BMC Endocrine Disorders

RESEARCH

Open Access

Type 2 diabetes and susceptibility to COVID-19: a machine learning analysis

Motahare Shabestari[1], Reyhaneh Azizi[2] and Akram Ghadiri-Anari[2*]

Abstract

Background Type 2 diabetes mellitus (T2DM) was one of the most prevalent comorbidities among patients with coronavirus disease 2019 (COVID-19). Interactions between different metabolic parameters contribute to the susceptibility to the virus; thereby, this study aimed to rank the importance of clinical and laboratory variables as risk factors for COVID-19 or as protective factors against it by applying machine learning methods.

Method This study is a retrospective cohort conducted at a single center, focusing on a population with T2DM. The patients attended the Yazd Diabetes Research Center in Yazd, Iran, from February 20, 2020, to October 21, 2020. Clinical and laboratory data were collected within three months before the onset of the COVID-19 pandemic in Iran. 59 patients were infected with COVID-19, while 59 were not. The dataset was split into 70% training and 30% test sets. Principal Component Analysis (PCA) was applied to the data. The most important components were selected using a 'sequential feature selector' and scored by a Linear Discriminant Analysis model. PCA loadings were then multiplied by the PCs' scores to determine the importance of the original variables in contracting COVID-19.

Results HDL-C, followed by eGFR, showed a strong negative correlation with the risk of contracting the virus. Higher levels of HDL-C and eGFR offer protection against COVID-19 in the T2DM population. But, the ratio of BUN to creatinine did not show any correlation. Conversely, the AIP, TyG index and TG showed the most positive correlation with susceptibility to COVID-19 in such a way that higher levels of these factors increase the risk of contracting the virus. The positive correlation of diastolic BP, TyG-BMI index, MAP, BMI, weight, TC, FPG, HbA1C, Cr, systolic BP, BUN, and LDL-C with the risk of COVID-19 decreased, respectively.

Conclusion The atherogenic index of plasma, triglyceride glucose index, and triglyceride levels are the most significant risk factors for COVID-19 contracting in individuals with T2DM. Meanwhile, high-density lipoprotein cholesterol is the most protective factor.

Keywords Machine learning (ML), Type 2 diabetes mellitus (T2DM), COVID-19, Insulin resistance, Atherogenic dyslipidemia

*Correspondence:
Akram Ghadiri-Anari
ghadiriam@yahoo.com
[1]Medical School, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
[2]Diabetes Research Center, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

Introduction

The coronavirus disease 2019 (COVID-19) caused a pandemic and significant health challenges all around the world. Among the infected population, hypertension and type 2 diabetes Mellitus (T2DM) were the most prevalent comorbidities [1]. The hallmarks of T2DM are insulin resistance (IR) and decreased tissue response to insulin's stimulation effect, leading to systemic inflammation, oxidative stress, vascular dysfunction, and impaired immune system reactions [2, 3]. These characteristics predispose individuals with T2DM to infection [4].

Diabetic dyslipidemia, also known as atherogenic dyslipidemia, is a macrovascular complication [5] and is closely associated with insulin resistance in the T2DM population. The dyslipidemia includes elevated levels of triglycerides and reduced levels of high-density lipoprotein cholesterol [6], which develops a chronic inflammation and causes a sustained release of cytokines [7]. This metabolic disturbance has been reported as an independent risk factor for adverse outcomes in COVID-19 patients [8].

Diabetic kidney disease impacts approximately 40% of individuals with T2DM [9]. This microvascular complication induces uremia, which in turn disturbs innate and adaptive immune systems and increases susceptibility to infection [10]. The estimated glomerular filtration rate as a measure of kidney function has demonstrated a negative association with the severity of COVID-19 [11].

Machine learning-based models have been progressively applied in the medical field to diagnose, treat, and evaluate the prognosis of various diseases, as well as to predict and score the risk of developing diseases [12]. Unlike conventional statistical methods, these algorithms can explore the complex relationships between different clinical variables and their interactions to achieve a good and accurate predictive performance [13].

Abnormalities of clinical and laboratory variables in individuals with T2DM make this population susceptible to contracting COVID-19. Most of the previous studies examined the association between vulnerability to COVID-19 and each of the clinical and laboratory features in isolation without considering potential interactions among the features. To address this gap, our study applies various machine learning algorithms to determine the relative importance of each feature's role in susceptibility to the virus. This helps to manage diabetic patients effectively during future pandemics.

Materials and methods

Study design

This retrospective cohort study was conducted at the Yazd Diabetes Research Center in Yazd, Iran, utilizing data collected from patients who attended the center between February 20, 2020, and October 21, 2020. The Research Ethics Council of Shahid Sadoughi University of Medical Sciences approved the study in Yazd, Iran (IR.SSU.REC.1401.097).

Patients and population

In this study,118 participants with T2DM aged between 30 and 60 were recruited. Clinical data and laboratory measurements were extracted from their medical records. Individuals who attended irregular follow-up visits or had a history of immunodeficiency, neoplasia, co-infection, and smoking were excluded. In order to minimize diabetic complications, the study specifically targeted individuals with a duration of T2DM between 3 and 7 years; therefore, none of the participants had macrovascular complications. Additionally, those without medical records within three months before the pandemic onset in Iran were not considered.

The "COVID-19 positive" group included 59 patients who tested positive for COVID-19 using the polymerase chain reaction (PCR) technique from February 20 to May 19, 2020, but were not hospitalized.

The "COVID-19 negative" group included 59 individuals with no documented history of COVID-19 infection before October 21, 2020. Gender and age matching were performed across the two groups.

Clinical variables and laboratory measures

Clinical data included age, gender, body mass index (BMI), and systolic and diastolic blood pressure. Blood samples obtained after 12 h of fasting and were analyzed for hemoglobin A1C (HbA1C), fasting plasma glucose (FPG), total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), blood urea nitrogen (BUN), and creatinine (Cr). The estimated glomerular filtration rate (eGFR) was calculated using the CKD-EPI formula based on creatinine [14]. The triglyceride glucose (TyG) index, triglyceride glucose-body mass index (TyG-BMI), and atherogenic index of plasma (AIP) were calculated using the following equations: [15, 16].

$$\begin{aligned}\text{TyG Index} &= \text{Ln}\ (\text{fasting glucose}\,(\text{mg/dL}) \\ &\times\ \text{triglycerides}\,(\text{mg/dL})/2)\end{aligned}$$

$$\text{TyG} - \text{BMI} = \text{BMI} \times \text{TyG index}$$

$$API = log\frac{Triglycerides}{HDL}$$

Statistical analysis

Statistical analyses were performed using SPSS version 27. The normality of data was evaluated using the Shapiro–Wilk test. Variables following a normal distribution

were presented as means±standard deviation (SD), while non-normally distributed variables were reported as medians and interquartile ranges. Differences in variables between the two groups were analyzed using the Independent Samples T-test or the Mann–Whitney U test. A P-value below 0.05 was considered statistically significant.

Machine learning to predict susceptibility to COVID-19

Machine learning (ML) models were used to identify the importance of each variable in predicting susceptibility to COVID-19. The dataset was split into 70% training and 30% test sets. All the computations were conducted using Python (version 3.11) and the scikit-learn library (version 1.2.2). The analysis proceeded through the following steps:

Standardization

Each variable was standardized using the 'StandardScaler' method to adjust them to have a mean of 0 and a standard deviation of 1 [17]. This procedure aimed to reduce biases resulting from differences in the measuring units.

Principal component analysis (PCA)

Due to high correlations among the original variables (Fig. 1), PCA was applied to transform the standardized variables into linearly uncorrelated components [18]. The number of principal components (PCs) was determined

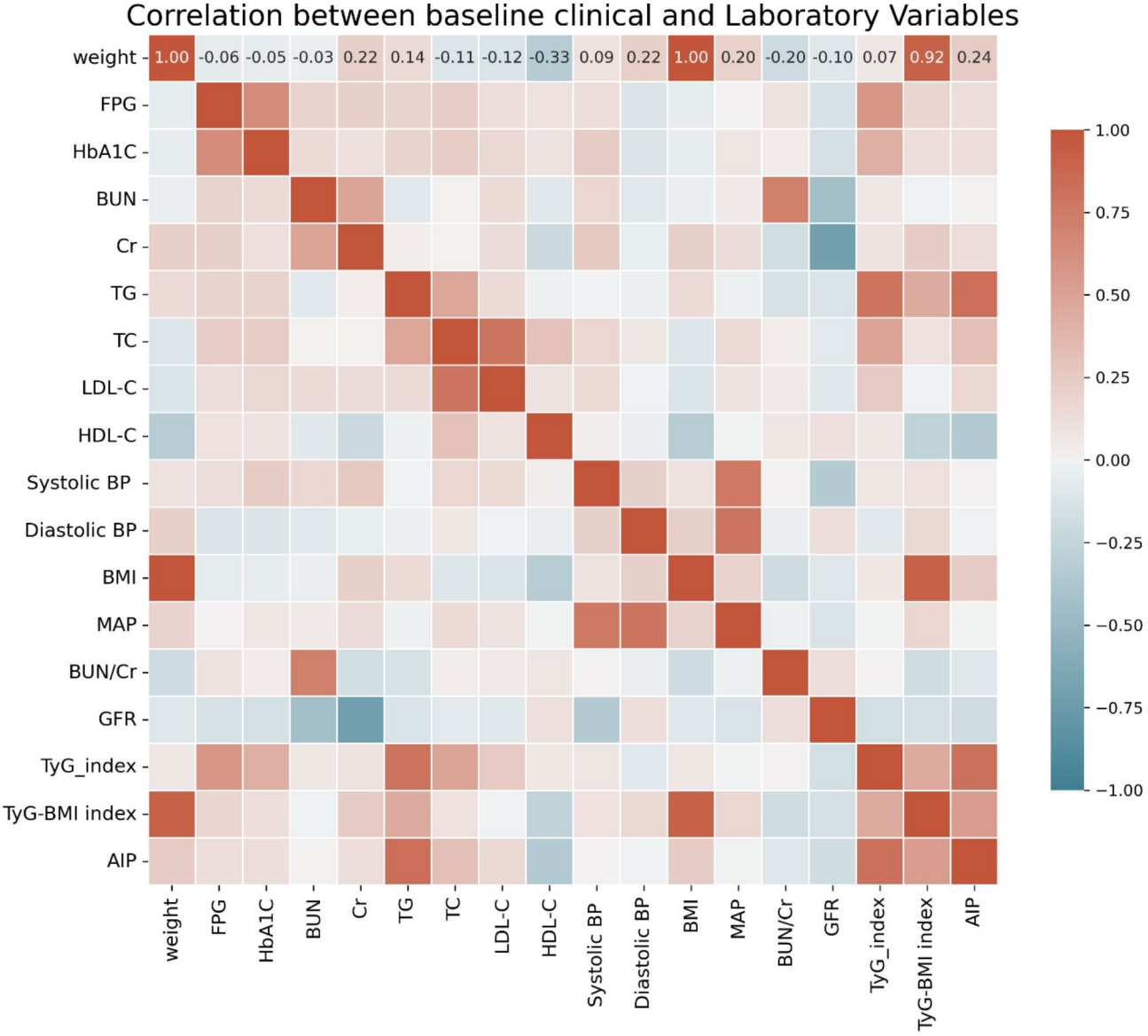

Fig. 1 A heatmap plot of the correlation between baseline variables

by setting the n_components parameter to 0.95, ensuring that 95% of the original variance was retained.

Features selection

A 'sequential feature selector' was applied to find the most predictive PCs [19] using a receiver operating characteristic (ROC) curve and area under the curve (AUC) as a scoring metric [20]. ROC-AUC was selected as the primary metric because it provides a comprehensive evaluation of model performance across all classification thresholds, which is important for susceptibility predictions. ROC-AUC was calculated using a repeated stratified k-fold cross-validation method (five-fold and ten-repeat). To find the best selector for the 'sequential feature selector,' Six different machine learning models were trained on the training set, including Linear Discriminant Analysis (LAD) [21], Logistic Regression (LR) [22], Support Vector Machine (SVM) [23], K-Nearest Neighbor (KNN) [24], Random Forest (RF) [25], eXtreme Gradient Boosting classifier (XGBoost classifier) [26]. Hyperparameter tuning of each model was performed with either 'grid search cv' or 'randomize search cv' to find the optimal parameters. The model with the highest ROC-AUC score on the test set was chosen as the selector. Feature selection was performed in both forward and backward directions, producing two sets of selected PCs.

Scoring selected PCs

Regression-based and tree-based models, including XGBoost, RF, Least Absolute Shrinkage and Selection Operator (LASSO) [27], LR and LDA were trained on the forward-selected and backward-selected PCs. Model evaluation on the test set was conducted using repeated stratified k-fold cross-validation (five-fold and ten-repeat) to calculate ROC-AUC scores. The model with the highest scores across both sets of PCs was chosen, and the set yielding the higher score was used to assess PCs' importance via the model's 'coef' attribute.

Translate PCA results back to the original features

PCA loadings were multiplied by the PCs' importance as identified by the chosen model from the previous step. This helped to determine the most critical clinical and laboratory variables that contributed to the prediction model and were strongly associated with COVID-19 susceptibility.

Table 1 Baseline clinical and laboratory characteristics of 59 patients in each group

Variables	COVID-19 positive ($n=59$)	COVID-19 negative ($n=59$)	*P*-value
Age (years old)	55 ± 8	56 ± 8	0.84
Female	$N=42$ (71.2%)	$N=42$ (71.2%)	-
Male	$N=17$ (28.8%)	$N=17$ (28.8%)	-
Weight [a] (kg)	81.8 ± 15	76.3 ± 10.2	**0.024**
BMI [a] (kg/m2)	28.9 ± 5.3	27 ± 3.6	**0.024**
Systolic BP (mmHg)	130 ± 20	130 ± 25	0.391
Diastolic BP (mmHg)	75 ± 10	70 ± 15	**0.039**
MAP [a] (mmHg)	93.4 ± 9.4	93.2 ± 9.2	0.502
FPG (mg/dL)	145 ± 61	145.4 ± 40 [a]	0.466
HbA1C [a] (%)	7.3 ± 1.4	7.2 ± 1.1	0.674
BUN (mg/dL)	13.534 ± 4.7	13.067 ± 6.5	0.842
Cr (mg/dL)	0.94 ± 0.32	0.91 ± 0.23	0.881
BUN/Cr	14.87 ± 4.88 [a]	15.75 ± 5.04	0.692
eGFR [a] (mL/min/1.73 $m2$)	73.3 ± 16.1	74.4 ± 18.6	0.722
TG (mg/dL)	202 ± 85	139 ± 72	**< 0.001**
TC (mg/dL)	170.5 ± 40	150 ± 65	**0.047**
LDL-C (mg/dL)	86.19 ± 33.12 [a]	74 ± 38	0.429
HDL-C (mg/dL)	47 ± 11.79	48 ± 14	0.794
AIP	0.57 ± 0.18 [a]	0.45 ± 0.21	**< 0.001**
TyG index	4.14 ± 0.22 [a]	3.98 ± 0.29	**< 0.001**
TyG-BMI index [a]	119.73 ± 21.12	107.44 ± 19.12	**0.001**

Note Values were presented as median ± Interquartile Range or number and percentage. [a] Normal distributed values were presented as mean ± Standard deviation. Abbreviations: COVID-19, coronavirus disease 2019; BMI, body mass index; BP, blood pressure; MAP, mean arterial pressure; FPG, fasting plasma glucose; HbA1c, glycated hemoglobin; BUN, blood urea nitrogen; Cr, creatinine; eGFR, estimated glomerular filtration rate; TG, triglyceride; TC, total cholesterol; TyG, triglyceride-glucose index; TyG-BMI, Triglyceride glucose-body mass index

Results

Baseline characteristics

Finally, 59 patients with type 2 diabetes in each group underwent analysis. Of these patients, 42 were female (71.2%) and 17 were male (28.8%). Weight ($P=0.024$), BMI ($P=0.024$), DBP (0.039), TG levels ($P<0.001$), TC levels ($P=0.047$), AIP ($P<0.001$), TyG index ($P<0.001$), and TyG-BMI index ($P=0.001$) were all higher in the 'COVID-19 positive' group due to statistical tests (Table 1).

Feature selection

The mean AUCs were calculated from six different machine-learning models in order to evaluate their performance (Table 2). The LR model demonstrated the best

Table 2 The ROC-AUC score was obtained after applying repeated stratified k-fold cross-validation (five-fold and ten-repeat) on models

Score\Models	SVC	LDA	KNN	LR	RF	XGBoosting
ROC-AUC	0.575	0.681	0.630	0.715	0.57125	0.454

Abbreviations SVC, Support Vector Classifier; LDA, Linear Discriminant Analysis; KNN, K-Nearest Neighbor; LR, Logistic Regression; RF, Random Forest; XGBoosting, eXtreme Gradient Boosting

performance, with a mean ROC-AUC of 0.715. Therefore, the LR was used as a selector in the 'sequential feature selector' method. Four PCs with indices of 0,3,6,8 were chosen from the primary nine PCs using a forward selection method. On the other hand, five PCs with indices of 0,3,5,6,8 were selected from the nine primary PCs using a backward selection method.

Scoring selected PCs

The performance of five regression-based and tree-based models on both sets of forward-selected and backward-selected PCs was evaluated through the mean AUCs (Fig. 2). The best performance, with a mean ROC-AUC of 0.714, was achieved when LDA was applied to the backward-selected PCs.

Map PCA components back to original features

The 'coef' attribute of the trained LDA model was used to obtain the importance of backward-selected PCs. The role of each variable in predicting susceptibility to COVID-19 was scored by multiplying the PCA loadings by the PCs' importance.

HDL-C, followed by eGFR, showed a strong negative correlation with the risk of contracting the virus. Higher levels of HDL-C and eGFR offer protection against COVID-19 in the T2DM population. But, the ratio of BUN to creatinine did not show any correlation. Conversely, the AIP, TyG index and TG showed the most positive correlation with susceptibility to COVID-19 in such a way that higher levels of these factors increase the risk of contracting the virus. The positive correlation of diastolic BP, TyG-BMI index, MAP, BMI, weight, TC, FPG, HbA1C, Cr, systolic BP, BUN, and LDL-C with the risk of COVID-19 decreased respectively (Fig. 3).

Discussion

In this study, we employed machine learning (ML) to identify the susceptibility to COVID-19 among the T2DM population based on clinical and laboratory variables. We obtained the best estimation performance when LDA was applied to backward-selected PCs, which contained five PCs out of the primary nine PCs. The model yielded a mean ROC-AUC of 0.714. Due to a small sample size, a ten-repeated stratified five-fold cross-validation approach was used to calculate the mean ROC-AUC. This mitigates potential overestimation or underestimation of the model's predictive capacity.

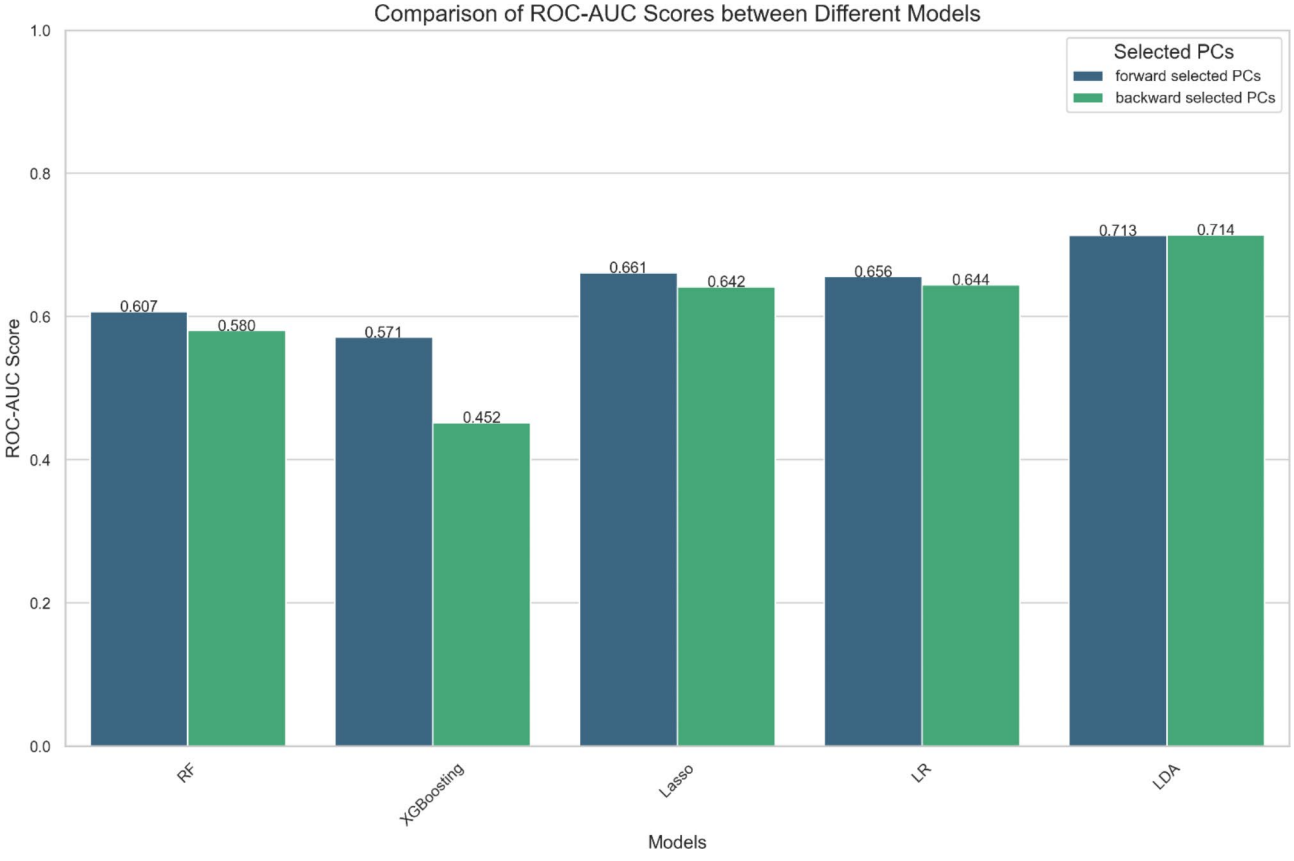

Fig. 2 A bar plot compares models' performance on forward-selected and backward-selected PCs to select the best model and the best set of PCs to predict COVID-19 susceptibility

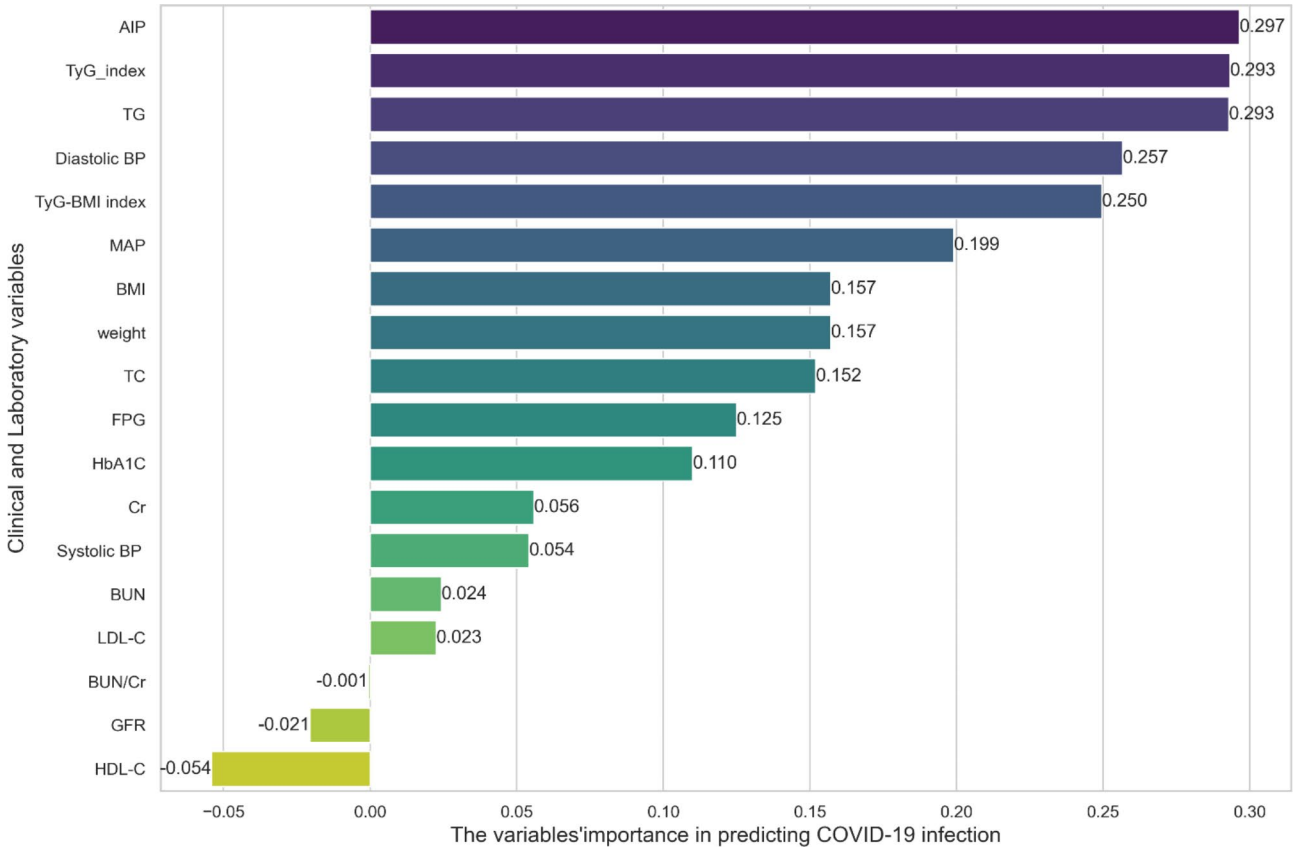

Fig. 3 The variables 'importance was calculated by multiplying the PCA loadings by the PCs' importance (the PCs' importance was derived from the 'coef' attribute of the LDA model trained on backward-selected PCs)

In the present study, HDL-C, followed by eGFR, was observed as a protective factor against COVID-19 infection in individuals with T2DM, While the AIP, TyG index and TG levels were the most significant risk factors for predicting susceptibility to COVID-19.

The findings obtained from the Mann- Whitney U test revealed that HDL-C and eGFR did not have statistically significant P-values between the 'COVID-19 positive' and 'COVID-19 negative' groups. Disparities in the outcomes may be due to the different analytic approaches of ML models and statistical tests. ML models are designed to enhance prediction accuracy by several mechanisms; they identify the complex interactions between variables, conduct concurrent analyses of multiple variables, and are able to handle widely fluctuating data. Conversely, statistical tests have limitations in dealing with these attributes [28].

Low levels of HDL-C have been identified as a risk factor for different types of infection [29–31]. HDL-C, along with its most important component, apolipoprotein A-I (Apo A-I), contributed to the susceptibility to COVID-19 [32–34]. An increase of 10 mg/dl in HDL-C or Apo A-I was able to reduce the risk of COVID-19 by 10% [34]. There are multiple mechanisms behind this link; for example, the receptor for HDL-C is called scavenger receptor protein B-I (SR B-I), which facilitated the entry of COVID-19 into cells that have angiotensin-converting enzyme 2 (ACE2) receptor [35]. Therefore, a higher concentration of HDL-C inhibited the virus' entry through the SR B-I pathway [36]. Additionally, Apo A-I disturbed the viral entrance into body cells independently [37]. Studies employing a genetic approach have suggested that low levels of HDL-C may have a causal effect on developing the infection [38, 39]. Our result about the protective effect of HDL-C was consistent with these findings.

In this study, it was determined that eGFR was the second protective factor against COVID-19. This was consistent with the findings of previous observational studies [11, 40, 41]. Lim et al. [40] reported an inverse association between eGFR and the risk of COVID-19, even among patients experiencing mild to moderate kidney dysfunction. Also, a study that used the Mendelian randomization analysis method found that kidney dysfunction causes increased susceptibility to contracting the virus [42]. However, we found the influence of eGFR on susceptibility to COVID-19 was almost comparable to that of BUN and approximately half of that of creatinine.

Unlike eGFR, both BUN and creatinine were identified as risk factors for contracting COVID-19 (Fig. 3). In fact, creatinine demonstrated a more pronounced effect than eGFR in this study, according to the PCA analysis, but relative to other variables acting as risk factors, creatinine was ranked as the least influential variable.

Reduced renal function was associated with impaired protein catabolism that, in turn, induces chronic oxidative stress and systemic inflammation. Moreover, the retention of toxic metabolites in kidney dysfunction inhibits immune cell activation and increases their apoptosis, which leads to systemic immunosuppression [43].

Insulin resistance (IR) is a condition in which tissues' response decreases to the stimulatory effect of insulin, leading to hyperinsulinemia and hyperglycemia. IR plays an important role in the pathogenesis of T2DM and makes this population more vulnerable to infection. IR and its related hyperglycemia increase the production of interleukin-6(IL-6), IL-1β, and TNF-α and develop a chronic inflammation; therefore, impair the function of the immune system [44]. Hyperinsulinemia increases membrane expression of ACE2, which serves as the receptor for COVID-19 in host cells [45], thereby facilitating viral entry and amplifying viral load. Conventional methods utilized for the assessment of IR are often costly and require technical expertise that may not be available across clinical settings [46]. Recently, cost-effective and valuable biomarkers have been introduced for the estimation of IR, including the TyG index [15], the TyG-BMI index [47], and the AIP [48].

We found the AIP and TyG index were two of the most critical risk factors for contracting COVID-19 among T2DM participants. A previous study on Iranian patients reported that the TyG index and TG/HDL-C ratio (AIP is equal to the base ten logarithm of this ratio) positively correlated with COVID-19 infection and prognosis [49]. Another study that considered diabetic patients showed the TyG index as a predictor for COVID-19 severity and mortality [50]. AIP was also shown to be associated with intubation and intensive care admission in hospitalized infected patients [51]. A few studies have investigated the relationship between the TyG-BMI index and COVID-19, such as a retrospective study on T2DM patients, which found that this index increased after contracting the viral infection [52]. We found the TyG-BMI index to be a risk factor for COVID-19 infection; however, its predictive value was lower than that of the AIP and TyG index.

IR promotes atherogenic dyslipidemia in the T2DM population by reducing HDL-C levels and increasing TG levels [53]. Furthermore, free fatty acids (FFA), which are generated during the breakdown of TG by lipoprotein lipase, exacerbate IR through numerous mechanisms [54]; FFA activates the proinflammatory pathway in the skeletal muscle known as nuclear factor (NF-κB), resulting in the secretion of proinflammatory cytokines and elevation in monocyte chemoattractant protein-1 (MCP-1) levels [55, 56]. MCP-1, in turn, leads to enhanced macrophage differentiation, thereby exacerbating the inflammatory state [57]. The present study found TG as a risk factor of equal importance to the AIP and TyG index for the development of COVID-19. Our findings are in line with previous studies that found TG levels as a predictor of the infection severity [58, 59].

Hypertension has been recognized as a potential risk factor for COVID-19, with a prevalence ranging from 27 to 34.6% among infected patients [60, 61]. Factors involved in blood pressure regulation, notably sodium level, aldosterone, and angiotensin II, contribute to the generation of reactive oxygen species. The background inflammation disturbs cell signaling and cell activation, especially within the immune system [62–65]. The dysregulated innate and adaptive immune cells produce several cytokines and worsen the inflammation [63, 65].

In addition to chronic inflammation, there are several factors related to hypertension that increase susceptibility to infection. These include a reduction in lymphocyte count [66], activation of inefficient CD8+cell types in antiviral defense [67], and vascular stiffness [68, 69]. Interestingly, we found that high diastolic blood pressure was approximately five times more important than high systolic blood pressure in predisposing individuals to COVID-19 (Fig. 3).

The present study found that BMI, weight, and TC levels had equivalent risks of contracting the infection. Following these factors, the importance of FPG, HbA1C, Creatinine, systolic blood pressure, BUN, and LDL-C levels decreased as risk factors for COVID-19, respectively.

To the best of our knowledge, this is the first study to employ ML algorithms to value metabolic and clinical parameters, which are commonly assessed during follow-up visits for individuals with T2DM, in predicting COVID-19 susceptibility. Therefore, the outcomes of the present study can help healthcare practitioners evaluate and score the risk factors associated with infection in this population, leading to better clinical decision-making.

The primary limitation of our study is the small sample size of each group because of our inclusion criteria. We enrolled individuals who received regular follow-up in our outpatient clinic at least once every three months and whose T2DM was well-controlled. Also, we only included those with laboratory measurements obtained within three months before the pandemic onset in Iran, which further limited the sample size. The lack of data on urine analysis and albuminuria resulted in their omission from the final model, which should be considered as a limitation in the interpretation of the findings. The ML method applied in this study follows standard analysis protocols. While our findings may be helpful for the

development of future diagnostic or therapeutic interventions, it is important to note that the current research does not involve any product, device, or therapeutic intervention subject to FDA regulations.

Conclusion

In this study, it was found that TG levels and IR markers (AIP and TyG-index) are the most significant risk factors for COVID-19 susceptibility in the population with T2DM. Furthermore, diastolic blood pressure, TyG-BMI index, BMI, weight, TC levels, FPG, HbA1C, creatinine, systolic blood pressure, BUN, and LDL-C levels showed decreased significance as risk factors. Conversely, HDL-C levels, followed by eGFR, were identified as the protective factor against the infection.

Abbreviations

T2DM	Type 2 diabetic mellitus
COVID-19	Coronavirus disease 2019
PCA	Principal Component Analysis
HbA1C	Hemoglobin A1C
FPG	Fasting plasma glucose
HDL-C	High-density lipoprotein cholesterol
LDL-C	Low-density lipoprotein cholesterol
TC	Total cholesterols
TG	Triglycerides
AIP	Atherogenic index of plasma
TyG index	Triglyceride glucose index
TyG-BMI index	Triglyceride glucose-body mass index
BP	Blood pressure
MAP	Mean arterial pressure
eGFR	Estimated glomerular filtration rate
ML	Machine learning
ROC-AUC	Receiver operating characteristic curve and area under the curve
LAD	Linear Discriminant Analysis
LR	Logistic Regression
SVM	Support vector machine
KNN	K-Nearest Neighbor
RF	Random forest
XGBoost classifier	eXtreme Gradient Boosting classifier
LASSO	Least Absolute Shrinkage and Selection Operator
Apo A-I	Apolipoprotein A-I
SR B-I	Scavenger receptor protein B-I
ACE2	Angiotensin-converting enzyme 2
NF-κB	Nuclear factor κB
MCP-1	Monocyte chemoattractant protein-1

Acknowledgements
Thanks to Doctor Foroozan Salari for her cooperation in collecting data.

Author contributions
M.S wrote the original draft, analyzed the data, and coded in Python R.A edited the original draft, prepared the tables and figures. A.G was the supervisor and manager of the project, designed the methodology, and reviewed the investigation.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability
All data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study was approved by the Research Ethics Committee of Shahid Sadoughi University of Medical Sciences, Yazd, Iran (IR.SSU.REC.1401.097). Written informed consent to participate was obtained from all enrolled subjects. The study was performed in accordance with the principles of the Helsinki Declaration.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 23 July 2024 / Accepted: 16 October 2024
Published online: 21 October 2024

References

1. Stoian AP, Banerjee Y, Rizvi AA, Rizzo M. Diabetes and the COVID-19 pandemic: how insights from recent experience might Guide Future Management. Metab Syndr Relat Disord. 2020;18(4):173–5.
2. Leahy JL. Pathogenesis of type 2 diabetes mellitus. Arch Med Res. 2005;36(3):197–209.
3. Ormazabal V, Nair S, Elfeky O, Aguayo C, Salomon C, Zuñiga FA. Association between insulin resistance and the development of cardiovascular disease. Cardiovasc Diabetol. 2018;17(1):122.
4. Shah BR, Hux JE. Quantifying the risk of infectious diseases for people with diabetes. Diabetes Care. 2003;26(2):510–3.
5. Farmer JA. Diabetic dyslipidemia and atherosclerosis: evidence from clinical trials. Curr Diab Rep. 2008;8(1):71–7.
6. Taskinen MR, Borén J. New insights into the pathophysiology of dyslipidemia in type 2 diabetes. Atherosclerosis. 2015;239(2):483–95.
7. Saltiel AR, Olefsky JM. Inflammatory mechanisms linking obesity and metabolic disease. J Clin Invest. 2017;127(1):1–4.
8. Bellia A, Andreadi A, Giudice L, De Taddeo S, Maiorino A, D'Ippolito I, et al. Atherogenic dyslipidemia on admission is Associated with poorer outcome in people with and without diabetes hospitalized for COVID-19. Diabetes Care. 2021;44(9):2149–57.
9. Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, Progress, and possibilities. Clin J Am Soc Nephrol. 2017;12(12):2032–45.
10. Syed-Ahmed M, Narayanan M. Immune Dysfunction and risk of infection in chronic kidney disease. Adv Chronic Kidney Dis. 2019;26(1):8–15.
11. Carlson N, Nelveg-Kristensen KE, Freese Ballegaard E, Feldt-Rasmussen B, Hornum M, Kamper AL, et al. Increased vulnerability to COVID-19 in chronic kidney disease. J Intern Med. 2021;290(1):166–78.
12. Li R, Chen Y, Ritchie MD, Moore JH. Electronic health records and polygenic risk scores for predicting disease risk. Nat Rev Genet. 2020;21(8):493–502.
13. Deo RC. Machine learning in Medicine. Circulation. 2015;132(20):1920–30.
14. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604–12.
15. Guerrero-Romero F, Simental-Mendía LE, González-Ortiz M, Martínez-Abundis E, Ramos-Zavala MG, Hernández-González SO, et al. The product of triglycerides and glucose, a simple measure of insulin sensitivity. Comparison with the euglycemic-hyperinsulinemic clamp. J Clin Endocrinol Metab. 2010;95(7):3347–51.
16. Niroumand S, Khajedaluee M, Khadem-Rezaiyan M, Abrishami M, Juya M, Khodaee G, Dadgarmoghaddam M. Atherogenic Index of Plasma (AIP): a marker of cardiovascular disease. Med J Islam Repub Iran. 2015;29:240.
17. Cabello-Solorzano K, Ortigosa de Araujo I, Peña M, Correia L, Tallón-Ballesteros J A, editors. The impact of data normalization on the accuracy of machine learning algorithms: a comparative analysis. International Conference on Soft Computing Models in Industrial and Environmental Applications; 2023: Springer.
18. Shlens J. A tutorial on principal component analysis. arXiv Preprint arXiv:14041100. 2014.
19. Kotsiantis S. Feature selection for machine learning classification problems: a recent overview. Artif Intell Rev. 2011;42(1):157–76.

20. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006;27(8):861–74.
21. Zhao H, Lai Z, Leung H, Zhang X, Zhao H, Lai Z et al. Linear discriminant analysis. Feature Learn Understanding: Algorithms Appl. 2020:71–85.
22. Van Houwelingen J, Le Cessie S. Logistic regression, a review. Stat Neerl. 1988;42(4):215–32.
23. Burges CJ. A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc. 1998;2(2):121–67.
24. Peterson LE. K-nearest neighbor. Scholarpedia. 2009;4(2):1883.
25. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
26. Chen T, Guestrin C, editors. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016.
27. Tibshirani R. Regression shrinkage and selection via the lasso. J Royal Stat Soc Ser B: Stat Methodol. 1996;58(1):267–88.
28. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of Conventional Statistical methods with Machine Learning in Medicine: diagnosis, Drug Development, and treatment. Med (Kaunas). 2020;56(9).
29. Catapano AL, Pirillo A, Bonacina F, Norata GD. HDL in innate and adaptive immunity. Cardiovasc Res. 2014;103(3):372–83.
30. Claxton AJ, Jacobs DR Jr., Iribarren C, Welles SL, Sidney S, Feingold KR. Association between serum total cholesterol and HIV infection in a high-risk cohort of young men. J Acquir Immune Defic Syndr Hum Retrovirol. 1998;17(1):51–7.
31. Madsen CM, Varbo A, Tybjærg-Hansen A, Frikke-Schmidt R, Nordestgaard BG. U-shaped relationship of HDL and risk of infectious disease: two prospective population-based cohort studies. Eur Heart J. 2018;39(14):1181–90.
32. Aung N, Khanji MY, Munroe PB, Petersen SE. Causal Inference for Genetic Obesity, Cardiometabolic Profile and COVID-19 susceptibility: a mendelian randomization study. Front Genet. 2020;11:586308.
33. Chidambaram V, Kumar A, Majella MG, Seth B, Sivakumar RK, Voruganti D, et al. HDL cholesterol levels and susceptibility to COVID-19. EBioMedicine. 2022;82:104166.
34. Hilser JR, Han Y, Biswas S, Gukasyan J, Cai Z, Zhu R, et al. Association of serum HDL-cholesterol and apolipoprotein A1 levels with risk of severe SARS-CoV-2 infection. J Lipid Res. 2021;62:100061.
35. Wei C, Wan L, Yan Q, Wang X, Zhang J, Yang X, et al. HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry. Nat Metab. 2020;2(12):1391–400.
36. Cho KH, Kim JR, Lee IC, Kwon HJ. Native high-density lipoproteins (HDL) with higher paraoxonase exerts a potent antiviral effect against SARS-CoV-2 (COVID-19), while glycated HDL lost the antiviral activity. Antioxid (Basel). 2021;10(2).
37. Oliveira C, Fournier C, Descamps V, Morel V, Scipione CA, Romagnuolo R, et al. Apolipoprotein(a) inhibits hepatitis C virus entry through interaction with infectious particles. Hepatology. 2017;65(6):1851–64.
38. Trinder M, Walley KR, Boyd JH, Brunham LR. Causal inference for genetically determined levels of high-density lipoprotein cholesterol and risk of Infectious Disease. Arterioscler Thromb Vasc Biol. 2020;40(1):267–78.
39. Trinder M, Wang Y, Madsen CM, Ponomarev T, Bohunek L, Daisely BA, et al. Inhibition of Cholesteryl Ester Transfer Protein Preserves High-Density Lipoprotein Cholesterol and improves survival in Sepsis. Circulation. 2021;143(9):921–34.
40. Lim Y, Lee MH, Lee SK, Jeong S, Han HW. Increased estimated GFR is negatively Associated with the risk of SARS-CoV-2 infection and severe COVID-19 within normal to mildly decreased levels: nested case-control study. J Korean Med Sci. 2023;38(49):e415.
41. Mirijello A, Piscitelli P, de Matthaeis A, Inglese M, D'Errico MM, Massa V et al. Low eGFR is a strong predictor of worse outcome in hospitalized COVID-19 patients. J Clin Med. 2021;10(22).
42. Li Q, Lin M, Deng Y, Huang H. The causal relationship between COVID-19 and estimated glomerular filtration rate: a bidirectional mendelian randomization study. BMC Nephrol. 2024;25(1):21.
43. Kurts C, Panzer U, Anders HJ, Rees AJ. The immune system and kidney disease: basic concepts and clinical implications. Nat Rev Immunol. 2013;13(10):738–53.
44. Muniyappa R, Montagnani M, Koh KK, Quon MJ. Cardiovascular actions of insulin. Endocr Rev. 2007;28(5):463–91.
45. Govender N, Khaliq OP, Moodley J, Naicker T. Insulin resistance in COVID-19 and diabetes. Prim Care Diabetes. 2021;15(4):629–34.
46. Singh B, Saxena A. Surrogate markers of insulin resistance: a review. World J Diabetes. 2010;1(2):36–47.
47. Er LK, Wu S, Chou HH, Hsu LA, Teng MS, Sun YC, Ko YL. Triglyceride glucose-body Mass Index is a simple and clinically useful surrogate marker for insulin resistance in nondiabetic individuals. PLoS ONE. 2016;11(3):e0149731.
48. Lioy B, Webb RJ, Amirabdollahian F. The Association between the Atherogenic Index of Plasma and Cardiometabolic Risk Factors: A Review. Healthcare (Basel). 2023;11(7).
49. Rohani-Rasaf M, Mirjalili K, Vatannejad A, Teimouri M. Are lipid ratios and triglyceride-glucose index associated with critical care outcomes in COVID-19 patients? PLoS ONE. 2022;17(8):e0272000.
50. Ren H, Yang Y, Wang F, Yan Y, Shi X, Dong K, et al. Association of the insulin resistance marker TyG index with the severity and mortality of COVID-19. Cardiovasc Diabetol. 2020;19(1):58.
51. Turgay Yıldırım Ö, Kaya Ş. The atherogenic index of plasma as a predictor of mortality in patients with COVID-19. Heart Lung. 2021;50(2):329–33.
52. Alshammari S, AlMasoudi AS, AlBuhayri AH, AlAtwi HM, AlHwiti SS, Alaidi HM, et al. Effect of COVID-19 on glycemic control, Insulin Resistance, and pH in Elderly patients with type 2 diabetes. Cureus. 2023;15(2):e35390.
53. Manoria PC, Chopra HK, Parashar SK, Dutta AL, Pinto B, Mullasari A, Prajapati S. The nuances of atherogenic dyslipidemia in diabetes: focus on triglycerides and current management strategies. Indian Heart J. 2013;65(6):683–90.
54. Boden G. Obesity, insulin resistance and free fatty acids. Curr Opin Endocrinol Diabetes Obes. 2011;18(2):139–43.
55. Boden G, She P, Mozzoli M, Cheung P, Gumireddy K, Reddy P, et al. Free fatty acids produce insulin resistance and activate the proinflammatory nuclear factor-kappab pathway in rat liver. Diabetes. 2005;54(12):3458–65.
56. Itani SI, Ruderman NB, Schmieder F, Boden G. Lipid-induced insulin resistance in human muscle is associated with changes in diacylglycerol, protein kinase C, and IkappaB-alpha. Diabetes. 2002;51(7):2005–11.
57. Weisberg SP, McCann D, Desai M, Rosenbaum M, Leibel RL, Ferrante AW. Jr. Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest. 2003;112(12):1796–808.
58. Masana L, Correig E, Ibarretxe D, Anoro E, Arroyo JA, Jericó C, et al. Low HDL and high triglycerides predict COVID-19 severity. Sci Rep. 2021;11(1):7217.
59. Zhong P, Wang Z, Du Z. Serum triglyceride levels and related factors as prognostic indicators in COVID-19 patients: a retrospective study. Immun Inflamm Dis. 2021;9(3):1055–60.
60. Reynolds HR, Adhikari S, Pulgarin C, Troxel AB, Iturrate E, Johnson SB, et al. Renin-angiotensin-aldosterone system inhibitors and risk of Covid-19. N Engl J Med. 2020;382(25):2441–8.
61. Vincent JL, Taccone FS. Understanding pathways to death in patients with COVID-19. Lancet Respir Med. 2020;8(5):430–2.
62. Guzik TJ, Hoch NE, Brown KA, McCann LA, Rahman A, Dikalov S, et al. Role of the T cell in the genesis of angiotensin II induced hypertension and vascular dysfunction. J Exp Med. 2007;204(10):2449–60.
63. Norlander AE, Madhur MS, Harrison DG. The immunology of hypertension. J Exp Med. 2018;215(1):21–33.
64. Patrick DM, Van Beusecum JP, Kirabo A. The role of inflammation in hypertension: novel concepts. Curr Opin Physiol. 2021;19:92–8.
65. Touyz RM, Rios FJ, Alves-Lopes R, Neves KB, Camargo LL, Montezano AC. Oxidative stress: a unifying paradigm in hypertension. Can J Cardiol. 2020;36(5):659–70.
66. Siedlinski M, Jozefczuk E, Xu X, Teumer A, Evangelou E, Schnabel RB, et al. White Blood cells and blood pressure: a mendelian randomization study. Circulation. 2020;141(16):1307–17.
67. Youn JC, Yu HT, Lim BJ, Koh MJ, Lee J, Chang DY, et al. Immunosenescent CD8 + T cells and C-X-C chemokine receptor type 3 chemokines are increased in human hypertension. Hypertension. 2013;62(1):126–33.
68. Rodilla E, López-Carmona MD, Cortes X, Cobos-Palacios L, Canales S, Sáez MC, et al. Impact of arterial stiffness on all-cause mortality in patients hospitalized with COVID-19 in Spain. Hypertension. 2021;77(3):856–67.
69. Safar ME, Asmar R, Benetos A, Blacher J, Boutouyrie P, Lacolley P, et al. Interaction between Hypertension and arterial stiffness. Hypertension. 2018;72(4):796–805.

Publisher's note