed, S.A., Shaltout, A.S., Abdel-Moniem, A.E., Ragheb, A.M.R., Attia, M.N., and Meshref, T.S. (2022). Combined blood indexes of systemic inflammation as a mirror to admission to intensive care unit in covid-19 patients: a multicentric study. *J. Epidemiol. Glob. Health* 12, 64–73.

47. Guzik, T.J., Mohiddin, S.A., Dimarco, A., Patel, V., Savvatis, K., Marelli-Berg, F.M., Madhur, M.S., Tomaszewski, M., Maffia, P., D'acquistio, F., et al. (2020). Covid-19 and the cardiovascular system: implications for risk assessment, diagnosis, and treatment options. *Cardiovasc. Res.* *116*, 1666–1687.
48. Bonow, R.O., O'Gara, P.T., and Yancy, C.W. (2020). Cardiology and covid-19. *JAMA* *324*, 1131–1132.
49. Lazar, D.R., Lazar, F.-L., Homorodean, C., Cainap, C., Focsan, M., Cainap, S., and Olinic, D.M. (2022). High-sensitivity troponin: A review on characteristics, assessment, and clinical implications. *Dis. Markers* *2022*, 9713326.
50. Li, H., Wang, F., and Huang, W. (2022). A novel, simple, and low-cost approach for machine learning screening of kidney cancer: an eight-indicator blood test panel with predictive value for early diagnosis. *Curr. Oncol.* *29*, 9135–9149.
51. Li, H., Cheng, Z.J., Liang, Z., Liu, M., Liu, L., Song, Z., Xie, C., Liu, J., Sun, B., and Sun, B. (2023). Novel nutritional indicator as predictors among subtypes of lung cancer in diagnosis. *Front. Nutr.* *10*, 1042047.
52. Noy, O., Coster, D., Metzger, M., Atar, I., Shenhar-Tsarfaty, S., Berliner, S., Rahav, G., Rogowski, O., and Shamir, R. (2022). A machine learning model for predicting deterioration of covid-19 inpatients. *Sci. Rep.* *12*, 2630.
53. Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). Tbttools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* *13*, 1194–1202.
54. Murphy, K.P. (2006). *Naive Bayes Classifiers*, 18 (University of British Columbia), pp. 1–8.
55. Kumar, R., and Indrayan, A. (2011). Receiver operating characteristic (roc) curve for medical researchers. *Indian Pediatr.* *48*, 277–287.
56. Kannan, R., and Vasanthi, V. (2019). *Machine Learning Algorithms with Roc Curve for Predicting and Diagnosing the Heart Disease (Soft Computing and Medical Bioinformatics)*, pp. 63–72.
57. Octaviani, T., and Rustam, D.Z. (2019). Random forest for breast cancer prediction. *AIP Conference Proceedings* *020050*, 1–6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
LABOSPECT 006 Automatic Biochemical Analyser	Hitachi, Ltd., Tokyo, Japan	https://www.hitachi.com/
Coulter Ac●T 5diff AL (Autoloader) Hematology Analyser	Beckman Coulter, Ltd, USA	https://www.beckmancoulter.com/
Quantum dots-dry fluorescence immunoassay kit	Kingfocus Co., Ltd., Shenzhen, China	https://www.king-focus.com/en/
The creatine kinase (CK) Activity Assay Kit	Solarbio Life Science Co., Ltd., Beijing, China	http://www.solarbio.net/
Lactate Dehydrogenase Activity Assay Kit	Sigma Aldrich, Ltd, USA	https://www.sigmaaldrich.com/US/en
Deposited data		
Covid-19 and Myocarditis patients' data	First Affiliated Hospital of Guangzhou Medical University, Dazhou Central Hospital	https://data.mendeley.com/datasets/chcnh9g262/1
Code for modeling and analysis	This paper	https://doi.org/10.17632/chcnh9g262.1
Software and algorithms		
Python 3.9	Python	python 3.7 version
R-4.2.1	R Core Team	https://mirrors.tuna.tsinghua.edu.cn/CRAN/
Rstudio	R desktop	Rstudio-desktop
sklearn 1.3	python package	https://scikit-learn.org/stable/
naivebayes	R package	https://github.com/majkamichal/naivebayes
Random Forest Classifier	python package	https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
GridSearchCV	python package	https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
Adobe Illustrator	Adobe	https://www.adobe.com/de/products/illustrator.html
TBtools	Chen et al. ⁵³	https://github.com/CJ-Chen/TBtools-II

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Baoqing Sun (sunbaoqing@vip.163.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data reported in this paper will be shared on Mendeley Data: <https://data.mendeley.com/datasets/chcnh9g262/1>.

All statistical analyses were run in R 4.2.1 using code that implement the algorithms described in the papers that introduced these statistical tests Code available at Mendeley Data: <https://doi.org/10.17632/chcnh9g262.1>.

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The study's sample comprised 5,230 COVID-19 patients (Omicron subvariant BA. 5.2 and E.G. 5) and 562 myocarditis patients, all of whom had recovered from COVID-19 between 1st December 2022 and 30th March 2023, all the sample were from Chinese. For variance control, we collected data from 730 EG.5 subvariant blood samples between July 15, 2023, and September 15, 2023 in [supplemental information \(Figure S6\)](#). These individuals had given informed consent to participate in the study. These patients were from two hospitals: the affiliated

hospital of Guangzhou Medical University and Dazhou Central Hospital. The gender and age distribution of the sample is depicted in Figure 2. Venous blood samples were collected from these patients, who had been medically tested and radiologically confirmed to have had COVID-19. Heparin was used as an anticoagulant during the collection process. The samples' corresponding index data were obtained using an automatic blood biochemical analysis instrument. The myocardial index and other markers were measured using the relevant kit. Any surplus blood was stored at 4 °C, centrifuged at 3000 rpm within 30 min of collection, and the supernatant was cleansed and packed before being stored at –80 °C for future use. In addition, 4,012 blood samples were randomly collected from healthy individuals who served as controls. In Figure 2B, the age distribution of the healthy population is uniform, and most of the patients with COVID-19 are elderly patients. Still, strangely, most of the patients with myocarditis, a sequela of COVID-19, are young people, and the proportion of people aged 20–30 years is the highest. This sequela has also become a significant threat to the lives of young people.

METHOD DETAILS

The outcomes evaluated in this study included Serum albumin (ALB), total protein (TP), hemoglobin (HB), blood platelets (PLT), neutrophil proportion (NEUTP), lymphocyte proportion (LYMPHP), and inflammation indicators such as C-reactive protein (CRP), high-sensitive C-reactive protein (hs-CRP), and myoglobin (MB), were meticulously collected from patients' blood. This data was procured using the LABOSPECT 006 Automatic Biochemical Analyser, a Hitachi, Ltd., Tokyo, Japan product. This sophisticated device can detect a wide range of substances, including ionic lipids, proteins, enzymes, hormones, and other metabolites. Moreover, the granulocyte count of the patients was determined using the Coulter Ac●T 5diG AL (Autoloader) Hematology Analyser, a product of Beckman Coulter, Ltd, USA. Medical System Biotechnology Co., Ltd., Ningbo, China, supplied the test kit for this purpose. Besides, the cardiac biomarkers of high-sensitive cardiac troponin-I (hs-cTn) were measured using a quantum dots-dry fluorescence immunoassay kit, a Kingfocus Co., Ltd., Shenzhen, China product. The creatine kinase (CK) Activity Assay Kit from Solarbio Life Science Co., Ltd., Beijing, China, was employed to measure Creatine Kinase. Lastly, the lactate dehydrogenase (LDH) levels were tested using the Lactate Dehydrogenase Activity Assay Kit, a product of Sigma Aldrich, Ltd, USA.

QUANTIFICATION AND STATISTICAL ANALYSIS

Continuous variables were reported as median [interquartile range (IQR)] or mean [standard deviation (SD)] values and were compared across groups using the Wilcoxon rank-sum test or Student's *t* test. *p* values were used for description rather than inference. The Hodges-Lehmann estimation was used to determine the 95% CI of the median differences for the non- normally distributed variables. The correlation coefficient measures the relationship between variables, with a positive correlation represented by a coefficient of 1 and a negative correlation represented by a coefficient of –1. Loess regression (local weight regression) for fitting the indicator distribution. The Kruskal–Wallis test is just the rank-sum test extended to more than two samples. In Table 1 the Kruskal Wallis test was used to test patients with COVID-19 and myocarditis, and corresponding *p* values between the two groups were obtained. Naive Bayes classifiers are a type of simple statistical classifiers that use Bayes' theorem and assume strong independence between features⁵⁴ and the conditional probability from the model can be decomposed as follows:

$$P(y|x) = \frac{P(y) \cdot P(x|y)}{P(x)} \quad (\text{Equation 1})$$

Where: $P(y|x)$ is the posterior probability of class (target variable) given predictor (feature), $P(y)$ is the prior probability of a class, $P(x|y)$ is the likelihood which is the probability of predictor given class, $P(x)$ is the "probability of the predictor. In the field of model research, the receiver operating characteristic (ROC) curve is a commonly used tool for evaluating the effectiveness of a model and determining its practical value.^{55,56} The "pROC" package (version 1.18) can display the ROC curve and calculate the area under the ROC Curve (AUC), which is a critical index in the ROC curve that measures whether positive outcomes are ranked higher than negative outcomes.

Prediction model development and validation

Our classification model is grounded in Bayes' theorem, the Naive Bayes model is particularly effective for large datasets and performs well in complex scenarios due to its simplicity and speed. It's especially useful when the dimensionality of the inputs is high, which is often the case in medical data. We normalised basic blood indicators to facilitate a more comprehensive understanding of the data. Normalisation is a scaling technique that adjusts the values measured on different scales to a notionally standard scale. This data was followed by "naivebayes" algorithm randomly dividing the data into a training set (70%) and a validation set (30%). The training set is used to build the model, while the validation set is used to test the model's predictive power. According to Octaviani and Rustam⁵⁷ there is a positive correlation between the prediction accuracy of the Random Forest (RF) model and the number of training sets used. The RF model is a popular machine learning algorithm that operates by constructing multiple decision trees and outputting the class, which is the mode of the categories or mean prediction of the individual trees. We implemented a Python 3.9 RF model using the "sklearn" library version 1.3 to optimise the results. The "GridSearchCV" module was employed to fine-tune the parameters of the RF model. This module exhaustively searches over specified parameter values for an estimator. We selected approximately 200 trees with 15 variables each and a maximum depth of 50. We gathered the results, selected the most efficient model, and evaluated its prediction accuracy using the test set. The number of variables chosen for each tree was also optimised. We employed cross-validation during the parameter adjustment process to prevent overfitting and ensure

the model's stability and practicality. Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The model's discrimination performance was evaluated using the Receiver Operating Characteristic (ROC) curve and the associated Area Under the Curve (AUC) value. The ROC curve is a plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. At the same time, the AUC measures the entire two-dimensional area underneath the ROC as a whole curve.

Data visualization

This study utilised various tools and techniques for data acquisition, statistical analysis, and visualisation. The primary software used for these tasks were R Core Team version 4.2.1 and Python 3.9, both widely used in the scientific community for their versatility and robustness. A range of graphical visualisations was generated to effectively communicate the findings, with specific software and packages employed for each type of visualisation. We used Adobe Illustrator to optimise color and typesetting, which is crucial for creating clear and aesthetically pleasing visualisations. This software provides a wide range of tools for creating and editing vector graphics. Column charts and box diagrams were created using the "Matplotlib" Python package (version 3.5). Additionally, pair plots were visualised with the assistance of the "seaborn" Python package (version 0.11.2). The heatscatter and statistically significant were generated using the "ggplot2" package (version 3.3.6) and the "LSD" package (version 4.1-0) in R. Tai chi distribution map was made by TBTtools software.⁵³ In order to compare differences between male and female medical records of various cancers, the "beanplot" package (version 1.3.1) was employed. Lastly, the correlation coefficients between data indices in this study were visualised using the "ggcorrplot" package (version 0.1.3).

Ethics

This study was conducted in line with the Declaration of Helsinki. It was approved by the Scientific Research Project Reviews Ethics Committee of the First Affiliated Hospital of Guangzhou Medical University (Document 2021 No. K25) and the ethics committee of Dazhou Central Hospital (protocol code 2023(034)).