

Original Paper

# A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study

---

Eric Luellen, MS, MPH

Bioinformatix, Interlochen, MI, United States

**Corresponding Author:**

Eric Luellen, MS, MPH

Bioinformatix

Faculty Lane, Box 3628

Interlochen, MI, 49643

United States

Phone: 1 4129157468

Email: [eluellen@bioinformatix.io](mailto:eluellen@bioinformatix.io)

**Related Articles:**

Preprint: <https://preprints.jmir.org/preprint/23582>

Peer-Review Report by Andy Chang (Reviewer G): <https://med.jmirx.org/2020/1/e24747/>

Peer-Review Report by Eric Abbott (Reviewer H): <https://med.jmirx.org/2020/1/e24765/>

Author Responses to Peer-Review Reports: <https://med.jmirx.org/2020/1/e24739/>

## Abstract

---

**Background:** Approximately 80% of those infected with COVID-19 are immune. They are asymptomatic unknown carriers who can still infect those with whom they come into contact. Understanding what makes them immune could inform public health policies as to who needs to be protected and why, and possibly lead to a novel treatment for those who cannot, or will not, be vaccinated once a vaccine is available.

**Objective:** The primary objectives of this study were to learn if machine learning could identify patterns in the pathogen-host immune relationship that differentiate or predict COVID-19 symptom immunity and, if so, which ones and at what levels. The secondary objective was to learn if machine learning could take such differentiators to build a model that could predict COVID-19 immunity with clinical accuracy. The tertiary purpose was to learn about the relevance of other immune factors.

**Methods:** This was a comparative effectiveness research study on 53 common immunological factors using machine learning on clinical data from 74 similarly grouped Chinese COVID-19-positive patients, 37 of whom were symptomatic and 37 asymptomatic. The setting was a single-center primary care hospital in the Wanzhou District of China. Immunological factors were measured in patients who were diagnosed as SARS-CoV-2 positive by reverse transcriptase-polymerase chain reaction (RT-PCR) in the 14 days before observations were recorded. The median age of the 37 asymptomatic patients was 41 years (range 8-75 years); 22 were female, 15 were male. For comparison, 37 RT-PCR test-positive patients were selected and matched to the asymptomatic group by age, comorbidities, and sex. Machine learning models were trained and compared to understand the pathogen-immune relationship and predict who was immune to COVID-19 and why, using the statistical programming language R.

**Results:** When stem cell growth factor-beta (SCGF- $\beta$ ) was included in the machine learning analysis, a decision tree and extreme gradient boosting algorithms classified and predicted COVID-19 symptom immunity with 100% accuracy. When SCGF- $\beta$  was excluded, a random-forest algorithm classified and predicted asymptomatic and symptomatic cases of COVID-19 with 94.8% AUROC (area under the receiver operating characteristic) curve accuracy (95% CI 90.17%-100%). In total, 34 common immune factors have statistically significant associations with COVID-19 symptoms (all  $c<.05$ ), and 19 immune factors appear to have no statistically significant association.

**Conclusions:** The primary outcome was that asymptomatic patients with COVID-19 could be identified by three distinct immunological factors and levels: SCGF- $\beta$  ( $>127,637$ ), interleukin-16 (IL-16) ( $>45$ ), and macrophage colony-stimulating factor

(M-CSF) (>57). The secondary study outcome was the suggestion that stem-cell therapy with SCGF- $\beta$  may be a novel treatment for COVID-19. Individuals with an SCGF- $\beta$  level >127,637, or an IL-16 level >45 and an M-CSF level >57, appear to be predictively immune to COVID-19 100% and 94.8% (AUROC) of the time, respectively. Testing levels of these three immunological factors may be a valuable tool at the point of care for managing and preventing outbreaks. Further, stem-cell therapy via SCGF- $\beta$  and M-CSF appear to be promising novel therapeutics for patients with COVID-19.

(*JMIRx Med* 2020;1(1):e23582) doi: [10.2196/23582](https://doi.org/10.2196/23582)

## KEYWORDS

infectious disease; SARS-CoV-2; COVID-19; public health; immunity; mass vaccinations; therapeutics; stem-cell growth factor-beta

## Introduction

Asymptomatic patients who are infected with SARS-CoV-2 have neither clinical symptoms nor abnormal chest imaging. However, these patients have the same infectivity as infected patients with symptoms [1]. Moreover, adult asymptomatic patients have been found to have the same viral loads as symptomatic patients [2]. Studies have shown that age appears to influence whether an infected person is susceptible to illness. Those under the age of 20 years have approximately half the morbidity probability as those over the age of 20 [3]. This improbability of becoming ill from SARS-CoV-2 infection is especially interesting because young children have been found to have 10 to 100 times the viral load as older children and adults, and disproportionately remain asymptomatic [4].

Stem cell growth factor-beta (SCGF- $\beta$ ) has been associated with H7N9 (Asian lineage avian influenza A subtype) and disassociated with H5N1 (a highly pathogenic avian influenza) [5,6]. Elevated SCGF- $\beta$  has also been associated with specific disease states of hepatocellular cancer, Chagas disease, cardiomyopathy, inflammation and insulin resistance, and unstable carotid plaques [7-10]. Interleukin-16 (IL-16), the second most important variable in predicting SARS-CoV-2 immunity or resistance, has been strongly associated with asthma [11].

Prior studies on the biomarkers associated with SARS-CoV-2 immune response and morbidity include interferon-gamma (IFN- $\gamma$ ), interferon-beta (IFN- $\beta$ ), and interleukin-8 (IL-8) [12]. Other previous research on immune parameters associated with SARS-CoV-2 severity and prognosis have involved interleukin-1 beta (IL-1 $\beta$ ) and interleukin-6 (IL-6). However, others found reduced immunoglobulin G levels in asymptomatic patients [13,14]. The general finding in prior research regarding the pathogen-immune relationship with SARS-CoV-2 is that symptomatic patients have considerably more inflammation and cytokine storm activity than asymptomatic patients [14].

What has been unknown for SARS-CoV-2 are three questions to which the answers are suggested in this study. First, which immunological variables are statistically significant, and how important is each in predicting asymptomatic status? Second, which of those variables, if any, have a strong negative correlation, or relationship, with disease severity (ie, asymptomatic patients' levels are significantly higher than symptomatic patients)? And third, is there an algorithmic or formulaic model of prognostic biomarkers that can accurately

predict morbidity—who will be asymptomatic if infected, and who is at risk of more severe symptoms and disease progression—and why?

## Methods

This study was based on secondary data published as a supplement in *Nature Medicine* in June 2020 [14]. Therein, immunological factors were measured in 74 patients in the Wanzhou District of China. They were diagnosed as SARS-CoV-2 positive by reverse transcriptase-polymerase chain reaction (RT-PCR) in the 14 days before observations were recorded. The median age of the 37 asymptomatic patients was 41 years (range 8-75 years); 22 were female and 15 were male. For comparison, 37 RT-PCR test-positive patients were selected and matched to the asymptomatic group by age, comorbidities, and sex [14].

In this study, five algorithms, or types, of machine learning—a kind of artificial intelligence employing robust brute-force statistical calculations—were applied to a data set of 74 observations of 34 immunological factors in order to attempt three things: (1) to develop a model to accurately predict which patients will be asymptomatic or symptomatic if infected with SARS-CoV-2; (2) to determine the relative importance of each immunological factor; and (3) to determine if there is any level of a subset of immunological factors that can accurately predict which patients are likely to be immune or resistant to SARS-CoV-2.

Minitab 19, version 19.2020.1 (Minitab LLC), was used to calculate means, 95% CIs, *P* values, and two-sample *t* tests of statistical significance. Correlation coefficients were also computed using Minitab via Spearman rho since the data were distributed nonparametrically. A second classification and regression tree (CART) algorithm was also applied in Minitab to cross-validate decision tree results from R in Rattle. Minitab's CART methodology was initially described by Stanford University and University of California Berkeley researchers in 1984 [15].

The Rattle library, version 5.3.0 (Togaware), in the statistical programming language R, version 3.6.3 (CRAN), was used to apply five machine learning algorithms—a decision tree, extreme gradient boosting (XGBoost), linear logistic model (LLM), random forest, and support vector machine (SVM)—to learn which model, if any, could predict asymptomatic status and how accurately. Rattle randomly partitioned the data to select and train on 80% (*n*=59), validate on 10% (*n*=7), and test

on 10% ( $n=7$ ) of observations. Two evaluation methods were used: (1) plots of linear fits of the predicted versus observed categorization; and (2) a pseudo- $R^2$  measure calculated as the square root of the correlation between the predicted and observed values. Pseudo- $R^2$  measure results were evaluated twice, each using for evaluation data that were held back by being randomly selected during partitioning and averaging the two accuracy findings for the final results.

Rattle's rpart decision tree was also used to identify if any levels of one or more immunological factors could accurately diagnose someone as asymptomatic (ie, via rules). The decision tree results reported here used 20 and 12 as the minimum number of observations necessary in nodes before the split (ie, minimum split). The trees used 7 and 4 as the minimum number of observations in a leaf node (ie, minimum bucket).

The random forest analysis in Rattle began by running a series of differently sized random forest algorithms, ranging from 50 to 500 decision trees, to learn the optimum number of trees to minimize error. Each random forest consisted of a minimum of six variables, which was closest to the square root of the number of statistically significant variables (ie, 34). The lowest error rate was approximately 200 decision trees.

The five machine learning models and CART classification trees were run, including and excluding SCGF- $\beta$  to identify if there were alternative prognostic biomarkers and levels in the immune profile that could accurately classify and predict SARS-CoV-2 immunity.

## Results

In total, 34 of the 53 immunological factors (64.2%) were indicated as statistically significant by  $P$  values  $<.05$  from a Spearman rho correlation. Of those 34 factors, 31 were statistically significant with  $P$  values  $<.01$ . Conversely, 35.9% of the 53 immune factors had no statistically significant association with whether a patient was asymptomatic or symptomatic to SARS-CoV-2.

The 22 factors positively correlated with being symptomatic ranged from a minimum coefficient of 0.205 (monocyte chemotactic protein-3 [MCP-3]) to a maximum of 0.781 (tumor necrosis factor-related apoptosis-inducing ligand [TRAIL]). The 11 factors negatively associated with being symptomatic ranged from a minimum of  $-0.866$  (SCGF- $\beta$ ) to a maximum of  $-0.276$  (interferon alpha-2 [IFN $\alpha$ 2]) (see [Table 1](#)).

**Table 1.** Immunological factors associated with SARS-CoV-2 morbidity ranked by Spearman correlation coefficients with 95% CIs and *P* values.

Immunological factor	Abbreviation	Pairwise Spearman correlation to asymptomatic (0) or symptomatic (1) status	95% CI	<i>P</i> value (<.05 target)
TNF <sup>α</sup> -related apoptosis-inducing ligand	TRAIL	0.781	0.654 to 0.865	<.001
Growth-related oncogene alpha	GRO-α	0.750	0.611 to 0.845	<.001
Macrophage-colony stimulating factor	M-CSF	0.748	0.608 to 0.843	<.001
Interleukin-6	IL-6	0.705	0.549 to 0.813	<.001
Granulocyte-colony-stimulating factor	G-CSF	0.697	0.539 to 0.808	<.001
Interleukin-2	IL-2	0.667	0.499 to 0.787	<.001
Nerve growth factor beta	NGF-β	0.651	0.479 to 0.775	<.001
Interleukin-10	IL-10	0.614	0.431 to 0.748	<.001
Monocyte chemoattractant protein-1	MCP-1	0.594	0.407 to 0.733	<.001
Stem-cell factor	SCF	0.586	0.397 to 0.728	<.001
Interleukin-15	IL-15	0.527	0.325 to 0.683	<.001
Interleukin-8	IL-8	0.514	0.311 to 0.673	<.001
Interferon-gamma	IFN-γ	0.464	0.252 to 0.633	<.001
Interleukin-7	IL-7	0.454	0.240 to 0.625	<.001
Interferon gamma inducible protein-10	INF-γ-IP-10	0.451	0.237 to 0.623	<.001
Interleukin-18	IL-18	0.438	0.223 to 0.613	<.001
Platelet-derived growth factor BB	PDGF-BB	0.436	0.220 to 0.611	<.001
Interleukin-2 receptor alpha	IL-2Rα	0.388	0.166 to 0.572	.001
Immunoglobulin G (convalescing)	IgG Conv	0.366	0.143 to 0.544	.001
Monokine-induced by gamma	MIG	0.364	0.140 to 0.552	.001
Immunoglobulin G (acute)	IgG Acute	0.330	0.103 to 0.524	.004
Macrophage migration inhibitory factor	MIF	0.237	0.006 to 0.444	.04
Monocyte chemotactic protein-3	MCP-3	0.205	-0.270 to 0.416	.08 <sup>b</sup>
Vascular endothelial growth factor	VEGF	0.184	-0.048 to 0.397	.12 <sup>b</sup>
N gene	N	0.180	-0.053 to 0.394	.13 <sup>b</sup>
Interleukin-3	IL-3	0.163	-0.070 to 0.379	.17 <sup>b</sup>
Interleukin-12-p40	IL-12(p40)	0.151	-0.082 to 0.368	.20 <sup>b</sup>
Interleukin-9	IL-9	0.149	-0.084 to 0.366	.21 <sup>b</sup>
Interleukin-1 beta	IL-1β	0.125	-0.107 to 0.345	.29 <sup>b</sup>
Days shed virions	Days shed	0.122	-0.110 to 0.342	.30 <sup>b</sup>
Stromal cell-derived factor-1 alpha	SDF-1α	0.098	-0.124 to 0.320	.41 <sup>b</sup>
Interleukin-12-p70	IL-12(p70)	0.083	-0.149 to 0.306	.48 <sup>b</sup>
Interleukin-17	IL-17	0.067	-0.164 to 0.291	.57 <sup>b</sup>
Interleukin-4	IL-4	0.020	-0.210 to 0.247	.87 <sup>b</sup>
Interleukin-13	IL-13	-0.022	-0.249 to 0.208	.86 <sup>b</sup>
Fibroblast growth factor	FGF	-0.078	-0.302 to 0.153	.51 <sup>b</sup>
Regulated upon activation, normal T-cell expressed and secreted	RANTES	-0.085	-0.308 to 0.146	.47 <sup>b</sup>

Immunological factor	Abbreviation	Pairwise Spearman correlation to asymptomatic (0) or symptomatic (1) status	95% CI	P value (<.05 target)
Macrophage inflammatory protein-1 beta	MIP-1 $\beta$	-0.109	-0.330 to 0.123	.35 <sup>b</sup>
ORF1ab gene	ORF1ab	-0.113	-0.334 to 0.119	.34 <sup>b</sup>
Macrophage inflammatory protein-1 alpha	MIP-1 $\alpha$	-0.138	-0.356 to 0.095	.24 <sup>b</sup>
Tumor necrosis factor-alpha	TNF- $\alpha$	-0.168	-0.383 to 0.065	.15 <sup>b</sup>
Tumor necrosis factor-beta	TNF- $\beta$	-0.197	-0.409 to 0.035	.09 <sup>b</sup>
Interferon alpha-2	IFN $\alpha$ 2	-0.276	-0.478 to -0.046	.02 <sup>c</sup>
Leukemia inhibitory factor	LIF	-0.312	-0.509 to -0.840	.007 <sup>c</sup>
Interleukin-5	IL-5	-0.316	-0.512 to -0.089	.006 <sup>c</sup>
Interleukin-1 alpha	IL-1 $\alpha$	-0.332	-0.526 to -0.106	.004 <sup>c</sup>
Granulocyte-macrophage colony-stimulating factor	GM-CSF	-0.359	-0.548 to -0.134	.002 <sup>c</sup>
Interleukin-1 receptor alpha	IL-1R $\alpha$	-0.359	-0.548 to -0.135	.002 <sup>c</sup>
Eotaxin	Eotaxin	-0.390	-0.576 to -0.169	.001 <sup>c</sup>
Cutaneous T-cell-attracting chemokine	CTACK	-0.456	-0.627 to -0.243	<.001 <sup>c</sup>
Hepatocyte growth factor	HGF	-0.594	-0.733 to -0.407	<.001 <sup>c</sup>
Interleukin-16	IL-16	-0.827	-0.895 to -0.721	<.001 <sup>c</sup>
Stem-cell growth factor-beta	SCGF- $\beta$	-0.866	-0.920 to -0.780	<.001 <sup>c</sup>

<sup>a</sup>TNF: tumor necrosis factor.

<sup>b</sup>Statistically insignificant.

<sup>c</sup>Statistically significant negative correlations.

When SCGF- $\beta$  was included in the machine learning analysis, two algorithms predicted and classified SARS-CoV-2 immunity or resistance by being asymptomatic with 100% accuracy: a decision tree and XGBoost. When SCGF- $\beta$  was excluded, a random-forest algorithm predicted and classified SARS-CoV-2 asymptomatic and symptomatic cases with 94.8% AUROC (area under the receiver operating characteristic) curve accuracy (95% CI 90.17%-100%) (see [Table 2](#)).

Notably, both the rpart decision trees and CART classification trees independently identified three prognostic biomarkers at

specific levels that could classify asymptomatic and symptomatic cases with 95%-100% accuracy. When SCGF- $\beta$  was included, all asymptomatic cases had levels >127,656.8, while all symptomatic cases had levels <127,656.8 ([Figure 1](#)). When SCGF- $\beta$  was excluded, as a type of contingency analysis to understand prognostic biomarker levels in other factors better, IL-16 accurately classified asymptomatic cases >44.59 and symptomatic cases <44.59 in 90.4% of the cases. In the remaining 9.6% of cases where IL-16 >44.59, all had macrophage colony-stimulating factor (M-CSF) >57.13 ([Figure 2](#)).

**Table 2.** Comparative accuracy of six machine learning algorithms in predicting SARS-CoV-2 asymptomatic status from immunological factors.

Machine learning model	Pseudo-R <sup>2</sup> (10% evaluation holdback sample 1) (%)	Pseudo-R <sup>2</sup> (10% evaluation holdback sample 2) (%)	Average Pseudo-R <sup>2</sup> (%)
<b>With SCGF-β<sup>a</sup></b>			
Decision tree	100.00	100.00	100.00
XGBoost <sup>b</sup>	100.00	100.00	100.00
GLM <sup>c</sup> (logistic)	100.00	98.89	99.45
Random forest	99.46	94.83	97.15
SVM <sup>d</sup>	78.81	96.99	87.90
<b>Without SCGF-β</b>			
Random forest	97.68	91.91	94.80
GLM (logistic)	100.00	85.96	92.98
SVM	77.76	89.69	83.73
XGBoost	99.42	54.27	76.85
Decision tree	100.00	2.22	51.11

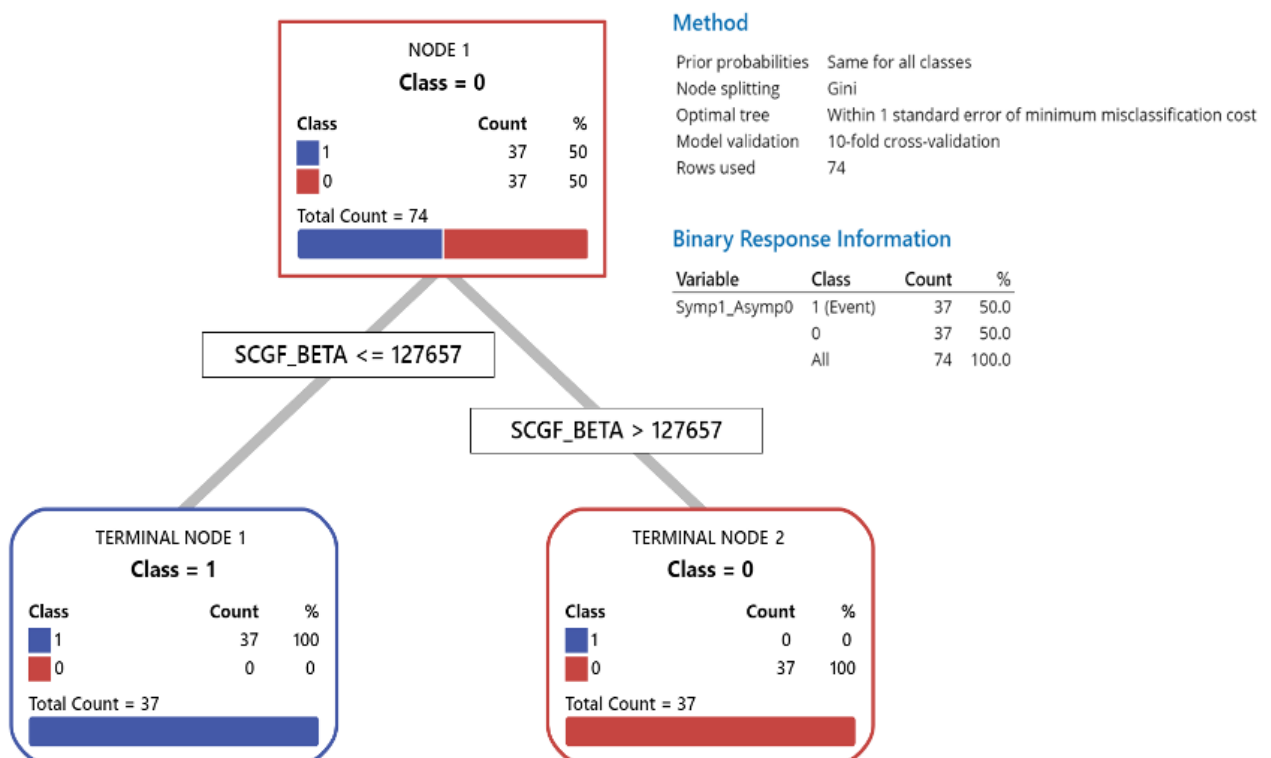
<sup>a</sup>SCGF-β: stem cell growth factor-beta.

<sup>b</sup>XGBoost: extreme gradient boosting.

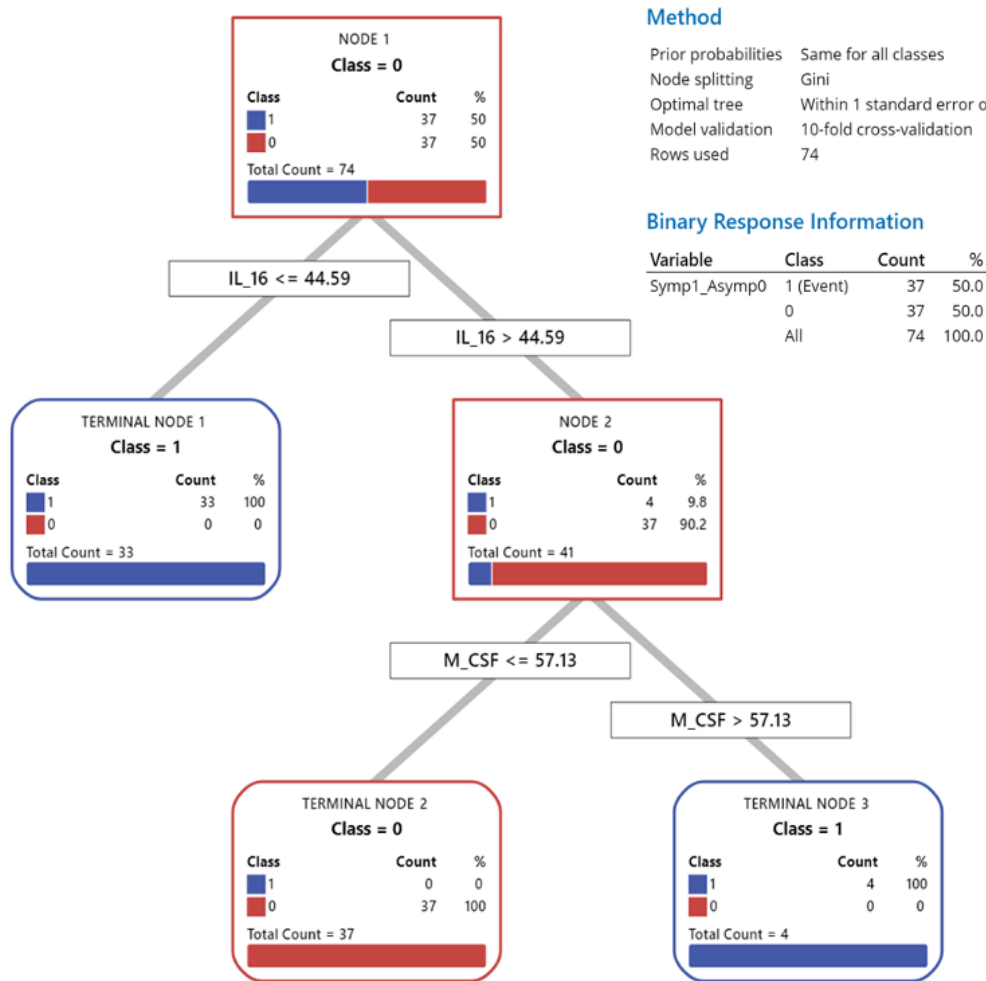
<sup>c</sup>GLM: generalized linear model.

<sup>d</sup>SVM: support vector machine.

**Figure 1.** Classification and regression tree (CART) of the role of stem cell growth factor-beta (SCGF-β) in predicting SARS-CoV-2 morbidity.



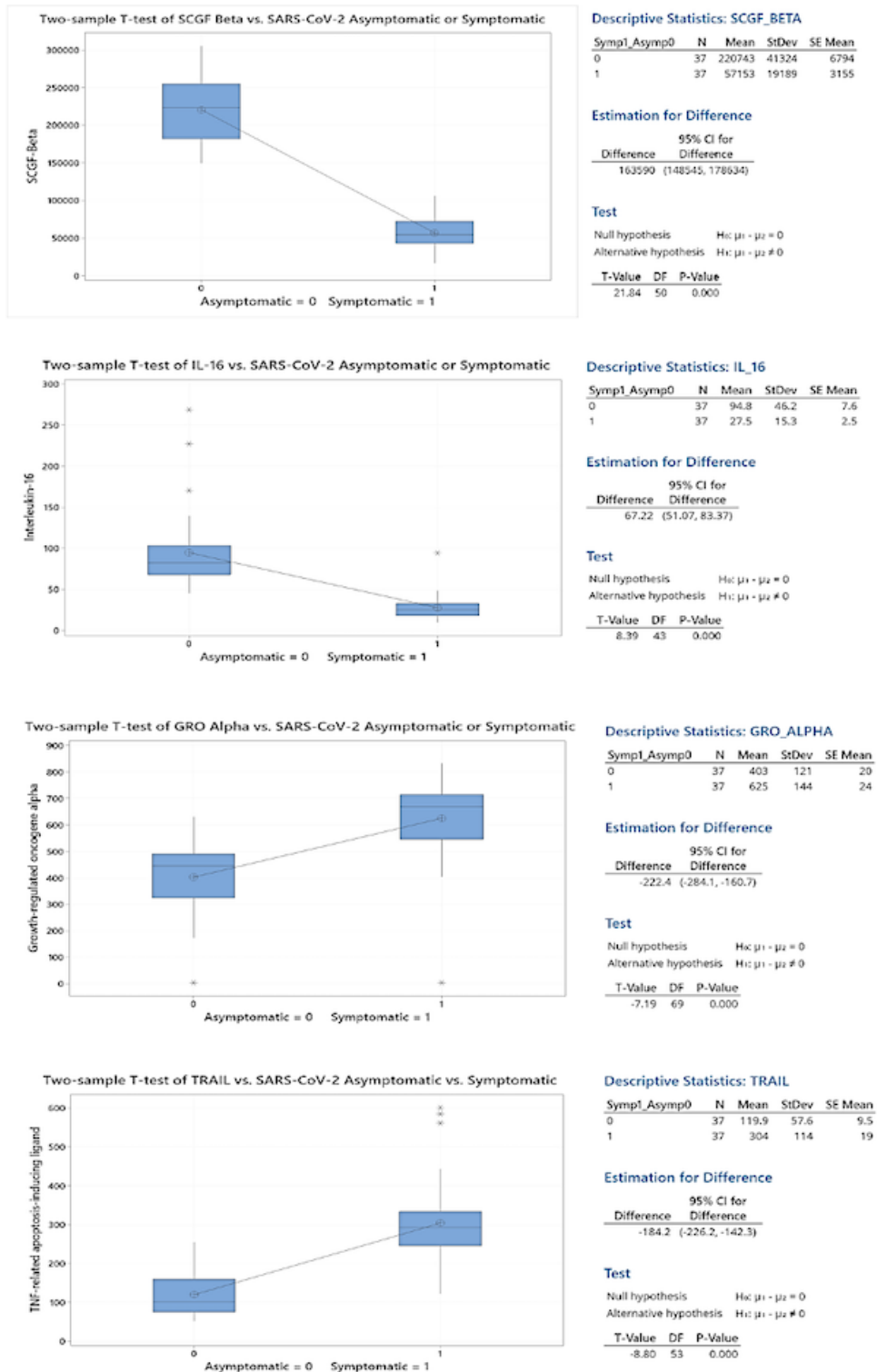
**Figure 2.** Classification and regression tree (CART) of the role of interleukin-16 (IL-16) and macrophage colony-stimulating factor (M-CSF) in predicting SARS-CoV-2 morbidity in the absence of stem cell growth factor-beta (SCGF-β).



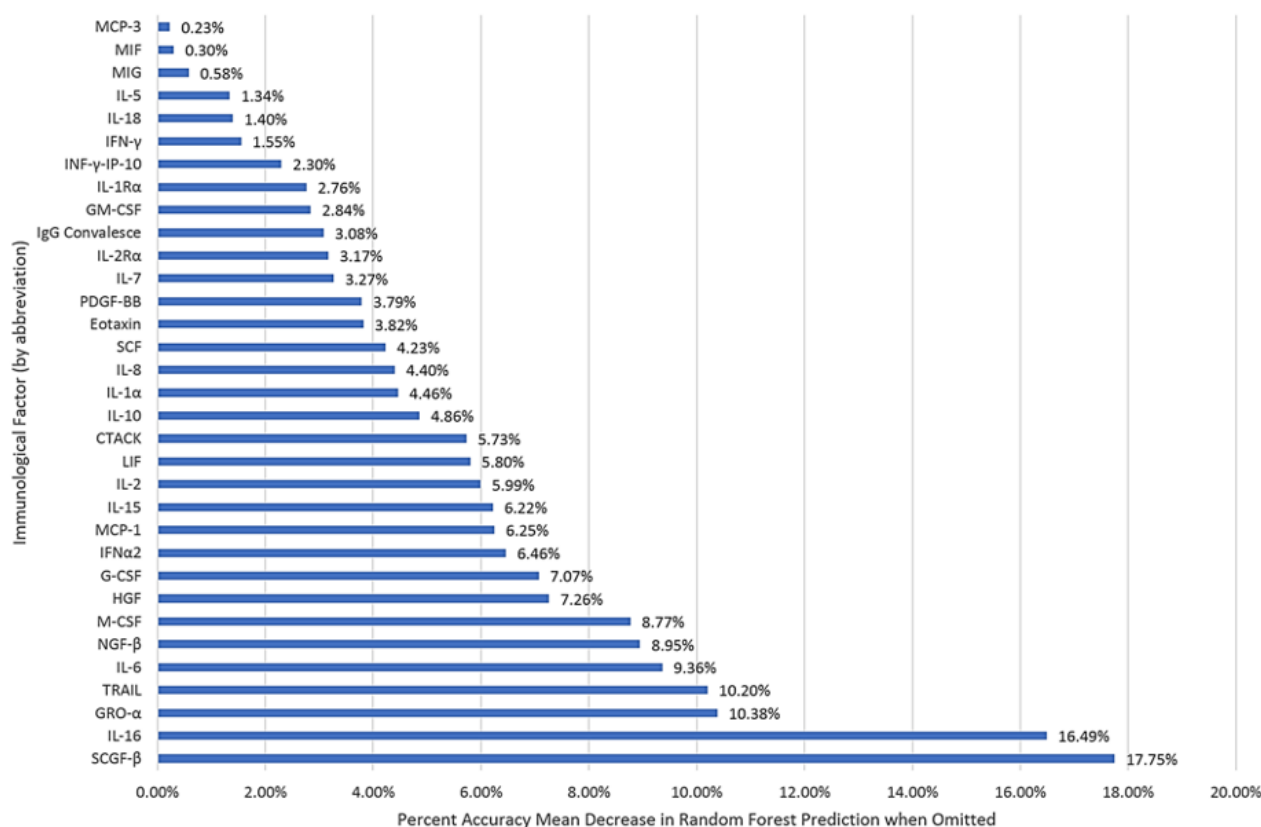
Two-sample *t* tests for the four factors with the highest positive and negative correlation coefficients, interquartile ranges, outliers, and levels between asymptomatic and symptomatic patients that were statistically significant were computed to ordinally rank factors by their correlation coefficients (Figure 3).

A random forest analysis of the most important variables to accurately classify and predict SARS-CoV-2 patients by binary morbidity ordinally ranked the 34 statistically significant factors. Unsurprisingly, SCGF-β, and IL-16, followed by growth-related oncogene alpha (GRO-α) and TRAIL, respectively, were the most critical factors in predicting morbidity (Figure 4).

**Figure 3.** Two-sample t tests of statistical significance of the difference in means of four leading prognostic biomarkers for asymptomatic or symptomatic SARS-CoV-2. SCGF-β: stem cell growth factor-beta; IL-16: interleukin-16; GRO-α: growth-related oncogene alpha; TRAIL: tumor necrosis factor-related apoptosis-inducing ligand.





**Figure 4.** Relative importance of immunological variables from random forest analysis in predicting SARS-CoV-2 morbidity.

Finally, the results suggest that IL-1β, 3, 4, 9, 12, 13, 17, and RANTES (regulated upon activation, normal T-cell expressed and secreted) are of low importance, or comparative irrelevance, in the pathogen-immune relationship and, that SCGF-β, IL-16, HGF, INFNα2, LIF, CTACK, IL-1α, Eotaxin, GM-CSF, IL-1Rα, and IL-5 are valuable in models to predict and classify asymptomatic or symptomatic SARS-CoV-2 cases accurately.

## Discussion

### Principal Findings

While it has been speculated that stem cells may play a role in SARS-CoV-2 and other zoonoses' resistance, prior research has focused on different stem cell involvement than SCGF-β [16-18]. Previous research has also established that stem cells can inhibit viral growth by expressing IFN-γ-stimulated genes and have been particularly effective against influenza A H5N1 virus and resulting lung injuries [19,20]. Stem cell therapy has been hypothesized as a treatment for SARS-CoV-2; however, there is no record in the literature specific as to which factors may influence SARS-CoV-2 infections, favorably or unfavorably, or to what degree until now [21].

Researchers have recently found that symptomatic patients generally have a more robust immune response to SARS-CoV-2 infection, culminating in cytokine storms in the worst cases. Conversely, asymptomatic patients have been found to have a weaker immune response [14]. Because infections are causal to immune response, of particular interest in this study were the most impactful immune-related variables that negatively correlated with asymptomatic status (ie, variables that were

greater for asymptomatic patients than symptomatic patients) (marked with a superscripted "c" in Table 1).

This paper's overarching importance is the identification of immunological factors for diagnoses, treatments, and preclinical prophylactic immune-based approaches to SARS-CoV-2 in the first 7 months of a pandemic that experts now opine will last decades [22]. Immunostimulant approaches are especially valuable because, unlike antivirals and vaccines, they may be given later in the course of the disease to optimize outcomes [21].

The primary importance of this work is machine learning algorithmic models that can predict with high accuracy whether someone, once infected, will be asymptomatic or symptomatic from SARS-CoV-2. This knowledge gives clinicians new tools to identify populations in advance who appear to be at higher risk of danger from the virus. Such devices, especially once reproduced in a more extensive study, may also inform policy decisions as to who needs to shelter in place. Finally, because of the scale of this pandemic and practical constraints as to how many vaccination doses can be manufactured and how quickly this can be done, such tools may become valuable in prioritizing vaccine administration to those in greatest need because they have a higher biological and immunological risk.

This work's secondary importance is a description of the cytokine and chemokine profile that is associated with asymptomatic or symptomatic SARS-CoV-2 infections. It enables a better understanding of the pathogen-immune relationship. These profiles provide insights into the biological pathways critical for SARS-CoV-2 progression.

As one example, stem cell factors secrete multiple factors that regulate immune cells and modulate them to restore tissue homeostasis. These results suggest that higher levels of SCF- $\beta$  (stem-cell factor-beta) may better control immune responses to prevent the more robust reactions universally associated so far with highly symptomatic patients and, further, prevent high morbidity and mortality cytokine storms. A better understanding of the pathogen-immune relationship may enable researchers to prevent and treat patients with SARS-CoV-2 infection more effectively with therapeutics currently untested and unused. This knowledge may also extend to similar zoonotic coronaviruses in the future.

The tertiary importance of this work is identifying three immune factors and precise levels that appear to be prognostic biomarkers as to whether someone, once infected with SARS-CoV-2, will be immune or resistant, as demonstrated by being asymptomatic or not. These insights also suggest new candidates for therapeutic research focused on the relatively newly identified and ill-understood SCGF- $\beta$  and its role in the immunological process.

The quaternary importance of this work is further proof that machine learning methods can accurately and quickly identify critical elements of disease dynamics that accelerate understanding and improve outcomes during pandemics. Moreover, it is an example of how a “dry” data science laboratory can link to clinical or “wet” laboratory science for real-world applications.

### Limitations

This study has several limitations. First, it is unknown from the data set how many days passed between exposure to the virus and immunological testing, or whether it was universally the same number of days. Second, because immune profiles are temporally sensitive, ideally, several tests would have been taken over several days, which did not occur (R Jankord, PhD, July 22, 2020). Third, immunological signaling and processing

are multifactorial and complex. Therefore, it is unclear why SCGF- $\beta$  levels are categorically high in asymptomatic patients and low in symptomatic patients, or whether they are causal to SARS-CoV-2 response. Fourth, combinatorial and sequential analysis of these immunological elements may be an important future research area to optimize therapeutic research outcomes. Fifth, at least one study in a leading journal, *The Lancet*, found that Chinese SARS-CoV-2 case data may have been misreported by as much as 400% [23]. That study, and much higher case and fatality numbers in over 200 countries, have created distrust and skepticism of SARS-CoV-2-related data originating from China.

Future research could ameliorate these limitations and focus on a more extensive study group to attempt to reproduce the results. Moreover, a prospective case-control study of patients with decreased SCGF- $\beta$  levels and supplementation that was protective against SARS-CoV-2 severity and symptoms would be invaluable validation.

### Conclusion

One implication of these findings is that if we can predict the 80% of society who may be immune or resistant to SARS-CoV-2, or asymptomatic, it may profoundly impact public health intervention decisions as to who needs to be protected and by how much. If, for example, 80% of the shelter-in-place orders and the resultant dramatic reduction in economic and social activity could have been prevented by accurately predicting who is at low risk of infection, the economic benefits alone may have been valued in US\$ trillions. The second implication of these findings is evidence that elevated levels of SCGF- $\beta$ , IL-16, and M-CSF may have a causal relationship with SARS-CoV-2 immunity or resistance, and may have utility as diagnostic determinants to (1) inform public health policy decisions to prioritize and reduce shelter-in-place orders to minimize economic and social impacts; (2) advance therapeutic research; and (3) prioritize vaccine distribution to benefit those with the greatest need and risks first.

---

### Acknowledgments

The author wishes to thank Dr Luka Fajs for his helpful comments on drafts of this paper.

---

### Conflicts of Interest

None declared.

---

### References

1. Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S, et al. A Systematic Review of Asymptomatic Infections with COVID-19. *J Microbiol Immunol Infect* 2020 May 15:e [FREE Full text] [doi: [10.1016/j.jmii.2020.05.001](https://doi.org/10.1016/j.jmii.2020.05.001)] [Medline: [32425996](https://pubmed.ncbi.nlm.nih.gov/32425996/)]
2. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med* 2020 Mar 19;382(12):1177-1179 [FREE Full text] [doi: [10.1056/NEJMc2001737](https://doi.org/10.1056/NEJMc2001737)] [Medline: [32074444](https://pubmed.ncbi.nlm.nih.gov/32074444/)]
3. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med* 2020 Aug;26(8):1205-1211. [doi: [10.1038/s41591-020-0962-9](https://doi.org/10.1038/s41591-020-0962-9)] [Medline: [32546824](https://pubmed.ncbi.nlm.nih.gov/32546824/)]
4. Heald-Sargent T, Muller WJ, Zheng X, Rippe J, Patel AB, Kociolek LK. Age-Related Differences in Nasopharyngeal Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Levels in Patients With Mild to Moderate Coronavirus Disease 2019 (COVID-19). *JAMA Pediatr* 2020 Jul 30:e. [doi: [10.1001/jamapediatrics.2020.3651](https://doi.org/10.1001/jamapediatrics.2020.3651)] [Medline: [32745201](https://pubmed.ncbi.nlm.nih.gov/32745201/)]

5. Guo J, Huang F, Liu J, Chen Y, Wang W, Cao B, et al. The Serum Profile of Hypercytokinemia Factors Identified in H7N9-Infected Patients can Predict Fatal Outcomes. *Sci Rep* 2015 Jun 01;5:10942 [FREE Full text] [doi: [10.1038/srep10942](https://doi.org/10.1038/srep10942)] [Medline: [26028236](https://pubmed.ncbi.nlm.nih.gov/26028236/)]
6. de Jong MD, Simmons CP, Thanh TT, Hien VM, Smith GJD, Chau TNB, et al. Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nat Med* 2006 Oct;12(10):1203-1207 [FREE Full text] [doi: [10.1038/nm1477](https://doi.org/10.1038/nm1477)] [Medline: [16964257](https://pubmed.ncbi.nlm.nih.gov/16964257/)]
7. Sukowati CHC, Patti R, Pascut D, Ladju RB, Tarchi P, Zanotta N, et al. Serum Stem Cell Growth Factor Beta for the Prediction of Therapy Response in Hepatocellular Carcinoma. *Biomed Res Int* 2018;2018:6435482 [FREE Full text] [doi: [10.1155/2018/6435482](https://doi.org/10.1155/2018/6435482)] [Medline: [30246025](https://pubmed.ncbi.nlm.nih.gov/30246025/)]
8. Wang Y, Khan A, Heringer-Walther S, Schultheiss H, Moreira MDCV, Walther T. Prognostic value of circulating levels of stem cell growth factor beta (SCGF beta) in patients with Chagas' disease and idiopathic dilated cardiomyopathy. *Cytokine* 2013 Mar;61(3):728-731. [doi: [10.1016/j.cyto.2012.12.018](https://doi.org/10.1016/j.cyto.2012.12.018)] [Medline: [23357302](https://pubmed.ncbi.nlm.nih.gov/23357302/)]
9. Schiro A, Wilkinson FL, Weston R, Smyth JV, Serracino-Inglott F, Alexander MY. Elevated levels of endothelial-derived microparticles, and serum CXCL9 and SCGF- $\beta$  are associated with unstable asymptomatic carotid plaques. *Sci Rep* 2015 Nov 13;5:16658 [FREE Full text] [doi: [10.1038/srep16658](https://doi.org/10.1038/srep16658)] [Medline: [26564003](https://pubmed.ncbi.nlm.nih.gov/26564003/)]
10. Tarantino G, Citro V, Balsano C, Capone D. Could SCGF-Beta Levels Be Associated with Inflammation Markers and Insulin Resistance in Male Patients Suffering from Obesity-Related NAFLD? *Diagnostics (Basel)* 2020 Jun 11;10(6):e [FREE Full text] [doi: [10.3390/diagnostics10060395](https://doi.org/10.3390/diagnostics10060395)] [Medline: [32545215](https://pubmed.ncbi.nlm.nih.gov/32545215/)]
11. Mathy NL, Scheuer W, Lanzendörfer M, Honold K, Ambrosius D, Norley S, et al. Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes. *Immunology* 2000 May;100(1):63-69. [doi: [10.1046/j.1365-2567.2000.00997.x](https://doi.org/10.1046/j.1365-2567.2000.00997.x)] [Medline: [10809960](https://pubmed.ncbi.nlm.nih.gov/10809960/)]
12. Geifman N, Whetton AD. A consideration of publication-derived immune-related associations in Coronavirus and related lung damaging diseases. *J Transl Med* 2020 Aug 03;18(1):297 [FREE Full text] [doi: [10.1186/s12967-020-02472-z](https://doi.org/10.1186/s12967-020-02472-z)] [Medline: [32746922](https://pubmed.ncbi.nlm.nih.gov/32746922/)]
13. Jesenak M, Brndiarova M, Urbancikova I, Rennerova Z, Vojtkova J, Bobcakova A, et al. Immune Parameters and COVID-19 Infection - Associations With Clinical Severity and Disease Prognosis. *Front Cell Infect Microbiol* 2020;10:364 [FREE Full text] [doi: [10.3389/fcimb.2020.00364](https://doi.org/10.3389/fcimb.2020.00364)] [Medline: [32695683](https://pubmed.ncbi.nlm.nih.gov/32695683/)]
14. Long QX, Tang XJ, Shi QL, Li Q, Deng HJ, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* 2020 Aug;26(8):1200-1204. [doi: [10.1038/s41591-020-0965-6](https://doi.org/10.1038/s41591-020-0965-6)] [Medline: [32555424](https://pubmed.ncbi.nlm.nih.gov/32555424/)]
15. Breiman L, Friedman J, Olshen R, Stone C. *Classification and Regression Trees*. Boca Raton, Florida: Chapman & Hall/CRC; 1984.
16. Yu F, Jia R, Tang Y, Liu J, Wei B. SARS-CoV-2 infection and stem cells: Interaction and intervention. *Stem Cell Res* 2020 Jul;46:101859 [FREE Full text] [doi: [10.1016/j.scr.2020.101859](https://doi.org/10.1016/j.scr.2020.101859)] [Medline: [32570174](https://pubmed.ncbi.nlm.nih.gov/32570174/)]
17. Golchin A, Seyedjafari E, Ardehshirylajimi A. Mesenchymal Stem Cell Therapy for COVID-19: Present or Future. *Stem Cell Rev Rep* 2020 Jun;16(3):427-433 [FREE Full text] [doi: [10.1007/s12015-020-09973-w](https://doi.org/10.1007/s12015-020-09973-w)] [Medline: [32281052](https://pubmed.ncbi.nlm.nih.gov/32281052/)]
18. Chrzanowski W, Kim SY, McClements L. Can Stem Cells Beat COVID-19: Advancing Stem Cells and Extracellular Vesicles Toward Mainstream Medicine for Lung Injuries Associated With SARS-CoV-2 Infections. *Front Bioeng Biotechnol* 2020;8:554 [FREE Full text] [doi: [10.3389/fbioe.2020.00554](https://doi.org/10.3389/fbioe.2020.00554)] [Medline: [32574317](https://pubmed.ncbi.nlm.nih.gov/32574317/)]
19. Wu X, Dao Thi VL, Huang Y, Billerbeck E, Saha D, Hoffmann H, et al. Intrinsic Immunity Shapes Viral Resistance of Stem Cells. *Cell* 2018 Jan 25;172(3):423-438.e25 [FREE Full text] [doi: [10.1016/j.cell.2017.11.018](https://doi.org/10.1016/j.cell.2017.11.018)] [Medline: [29249360](https://pubmed.ncbi.nlm.nih.gov/29249360/)]
20. Chan MCW, Kuok DIT, Leung CYH, Hui KPY, Valkenburg SA, Lau EHY, et al. Human mesenchymal stromal cells reduce influenza A H5N1-associated acute lung injury in vitro and in vivo. *Proc Natl Acad Sci U S A* 2016 Mar 29;113(13):3621-3626. [doi: [10.1073/pnas.1601911113](https://doi.org/10.1073/pnas.1601911113)] [Medline: [26976597](https://pubmed.ncbi.nlm.nih.gov/26976597/)]
21. Florindo HF, Kleiner R, Vaskovich-Koubi D, Acúrcio RC, Carreira B, Yeini E, et al. Immune-mediated approaches against COVID-19. *Nat Nanotechnol* 2020 Aug;15(8):630-645 [FREE Full text] [doi: [10.1038/s41565-020-0732-3](https://doi.org/10.1038/s41565-020-0732-3)] [Medline: [32661375](https://pubmed.ncbi.nlm.nih.gov/32661375/)]
22. Farrar J. Coronavirus: 'Infection here for many years to come'. BBC. 2020 Jul 21. URL: <https://www.bbc.com/news/uk-53488142> [accessed 2020-07-23]
23. Tsang T, Wu P, Lin Y, Lau E, Leung G, Cowling B. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *The Lancet Public Health* 2020 May;5(5):e289-e296. [doi: [10.1016/S2468-2667\(20\)30089-X](https://doi.org/10.1016/S2468-2667(20)30089-X)]

## Abbreviations

- AUROC:** area under the receiver operating characteristic
- GRO- $\alpha$ :** growth-related oncogene alpha
- IFN- $\beta$ :** interferon-beta
- IFN- $\gamma$ :** interferon-gamma
- IFN $\alpha$ 2:** interferon alpha-2

**IL-1 $\beta$** : interleukin-1 beta  
**IL-6**: interleukin-6  
**IL-8**: interleukin-8  
**IL-16**: interleukin-16  
**LLM**: linear logistic model  
**M-CSF**: macrophage colony-stimulating factor  
**MCP-3**: monocyte chemotactic protein-3  
**RANTES**: regulated upon activation, normal T-cell expressed and secreted  
**RT-PCR**: reverse transcriptase-polymerase chain reaction  
**SCF- $\beta$** : stem-cell factor-beta  
**SCGF- $\beta$** : stem cell growth factor-beta  
**SVM**: support vector machine  
**TRAIL**: tumor necrosis factor-related apoptosis-inducing ligand  
**XGBoost**: extreme gradient boosting

*Edited by G Eysenbach; submitted 26.08.20; peer-reviewed by A Chang, E Abbott; comments to author 23.09.20; revised version received 24.09.20; accepted 30.09.20; published 19.10.20*

*Please cite as:*

*Luellen E*

*A Machine Learning Explanation of the Pathogen-Immune Relationship of SARS-CoV-2 (COVID-19), and a Model to Predict Immunity and Therapeutic Opportunity: A Comparative Effectiveness Research Study*

*JMIRx Med 2020;1(1):e23582*

*URL: <https://med.jmirx.org/2020/1/e23582/>*

*doi: [10.2196/23582](https://doi.org/10.2196/23582)*

*PMID: [33711083](https://pubmed.ncbi.nlm.nih.gov/33711083/)*

©Eric Luellen. Originally published in JMIRx Med (<https://med.jmirx.org>), 19.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the JMIRx Med, is properly cited. The complete bibliographic information, a link to the original publication on <https://med.jmirx.org/>, as well as this copyright and license information must be included.