Biological Data 2026, 1(1), 1-19.
https://doi.org/????

Original Article

Association Rule Mining Reveals IgG–Symptom Relationships in COVID-19 Patients: Insights for Personalized Medicine

Hamid Ghaznavi[1*]

[1] Department of Radiology, Kurdistan University of Medical Sciences, Sanandaj, Iran

***** Corresponding Author: Hamid Ghaznavi
Email: ghaznavi2691@gmail.com

CITATION

Ghaznavi, H. (2026). Association rule mining reveals IgG–symptom relationships in COVID-19 patients: Insights for personalized medicine. Biological Data Journal, 1(1), 1–19. https://doi.org/xxxxxxxxx.

ARTICLE INFO

Received: February 23, 2026
Revised: March 22, 2026
Accepted: April 9, 2026
Available online: April 12, 2026

COPYRIGHT

© The author(s) retain the copyright.
Publisher: University of Kurdistan Press.
This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract: The COVID-19 pandemic has underscored the variability in immune responses and symptom presentations due to SARS-CoV-2, with Immunoglobulin G (IgG) levels critical for long-term immunity. This study explores the relationship between IgG levels and COVID-19 symptoms using association rule mining (ARM) to address a research gap. A cross-sectional study was conducted during the third wave of COVID-19 in western Iran, enrolling 237 RT-PCR-confirmed patients, excluding those with autoimmune conditions or recent infections. Data on demographics, IgG levels (measured via ELISA), and symptoms (e.g., fever, cough, loss of smell) were analyzed using the ORANGE software for machine learning with the Apriori algorithm, applying 90% confidence, 0.001 support, and lift > 1 thresholds. Significant associations emerged: low IgG levels correlated with milder symptoms (e.g., absence of cough, tiredness), moderate IgG with intermediate symptoms (e.g., loss of smell without fever), and high IgG with severe clusters (e.g., fever, cough). Younger females (20-40 years) and blood group A were linked to higher IgG, while older patients (>65 years) with blood group O and Rh-negative status showed lower IgG. Confidence reached 100%, with lift values up to 3.607. These findings enhance understanding of immune response dynamics, offering insights for personalized diagnostics, treatment, and long COVID management. ARM application highlights its potential for diagnostic modeling in infectious diseases, improving public health preparedness.

Keywords: COVID-19, IgG Levels, Association Rule Mining, Symptoms, Personalized Medicine, Immune Response

1. Introduction

In The COVID-19 pandemic has resulted in an increase in the rate of hospitalizations due to pneumonia and multisystem organ failure caused by the SARS-CoV-2 virus (Park, 2020). The clinical presentation of COVID-19 varies widely, ranging from asymptomatic cases to severe manifestations, including respiratory distress and death (Alimohamadi et al., 2020; Silva et al., 2020). Common symptoms reported by patients include fever, persistent dry cough, and shortness of breath, with severity often escalating in critical cases (Altuna et al., 2021; Sobhani et al., 2021). The immune response to SARS-CoV-2 is mediated by Immunoglobulin M (IgM), which appears early in infection, and Immunoglobulin G (IgG), which provides long-term immunity by neutralizing the virus and facilitating immune recognition (Abril et al., 2024; Jia et al., 2021). Whether IgG or IgM are important in diagnosis depends on when they are tested and when the relevant infection occurred. Testing for IgM early in an infection timeline is more valuable in terms of current or recently as a part of the total infection process; and testing for IgG is more informative for distinct past infections or immunity information (Bastos et al., 2020). Recent studies have demonstrated that IgG levels may fluctuate dynamically across patient demographics and disease stages, underscoring the need for advanced analytical methods to decode these patterns (Xu et al., 2024). Additionally, IgG is being explored as a biomarker for prolonged immune activation, which could be crucial for post-acute care management (Vojdani et al., 2024). Assessing the link between IgG levels and COVID-19 symptoms have three important reasons to undertake this analysis. First, it might shed light on the immune response dynamics to SARS-CoV-2, or the body's response to the virus. In fact, high IgG levels correlate with more severe disease and robust immune response. Such details can help predict severity and patient outcomes (de Oliveira et al., 2024; Marklund et al., 2020). Second, determining which symptoms were specifically associated with high IgG levels, could inform future treatment and vaccine plans. Fever, body aches, and low appetite might suggest more substantial IgG levels, which indicates the symptoms acted as a marker for a strong immune response (Amjadi et al., 2021). Lastly, understanding the connection and relationship between IgG levels and symptoms might improve diagnostics and management for long COVID, as persistent symptoms might correlate with IgG (Binswanger et al., 2024). Such knowledge is vital both for public health practice and for improving the acute care of patients. Despite these advancements, the application of computational methods to investigate the specific relationship between IgG levels and the varied symptom profiles of COVID-19 remain largely unexplored. Traditional analyses often focus on individual symptoms or broad immune markers, missing nuanced interactions. Association rule mining (ARM), a data mining technique, has emerged as a powerful tool for uncovering hidden patterns in healthcare data (Lakshmi & Vadivu, 2017), with applications in identifying symptom clusters in COVID-19 (Ye et al., 2024) and enhancing predictive modeling (AbdulLateef et al., 2024). Previous studies have successfully used ARM to reveal prevalent symptom combinations, yet few have integrated IgG levels as a key variable (Dehghani & Yazdanparast, 2023; Singh et al., 2023). However, as highlighted in a recent review, artificial intelligence offers transformative potential in infectious disease diagnosis and prognosis. AI algorithms can analyze diverse datasets, including clinical symptoms and laboratory results, to significantly enhance early detection and personalized treatment strategies (Hormaty et al., 2025). AI-driven models improve diagnostic accuracy and predict patient outcomes, addressing challenges in traditional methods, such as delayed detection and limited accuracy (Hudu et al., 2025). Integrating AI with techniques like ARM could provide a more nuanced understanding of the relationship between IgG levels and COVID-19 symptoms, potentially revealing patterns that traditional methods might miss. This study addresses this gap by leveraging ARM to uncover novel IgG-symptom associations, offering insights into immune response dynamics and their clinical implications. Specifically, it demonstrates ARM's potential for predictive modeling in infectious

diseases, enhances diagnostic frameworks for disease severity and long COVID management, and provides guidance for personalized treatment strategies and being prepared for future pandemics. Therefore, this study leverages ARM to investigate the dependencies between IgG levels and COVID-19 symptoms, aiming to provide actionable insights for clinical management, long COVID strategies, and being prepared for future pandemics.

2. Related Works

Research on COVID-19 has leveraged data-driven techniques to explore symptom patterns, immune responses, and resource management, providing a foundation for understanding the disease's complexity. This section reviews relevant studies, focusing on their methodologies and findings, to highlight the gap addressed by the current research. Several studies have employed ARM to uncover symptom patterns in COVID-19. Tandan et al. analyzed the Wolfram Data Repository, identifying fever, cough, pneumonia, sore throat, and respiratory difficulties as prevalent symptoms, revealing common symptom clusters (Tandan et al., 2021). Similarly, Singh et al. used ARM on GitHub Data Repository data, finding that fever, cough, and malaise/body pain frequently co-occurred, aiding in early diagnosis and patient management (Singh et al., 2023). Matharaarachchi et al. applied ARM to Twitter data, uncovering brain fog, fatigue, lung problems, heart issues, depression, and general pains as key symptoms, particularly linked to long COVID, demonstrating the value of social media for symptom trend analysis (Matharaarachchi et al., 2022). Sogandi combined ARM with supervised machine learning on 3,490 Kaggle patient records, identifying symptom clusters like fever, cough, and fatigue, and showing how hybrid methods can enhance diagnostic accuracy (Sogandi, 2024). These studies demonstrate ARM's effectiveness in identifying symptom relationships but largely overlook the role of immune markers like IgG. Other research has focused on immune markers, including IgG, to predict COVID-19 outcomes. Zhuang et al. developed a predictive algorithm for severe COVID-19 in the elderly using IgG subtypes and clinical features, with logistic regression outperforming XGBoost for early risk identification (Zhuang et al., 2023). Alathari et al. proposed a hybrid deep learning model combining convolutional neural networks and bi-directional long short-term memory) to detect IgG antibodies, achieving 89% accuracy and highlighting its potential for diagnostic applications (Alathari et al., 2024). Khudhur et al. used machine learning to predict infection timing based on IgG and IgM levels, achieving 80–100% accuracy, underscoring the diagnostic value of antibody profiles (Khudhur & Khudhur, 2022). Sun et al. introduced the NEAR model (Neutrophil, Eosinophil, and IgA Ratio), which achieved an AUC of 0.961 for predicting COVID-19 severity, aiding timely interventions (Sun et al., 2021). Elkin et al. employed XGBoost to predict COVID-19 test outcomes using IgG, IgM, IgA levels, and demographic data from 2,467 donors, achieving 76.85% accuracy and an AUC of 81.4%, demonstrating the utility of immune markers in screening (Elkin & Zhu, 2022). Kurano et al. applied the Apriori algorithm to a Kaggle dataset, combining supervised and unsupervised methods to identify symptom patterns, reinforcing ARM's role in improving diagnostic accuracy (Kurano et al., 2022). While these studies highlight the importance of IgG and other immune markers, they primarily focus on severity prediction or antibody detection, not symptom associations. Beyond symptom and immune analysis, some studies have explored broader applications. Riazi et al. used K-means clustering to analyze immunoglobulin (Ig) utilization patterns among 4,112 recipients, identifying six patient clusters with distinct usage characteristics (e.g., Cluster 1 accounted for 62.2% of Ig use), offering insights for resource allocation during shortages (Riazi et al., 2023).

Table 1. Summary of computational methods on COVID-19 symptoms.

Study	Data Source	Symptoms Analyzed	Methodology	Features Used in ARM/ML	Findings
(Tandan et al., 2021)	Wolfram Data Repository	Fever, cough, pneumonia, sore throat, respiratory difficulties	ARM (Apriori)	Symptom co-occurrence patterns	Identified common symptom clusters in COVID-19 patients.
(Singh et al., 2023)	GitHub Data Repository	Fever, cough, malaise/body pain	ARM (Apriori)	Frequency of symptom co-occurrences	Revealed frequent symptom co-occurrences, aiding early diagnosis.
(Matharaarachchi et al., 2022)	Twitter data	Brain fog, fatigue, lung problems, heart issues, depression, pains	ARM (Apriori)	Symptom patterns from social media data	Highlighted long COVID symptoms via social media trends.
(Dehghani & Yazdanparast, 2023)	Patient records (2,875)	Apnea, cough, fever, weakness, myalgia, sore throat	ARM (Apriori)	Symptom co-occurrence patterns	Identified prevalent symptoms for treatment optimization.
(Sogandi, 2024)	Kaggle repository (3,490)	Fever, cough, fatigue, myalgia, headache	ARM + supervised ML	Symptom patterns with hybrid methods	Enhanced diagnostic accuracy through combined ML approaches.
(Zhuang et al., 2023)	Patient records	N/A (focused on severity)	Logistic regression, XGBoost	IgG subtypes, clinical features (e.g., IL-2)	Logistic regression best predicted severe COVID-19 in the elderly.
(Alathari et al., 2024)	Fiber optic data	N/A (focused on IgG detection)	CNN-BiLSTM	SARS-CoV-2 IgG antibodies	Achieved 89% accuracy in IgG detection for COVID-19 diagnosis.
(Khudhur & Khudhur, 2022)	Patient records	N/A (focused on infection timing)	Machine learning (unspecified)	IgG, IgM levels	80–100% accuracy in predicting infection timing using antibody levels.
(Sun et al., 2021)	Patient records	N/A (focused on severity)	NEAR model	Neutrophil, eosinophil, IgA levels	NEAR model (AUC 0.961) identified high-risk patients for severe COVID-19.
(Elkin & Zhu, 2022)	Donor dataset (2,467)	Fever, chills, sore throat	XGBoost	IgG, IgM, IgA, age, gender, symptoms	Achieved 76.85% accuracy in predicting COVID-19 test outcomes.
(Riazi et al., 2023)	Patient records (4,112)	N/A (focused on Ig utilization)	K-means clustering	Ig utilization patterns	Identified patient clusters for optimized Ig resource allocation.

This work underscores the potential of machine learning to optimize healthcare logistics, though it does not directly address clinical manifestations like symptoms or immune responses. While existing studies have advanced our understanding of COVID-19 through ARM and machine learning, a notable gap remains in exploring the specific association between IgG levels and symptom profiles. Symptom-focused studies using ARM (Sogandi, 2024; Dehghani & Yazdanparast, 2023; Singh et al., 2023; Matharaarachchi et al., 2022; Tandan et al., 2021) provide valuable insights into clinical presentations but rarely incorporate immune markers, while immune-focused studies (Alathari et al., 2024; Zhuang et al., 2023; Elkin & Zhu, 2022; Khudhur & Khudhur, 2022; Kurano et al., 2022; Sun et al., 2021) prioritize severity prediction over symptom relationships. This lack of integration limits the ability to understand how IgG levels correlate with diverse symptom presentations, a critical factor for diagnostics, treatment personalization, and long COVID management. Motivated by this gap, the current study employs ARM to uncover hidden associations between IgG levels and COVID-19 symptoms, aiming to enhance patient management and inform public health strategies. Table 1 summarizes the key elements of the reviewed studies.

3. Methodology and Data Sets

3.1. Problem Formulation and Data Collection

Let $D = \{x_i\}_{i=1}^{N}$ be the dataset of N=237 identified patient records. Each record x_i contains demographic attributes (age, sex, blood group, Rh), binary symptom indicators (fever, cough, body pain, tiredness, loss of smell, loss of taste, lung involvement on CT), and a serological measure of IgG. The primary, exploratory objective is to discover interpretable association rules of the form $X \rightarrow Y$ that relate itemsets X (combinations of symptoms and demographic attributes) to IgG categories $Y \in \{\text{Low}, \text{Moderate}, \text{High}\}$. ARM (association rule mining) is used for hypothesis generation (to surface subgroup patterns) rather than for building a predictive production model. This cross-sectional study was conducted in western Iran during the third COVID-19 wave (July–September 2021). The dataset analyzed here derives from a prior IRB-approved project; ethical approval for the original data collection and secondary analyses was granted by the Kurdistan University of Medical Sciences Ethics Committee, Sanandaj, Iran (Approval No. IR.MUK.REC.1399.313) (Ghaderi et al., 2024). All participants in the original study provided informed consent; data used here were anonymized prior to analysis. Inclusion criteria: symptomatic patients with RT-PCR confirmed SARS-CoV-2 infection (nasal/pharyngeal swab). Exclusion criteria: negative RT-PCR, autoimmune disease, or documented non-COVID infection within the previous three months. After applying inclusion/exclusion criteria the final dataset included N = 237 patients.

3.2. Serological Testing

With informed consent, 3 mL of peripheral blood was drawn from each patient into serum separator tubes (SSTs). Anti-COVID-19 IgG levels were measured using a commercial enzyme-linked immunosorbent assay (ELISA) kit (GenScript SARS-CoV-2 Spike IgG ELISA kit manufactured by GenScript Biotech Corporation, China, Catalog No. L00844). An indirect ELISA was performed according to the manufacturer's instructions to detect SARS-CoV-2 anti-S IgG antibodies. Briefly, 96-well ELISA plates pre-coated with purified SARS-CoV-2 antigens were used. After adding 10 µL of serum to each well, the plates were incubated at 37°C for 30 minutes. Following a series of washes, horseradish peroxidase-labeled anti-human IgG antibody was added, and incubation continued for 20 minutes. The reaction was stopped, and optical densities were measured at 450 nm. A cutoff value for positivity was calculated as the mean absorbance of negative control samples (obtained from pre-pandemic sera) plus two standard deviations. IgG positivity was determined according to the

manufacturer's guideline, using the cutoff formula: mean OD of negative control wells + 0.1, which aligns with GenScript's recommended procedure for distinguishing non-reactive from reactive samples in SARS-CoV-2 ELISA assays. IgG levels were categorized into low, moderate, and high based on tertiles of the optical density distribution: low (<0.5), moderate (0.5–1.0), and high (>1.0).

3.3. Data Management and Analysis

The ORANGE data mining toolkit was employed for data preprocessing, analysis, and ARM (Demšar et al., 2013). ORANGE's visual programming interface facilitated efficient data management without requiring extensive coding expertise. The toolkit was used to process raw patient data from medical records and serological tests. Data cleaning was performed using ORANGE's preprocessing tools. Missing values were imputed using the mean for numerical variables (e.g., IgG levels) and the mode for categorical variables (e.g., symptoms), while outliers were identified and removed using the interquartile range (IQR) method (values beyond 1.5 × IQR). This ensured the dataset was consistent and suitable for analysis. The dataset comprised demographic and clinical features, including age, sex, IgG levels, Rh factor, blood groups, and symptoms such as fever, cough, body pain, tiredness, loss of smell and taste, and lung involvement (assessed via CT). Age was categorized into three groups (20–40, 40–65, >65 years) based on clinical relevance, and IgG levels were categorized as low, moderate, and high (as described in Section 3.2) to facilitate ARM analysis. Sample size was chosen to achieve 80% power at $\alpha = 0.05$ to detect medium-sized associations between categorized IgG levels and symptom clusters. Sample size estimation was performed using G*Power v3.1 (Faul et al., 2007). The calculation parameters were: test family= χ^2 tests; statistical test=contingency tables (goodness-of-fit / independence); effect size w = 0.20 (Cohen's small–medium), α= .05, power (1–β)=0.80, degrees of freedom=2 (three IgG categories). Under these parameters G*Power estimates a required total sample of N ≈ 219. We enrolled N=237 patients, which exceeds the calculated requirement and thus provides the planned power for detection of the specified effect size. Association rule mining was conducted using ORANGE's ARM functionality (Orange software, version 3.38.1), which implements the Apriori algorithm. This analysis identified significant patterns and correlations between categorized IgG levels and symptoms, providing insights into the relationship between immune response and clinical presentation in COVID-19 patients. By incorporating both demographic and clinical data, the analysis enabled a comprehensive exploration of symptom clusters and their dependence on IgG levels.

3.4. Association Rule Mining (ARM)

ARM identifies patterns of frequently occurring items or events within a dataset, including relationships between them (Zhang & Zhang, 2002). These patterns reveal combinations of items or events that often occur together. In healthcare, this can be useful for understanding how diseases are linked, such as the connection between diabetes and hypertension. Association rules are expressed as "$X \rightarrow Y$," where X and Y are distinct sets of symptoms or diseases (X does not overlap with Y). X is the antecedent (the "if" part) and Y is the consequent (the "then" part). The effectiveness of discovered rules is evaluated using support, confidence, and lift.

Support: measures how frequently a rule occurs in the dataset.
Support ($X \rightarrow Y$) = $P(X \cap Y)$

- $P(X \cap Y)$: Probability of both X and Y occurring.

Confidence: indicates how often Y appears in individuals who have X.
Confidence (X → Y) = P(Y | X) = P(X ∩ Y) / P(X)

- P(Y | X): Probability of Y given X.

Lift: determines how often Y appears when X appears, considering the overall frequency of Y.
Lift (X → Y) = P(Y | X) / P(Y)

- X is the antecedent (the "if" part of the rule).
- Y is the consequent (the "then" part of the rule).

A lift value of 1 suggests no correlation, greater than 1 indicates a positive relationship, and less than 1 indicates a negative relationship. One limitation of the confidence measure is that it can overestimate the importance of an association if the consequent (Y) is very common. Lift addresses this by considering the overall frequency of Y. For example, in our study, a rule such as "High IgG, Fever → Cough" might have a support of 15% (15% of patients exhibit this combination), a confidence of 80% (80% of patients with high IgG and fever also have a cough), and a lift of 2.5 (the likelihood of cough is 2.5 times higher in patients with high IgG and fever compared to the general population). The ARM algorithm was applied to identify frequent item sets meeting user-defined thresholds. A confidence threshold of 0.9 (90%) was chosen to ensure high reliability of the rules, a minimum support threshold of 0.001 was set to capture even rare but meaningful patterns given the dataset size, and a lift > 1 was required to focus on positively correlated rules (Mutter et al., 2004)**.** The Apriori algorithm in ORANGE iteratively generated frequent item sets by first identifying single items meeting the support threshold, then combining them into larger sets, and finally deriving rules that met the confidence and lift criteria (Zakur & Flaih, 2023). Only the top 10 rules with the highest support scores are presented to highlight the most frequent and impactful associations. This approach enabled the identification of patterns between categorized IgG levels (low, moderate, high) and symptoms, providing insights into immune response dynamics in COVID-19. Figure 2 illustrated ARM Process for exploring IgG-symptom associations in COVID-19**.** This flowchart outlines the methodology to uncover associations between IgG levels and symptoms in 237 RT-PCR-confirmed COVID-19 patients. The process starts with a dataset including demographics, IgG levels (via ELISA), and symptoms. Data preprocessing cleans the data (e.g., imputing missing values) and categorizes IgG into low, moderate, and high tertiles. The Apriori algorithm, applied via the ORANGE toolkit with support ≥ 0.001, confidence ≥ 90%, and lift > 1, generates association rules. Results include rules like “If Cough=Yes and Fever=Yes, then IgG=High” (Support: 0.004, Confidence: 1.000, Lift: 2.788), linking symptoms to IgG levels. Figure 1, demonstrates the overall workflow of the research. We selected a high confidence threshold (≥ 0.90) to emphasize rules that would be most trustworthy in a clinical context. This increases specificity of the reported rules at the expense of sensitivity (fewer, but more reliable rules). The minimum support threshold (0.001) was intentionally low to allow detection of rare but potentially meaningful patterns in a modest-sized dataset; however, low support increases the chance that a rule is idiosyncratic to the sample. Lift values > 1 were used to indicate positive association strength; we caution readers that high lift in combination with very low support (e.g., subgroup n = 10) may not generalize and should be interpreted as hypothesis-generating rather than confirmatory.

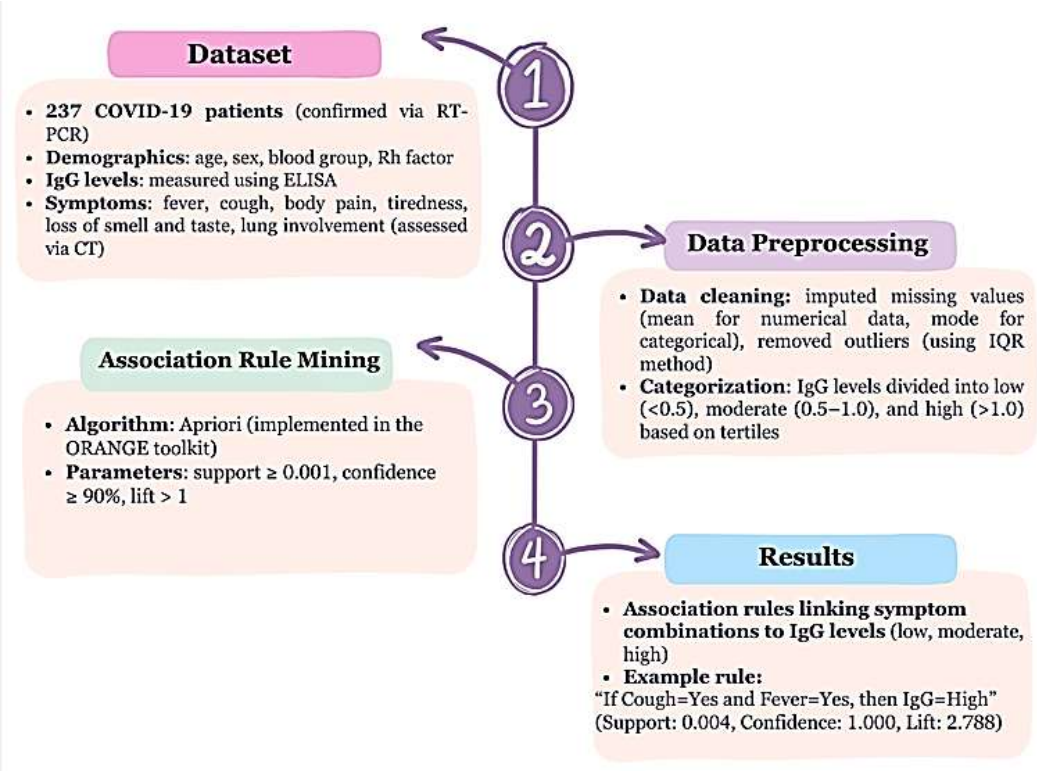


Figure 1. Illustrated workflow of association rule mining process for exploring igg-symptom associations in COVID-19.

4. Results

4.1. Demographic Characteristics and Symptom Prevalence

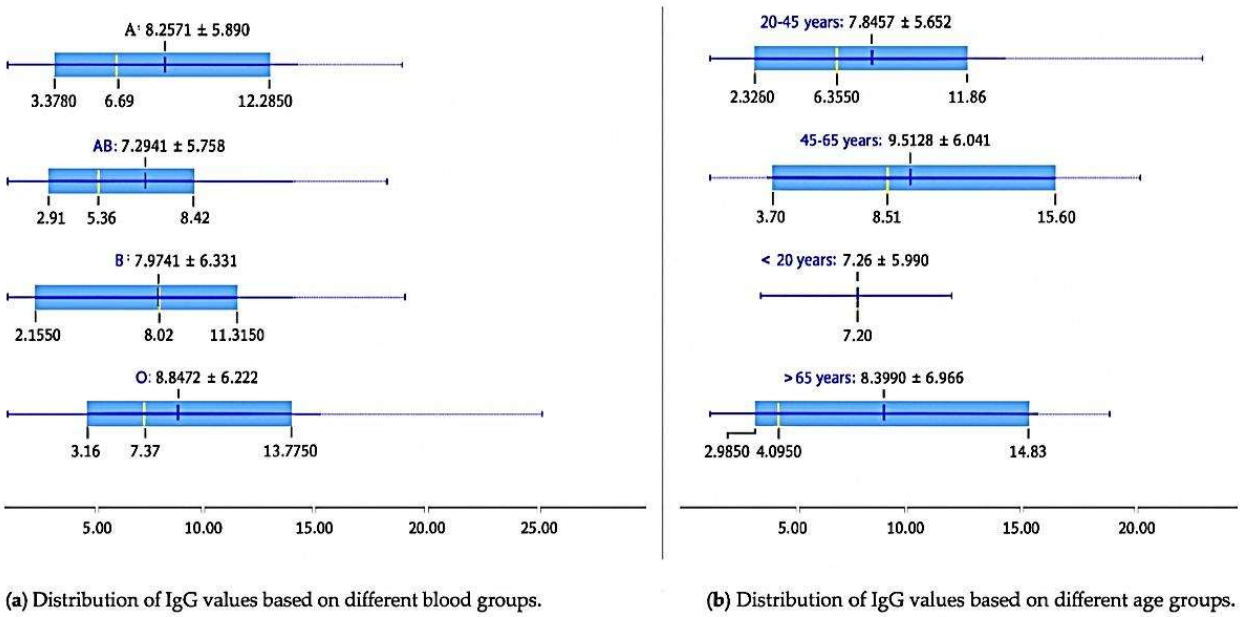
A total of 237 RT-PCR–confirmed COVID-19 patients were included in the analysis. The demographic, clinical, and serological characteristics of the cohort are summarized in Table 2. The median age was 40 years (IQR 31–54), and the majority of participants were between 20 and 40 years old (65%). Males accounted for 61.6% of the cohort (n = 146). Blood group distribution showed that types A (35.4%) and O (32.1%) were most common, and 89% of participants were Rh-positive. Fever (77.2%), cough (66.7%), and tiredness (62.9%) were the most frequent symptoms reported. Loss of smell (45.6%) and loss of taste (43.9%) were also prominent. IgG levels showed substantial variability across individuals, with a mean 7.06 ± 5.88 and a median of 5.10 (IQR 2.40–9.90). Categorization of IgG into tertiles resulted in nearly equal distribution of low, moderate, and high IgG groups, supporting subsequent association rule mining analyses.

4.2. IgG Levels and Statistical Associations

The mean IgG level was 7.06 ± 5.88 units (confidence interval for the mean is 6.31- 7.81), ranging from 1.12 to 24.59 units. A Mann-Whitney U test revealed significantly higher IgG levels in males (mean = 7.82, SD ± 6.01) compared to females (mean = 5.92, SD ± 5.52; p = 0.004), indicating sex-based differences in immune response. A chi-square test showed a significant association between higher IgG levels (categorized as moderate or high) and the presence of cough (p = 0.048), though this p-value was not adjusted for multiple comparisons. Figure 2 illustrates the distribution of IgG levels across blood groups and age groups, showing higher median IgG levels in the 40–65 age group (median = 8.12, IQR = 4.23–12.56) and among patients with blood group O (median = 7.45, IQR = 3.89–11.32) compared to other groups.

Table 2. Demographic, clinical, and serological characteristics of the study population.

Variable	Category / Statistic	Value
Age (years)	Mean ± SD	42.8 ± 14.7
	Median (IQR)	40 (31–54)
	Age < 20	n = 1 (0.4%)
	Age 20–40	n = 154 (65.0%)
	Age 40–65	n = 72 (30.4%)
	Age > 65	n = 10 (4.2%)
Sex	Male	n = 146 (61.6%)
	Female	n = 91 (38.4%)
Blood group	A	n = 84 (35.4%)
	B	n = 59 (24.9%)
	AB	n = 18 (7.6%)
	O	n = 76 (32.1%)
Rh factor	Rh-positive	n = 211 (89.0%)
	Rh-negative	n = 26 (11.0%)
Common symptoms	Fever	183 (77.2%)
	Cough	158 (66.7%)
	Tiredness	149 (62.9%)
	Body pain	127 (53.6%)
	Loss of smell	108 (45.6%)
	Loss of taste	104 (43.9%)
	Lung involvement (CT)	92 (38.8%)
IgG	Mean ± SD	7.06 ± 5.88
	Median (IQR)	5.10 (2.40–9.90)
	Range	0.35–32.7
IgG category	Low (<0.5)	n = 78 (32.9%)
	Moderate (0.5–1.0)	n = 79 (33.3%)
	High (>1.0)	n = 80 (33.8%)



(a) Distribution of IgG values based on different blood groups. (b) Distribution of IgG values based on different age groups.

Figure 2. IgG distribution by blood groups (a) and age groups (b).

4.3. Symptom Rules Across IgG Levels

ARM was applied to evaluate the relationship between categorized IgG levels (low, moderate, high) and symptoms, revealing distinct patterns (Tables 3a–3c). Confidence values of 1.000 indicate that all patients meeting the antecedent criteria exhibited the consequent IgG level, while lift values > 1 (e.g., 3.537) suggest a strong positive association, meaning the likelihood of the consequent is 3.537 times higher than expected by chance. For patients with low IgG levels (Table 3a), rules consistently showed milder symptom profiles, such as the absence of cough combined with loss of smell and taste and tiredness (support=0.008, confidence=1.000, lift=3.537). This suggests that patients with these symptoms may have a diminished immune response, potentially impacting diagnostic sensitivity for active infection. Patients with moderate IgG levels (Table 3b) exhibited intermediate symptom patterns, with the highest lift for rules involving the absence of fever but presence of loss of smell and taste and body pain (support=0.004, confidence=1.000, lift=2.788). This indicates a mixed symptom profile, reflecting an intermediate immune response. High IgG levels (Table 3c) were associated with more severe symptom clusters, such as cough, fever, and loss of smell and taste (support=0.004, confidence=1.000, lift=2.788). For example, patients with fever and loss of smell and taste but without body pain showed a strong association with high IgG levels (support=0.013, confidence=0.750, lift=2.091), suggesting a robust immune response linked to diverse symptoms. These findings align with the statistical association between higher IgG levels and cough ($p = 0.048$), reinforcing the ARM results. Findings highlight the complex relationships between IgG levels and symptomatology in patients. Low IgG levels were consistently associated with milder or more specific symptom combinations, potentially indicating an impaired immune response. In contrast, higher IgG levels were linked to more diverse symptoms, reflecting the immune system's active role in combating infection. The clear distinctions between these groups, particularly in the consistent patterns observed for both low and high IgG levels, provide valuable insights for clinical management and patient monitoring.

Table 3a. Top 5 rules for Low IgG levels.

Antecedents (Symptoms)	Consequent (IgG Level)	Support	Confidence	Lift
Cough=No, Loss of smell and taste=Yes, Tiredness=Yes	Low	0.008	1.000	3.537
Cough=No, Fever=Yes, Loss of smell and taste=Yes	Low	0.008	1.000	3.537
Fever=Yes, Loss of smell and taste=Yes, Body Pain=Yes	Low	0.008	1.000	3.537
Fever=Yes, Loss of smell and taste=Yes, Tiredness=No, Body Pain=No	Low	0.008	1.000	3.537
Cough=Yes, Fever=Yes, Loss of smell and taste=Yes	Low	0.008	1.000	3.537

4.4. Age-Specific Patterns

ARM identified age-related patterns in symptom presentation and IgG levels (Table 4). The 20–40 age group (n=154) was the most represented, showing diverse symptoms like cough, fever, and loss of smell, often with moderate to high IgG levels (e.g., Fever=Yes, Body Pain=No, IgG=Moderate; support=0.025, confidence=1.000, lift=1.55). The 40–65 age group (n=72) exhibited a mix of moderate and severe symptoms, with rules like Loss of smell=Yes, Body Pain=No, IgG=Moderate (support=0.004, confidence=1.000, lift = 3.29), indicating stronger associations (higher lift) compared to the 20–40 group. The >65 age group (n=10) had lower support values due to its small size, but rules suggested milder symptoms, such as Body Pain=No, Rh=Negative, Blood Group=B

(support=0.005, confidence=0.500, lift=13.75), with high lift values reflecting the rarity of this demographic. Notably, the <20 age group (n=1) had insufficient data for meaningful ARM analysis and is not included in Table 4.

Table 3b. Top 5 rules for moderate IgG levels.

Antecedents (Symptoms)	Consequent (IgG Level)	Support	Confidence	Lift
Fever=No, Loss of smell and taste=Yes, Tiredness=No	Moderate	0.004	1.000	2.788
Cough=No, Fever=No, Loss of smell and taste=Yes, Body Pain=No	Moderate	0.004	1.000	2.788
Cough=Yes, Fever=Yes, Loss of smell and taste=No, Tiredness=No	Moderate	0.004	1.000	2.788
Cough=Yes, Fever=Yes, Tiredness=No, Body Pain=Yes	Moderate	0.004	1.000	2.788
Cough=No, Fever=No, Loss of smell and taste=Yes, Tiredness=No, Body Pain=No	Moderate	0.004	1.000	2.788

Table 3c. Top 5 rules for high IgG levels.

Antecedents (Symptoms)	Consequent (IgG Level)	Support	Confidence	Lift
Cough=Yes, Loss of smell and taste=Yes, Tiredness=Yes, Body Pain=No	High	0.008	1.000	2.788
Cough=No, Fever=Yes, Loss of smell and taste=Yes, Tiredness=No, Body Pain=No	High	0.008	1.000	2.788
Cough=Yes, Fever=Yes, Loss of smell and taste=Yes	High	0.004	1.000	2.788
Fever=Yes, Loss of smell and taste=Yes, Body Pain=No	High	0.013	0.750	2.091
Cough=Yes, Loss of smell and taste=Yes, Tiredness=Yes, Body Pain=Yes	High	0.017	0.571	1.593

Table 4. Rules for different age ranges.

Antecedent	Consequent	Support	Confidence	Lift
Cough=Yes, Tiredness=No, Blood Group=A	Age 20–40 years	0.017	1.000	1.55
Fever=Yes, Body Pain=No, IgG=Moderate	Age 20–40 years	0.025	1.000	1.55
Cough=Yes, Tiredness=No, IgG=High	Age 20–40 years	0.013	1.000	1.55
Fever=Yes, Loss of smell=Yes, Body Pain=No	Age 20–40 years	0.021	1.000	1.55
Cough=Yes, Fever=No, Loss of smell=Yes	Age 20–40 years	0.017	1.000	1.55
Cough=Yes, Body Pain=No, Blood Group=A	Age 40–65 years	0.004	1.000	3.29
Fever=No, Tiredness=Yes, Blood Group=B	Age 40–65 years	0.004	1.000	3.29
Loss of smell=Yes, Body Pain=No, IgG=Moderate	Age 40–65 years	0.004	1.000	3.29
Rh=Negative, Fever=No, Tiredness=Yes	Age 40–65 years	0.004	1.000	3.29
Sex=Male, Rh=Negative, Blood Group=O	Age 40–65 years	0.004	1.000	3.29
Body Pain=No, Rh=Negative, Blood Group=B	Age >65 years	0.005	0.500	13.75
Fever=No, Body Pain=No, Rh=Negative	Age >65 years	0.005	0.333	9.17
Loss of smell=No, Body Pain=No, IgG=Moderate	Age >65 years	0.005	0.250	6.87
Fever=No, IgG=Moderate, Rh=Negative	Age >65 years	0.005	0.250	6.87
Tiredness=No, IgG=Moderate, Blood Group=B	Age >65 years	0.005	0.250	6.87

4.5. IgG Levels and Demographic Factors

ARM identified patterns between IgG levels and demographic factors (Table 5). Key findings include a strong association between low IgG levels and females aged >65 years with blood group O (support= 0.009, confidence=1.000, lift=3.607), suggesting a diminished immune response in this subgroup. Moderate IgG levels were strongly linked to Rh-negative patients with blood group A, particularly in the 20–45 age group (support=0.014, confidence=1.000, lift=2.821). High IgG levels were associated with Rh-negative patients with blood group O, especially males aged 20–45 years (support=0.009, confidence=1.000, lift=2.716). These patterns suggest that demographic factors like sex, age, and blood group may influence immune response, with females and younger patients more likely to exhibit higher IgG levels, consistent with the statistical finding of higher IgG in males ($p = 0.004$), which may reflect differences in symptom severity or exposure.

Table 5. Rules for IgG levels and demographic factors.

Antecedent	Consequent	Support	Confidence	Lift
Female, Age >65 years	IgG Categorized=Low	0.009	1.000	3.607
Female, Rh-positive, Age >65 years	IgG Categorized=Low	0.009	1.000	3.607
Female, Blood Group=O, Age >65 years	IgG Categorized=Low	0.009	1.000	3.607
Female, Rh-positive, Blood Group=O	IgG Categorized=Low	0.086	0.463	1.671
Female, Blood Group=O	IgG Categorized=Low	0.086	0.442	1.594
Rh-negative, Blood Group=A	IgG Categorized=Moderate	0.023	1.000	2.821
Male, Rh-negative, Blood Group=A	IgG Categorized=Moderate	0.014	1.000	2.821
Female, Rh-negative, Blood Group=A	IgG Categorized=Moderate	0.009	1.000	2.821
Rh-negative, Blood Group=A, Age 20-45 years	IgG Categorized=Moderate	0.014	1.000	2.821
Female, Rh-negative, Age 45-65 years	IgG Categorized=Moderate	0.009	1.000	2.821
Rh-negative, Blood Group=O	IgG Categorized=High	0.014	1.000	2.716
Male, Rh-negative, Blood Group=O	IgG Categorized=High	0.005	1.000	2.716
Female, Rh-negative, Blood Group=O	IgG Categorized=High	0.009	1.000	2.716
Rh-negative, Blood Group=O, Age 20-45 years	IgG Categorized=High	0.009	1.000	2.716
Male, Rh-negative, Age 45-65 years	IgG Categorized=High	0.009	1.000	2.716

5. Discussion

This study leveraged ARM to uncover novel associations between IgG levels and COVID-19 symptoms, providing insights into immune response dynamics during the third wave of the pandemic in western Iran (July–September 2021). Our findings reveal distinct symptom profiles linked to low, moderate, and high IgG levels, with implications for diagnostics, treatment personalization, and public health strategies. These results build on prior ARM-based symptom analyses (Singh et al., 2023; Matharaarachchi et al., 2022; Tandan et al., 2021) by integrating IgG as a key immunological variable, addressing a gap in understanding how antibody responses correlate with clinical presentations. Patients with low IgG levels exhibited milder symptom profiles, such as the absence of cough paired with loss of smell and taste or tiredness (confidence=1.000, lift=3.537). This suggests a potentially diminished immune response, which may reflect early-stage infection, immune suppression, or lower viral load—hypotheses warranting further investigation. Low IgG levels could also indicate a more efficient early viral clearance driven by innate immunity, as suggested by research indicating that a rapid innate response can limit viral replication and subsequent adaptive immune activation (Lucas et al., 2020). Conversely, high IgG levels were associated with severe symptom clusters, including fever, cough, and loss of smell (confidence=1.000, lift=2.788), aligning with prior evidence that robust IgG responses correlate with greater disease severity (de Oliveira et

al., 2024; Marklund et al., 2020). This observation is consistent with studies demonstrating that excessive inflammation and immune dysregulation contribute to severe COVID-19, potentially driving higher IgG production as a consequence of prolonged antigen exposure (Fajgenbaum & June, 2020). Moderate IgG levels showed intermediate patterns, such as loss of smell without fever (confidence=1.000, lift=2.788), indicating a transitional immune state. These graded associations underscore IgG's role as a biomarker for disease progression and severity, extending beyond its established function in long-term immunity (de Oliveira et al., 2024; Sołkiewicz et al., 2024; Ahmed et al., 2023). Demographic factors further modulated these patterns. Younger females (20–40 years) and those with blood group A exhibited higher IgG levels, consistent with statistical findings of sex-based differences (p=0.004), while older patients (>65 years) with blood group O and Rh-negative status showed lower IgG. These trends align with studies suggesting age and blood group influence immune responses to SARS-CoV-2 (Zhuang et al., 2023; Sun et al., 2021) though our ARM approach uniquely ties these factors to specific symptom clusters. Specifically, research suggests that estrogen may enhance humoral immunity, potentially explaining higher IgG levels in younger females (Klein & Flanagan, 2016). For instance, the strong association between low IgG and females >65 years with blood group O (lift=3.607) highlights a vulnerable subgroup that may require targeted monitoring or intervention. This finding resonates with studies indicating that older adults exhibit immunosenescence, characterized by a decline in immune function and responsiveness to vaccination (Nikolich-Žugich, 2018). Further, individuals with blood group O have a lower risk of coronavirus infection, a trend seen with SARS-CoV and confirmed in COVID-19, possibly due to the Spike protein's poor binding to O antigens and anti-A antibody neutralization. Conversely, blood group A increases SARS-CoV-2 infection risk due to Spike-A antigen binding. However, ABO blood groups show no clear link to COVID-19 severity or outcomes, and no research has yet examined their role in long COVID (Franchini et al., 2024). Although males showed higher average IgG levels (p = 0.004), several ARM rules linked high IgG to younger females. These results are not contradictory—they simply reflect different types of analysis. The statistical test compares overall IgG averages between sexes, while ARM looks for specific combinations of symptoms and demographics that tend to appear together. It is possible that a few males with very high IgG increased the overall male mean, while certain symptom patterns were more common among females aged 20–40, placing them in the "high IgG" tertile used for ARM. Differences in group sizes, the use of IgG tertiles instead of continuous values, and confounding factors such as symptoms or blood group may also contribute. Because of this, ARM findings should be viewed as subgroup signals rather than population-wide effects. Future work using multivariable models and sex-stratified ARM will help clarify these relationships. Compared to prior ARM studies, our work advances the field by integrating immune markers with symptom data (Singh et al., 2023; Tandan et al., 2021). identified prevalent symptom combinations (e.g., fever, cough, malaise), but lacked immunological context. Similarly, Matharaarachchi et al. linked ARM to long COVID symptoms via social media, yet did not explore antibody levels. Our study bridges this gap, demonstrating ARM's utility in decoding complex IgG–symptom interactions. The high confidence (up to 100%) underscores the reliability of these findings (Alathari et al., 2024; Khudhur & Khudhur, 2022). It is important to note that while high confidence values indicate strong associations within our dataset, they may also reflect the homogeneity of our study population or the specific parameter settings used in the ARM analysis. Therefore, these findings should be interpreted cautiously and validated in other cohorts. Some of the ARM rules in our study showed a confidence of 1.000, which indicates that all patients who exhibited the antecedent symptoms also fell into the corresponding IgG category within our dataset. While this appears strong statistically, perfect confidence often arises in small or highly specific subgroups, making such rules more sensitive to sampling variation than true biological certainty. Similarly, lift

values greater than 1 suggest a positive association, but lift can be artificially inflated when support is low, as seen in older age groups with limited sample sizes. Thus, rules with high confidence but low support should be interpreted as preliminary signals rather than robust clinical patterns. Additionally, because this study is cross-sectional, the relationships identified represent correlations only; we cannot infer whether particular symptoms cause changes in IgG levels or vice versa. Longitudinal follow-up would be required to confirm the temporal and causal nature of these associations. Clinically, these findings have several implications. First, symptom-IgG patterns could enhance diagnostic frameworks by identifying patients at risk of severe outcomes (high IgG) versus those with milder disease (low IgG), complementing RT-PCR and imaging. Second, the association of high IgG with severe symptoms like fever and cough could guide treatment strategies, such as prioritizing anti-inflammatory therapies in these patients. Third, demographic insights—e.g., higher IgG in younger females—may inform risk stratification and vaccination prioritization, particularly in resource-limited settings like western Iran during the third wave. Specifically, in resource-constrained settings, these ARM-derived rules could be incorporated into simple clinical decision support tools to aid in risk stratification and resource allocation. A key strength of this study is the use of ARM, which produces clear and understandable patterns that clinicians and public health teams can easily interpret. Unlike more advanced approaches such as deep learning or ensemble models—which often act as "black boxes"—ARM generates simple, transparent rules (for example, *"cough and fever → high IgG"*) that make the relationships in the data immediately visible. This level of interpretability is especially important in settings where computational resources are limited, dataset sizes are modest, or rapid, intuitive decision support is needed. ARM also works well with the mix of categorical and clinical variables in our dataset and can capture multi-symptom patterns without complex modeling assumptions or training procedures. Beyond the choice of method, the study benefits from standardized IgG testing, an a priori power calculation, and careful preprocessing, including stability and sensitivity checks, all of which help increase confidence in the patterns we identified. This study has several important limitations. First, the cross-sectional design captures a single time point per patient, so we can report associations but cannot establish temporality or causation. Second, subgroup sizes were unequal and in some cases very small (age > 65: n=10; age < 20: n=1), which reduces statistical power and can make ARM results unstable. In particular, rules based on tiny subgroups may show confidence=1.000 simply because every person in that subgroup shares the pattern; likewise, lift can be inflated when support is low. Such rules should therefore be treated as preliminary, hypothesis-generating signals rather than conclusive findings. Third, p-values reported in this manuscript (for example, the chi-square result linking cough to higher IgG, p=0.048) are unadjusted for multiple comparisons, increasing the risk of false positives; future confirmatory analyses should apply corrections such as Bonferroni or Benjamini–Hochberg. Fourth, although we reported the ELISA kit and cutoff method, assay differences between laboratories may limit direct comparisons with other studies. Finally, our sample was drawn from a single region (western Iran) during a specific pandemic wave, and we did not account for all potential confounders (e.g., comorbidities, medication use). Future work should use larger, multi-site and longitudinal cohorts, perform multivariable adjustment and stratified ARM (e.g., by sex), and apply multiple-comparison corrections to validate these findings. In addition, although ARM offers clear advantages in terms of transparency and ease of interpretation, it also has methodological limitations. ARM is well suited for exploratory work and for generating simple, clinically understandable rules, but it is not designed to optimize predictive accuracy in the same way as more advanced machine-learning methods (such as random forests, gradient boosting, or deep learning). These models can capture complex nonlinear interactions and often perform better in large, diverse datasets; however, they typically require substantially more data, involve iterative training with tuning procedures, and produce outputs that

are less intuitive without additional explainability tools. Given the modest sample size and the exploratory aims of the present study, ARM allowed us to prioritize interpretability and minimize overfitting. Nonetheless, future work—especially with larger or longitudinal datasets—should complement ARM with multivariable modeling and modern machine-learning approaches to evaluate whether more complex models can improve prediction while maintaining clinical interpretability. Future research should translate these ARM findings toward practical applications and stronger evidence. First, prospective, longitudinal cohorts with repeated symptom and antibody measurements would clarify temporal relationships and help identify patterns associated with long COVID. Adding other immune markers (e.g., IgM, neutralizing antibodies, cytokine profiles) and viral load measures would enrich the feature space and likely improve rule stability and predictive value. We also recommend validating the discovered rules in larger, multi-site cohorts that include diverse demographics and viral variants (e.g., Delta, Omicron) to assess generalizability. From a translational perspective, ARM-derived rules could be combined with clinical variables in a simple decision-support prototype (for example, an EHR-integrated checklist or scoring tool) to triage patients or flag those at risk for severe disease or prolonged symptoms — subject, of course, to rigorous external validation and usability testing. Finally, exploring complementary modeling approaches (Bayesian networks, decision trees, or ensemble methods) and making the code and de-identified data available will help refine the models and facilitate reproducibility and clinical adoption.

6. Conclusion

In conclusion, this study successfully applied ARM to reveal significant associations between IgG levels and COVID-19 symptoms. The analysis of data from 237 patients during the third wave in western Iran highlighted distinct symptom profiles corresponding to low, moderate, and high IgG levels. Patients with low IgG levels typically presented milder symptoms, such as the absence of cough combined with loss of smell and taste, suggesting a potentially diminished immune response in this group. Conversely, high IgG levels were associated with more severe symptom clusters, including fever, cough, and loss of smell, indicative of a robust immune response often linked to greater disease severity. Moderate IgG levels were characterized by intermediate symptom patterns, reflecting a transitional state in the immune response. Furthermore, the study identified demographic factors influencing IgG levels and symptom presentation. Younger females (20–40 years) and individuals with blood group A tended to exhibit higher IgG levels, while older patients (>65 years) with blood group O and Rh-negative status showed lower IgG levels. These findings underscore the complex interplay between immune response, symptom manifestation, and patient demographics in COVID-19. The application of ARM proved instrumental in uncovering these nuanced associations, demonstrating its potential for enhancing diagnostic frameworks, guiding personalized treatment strategies, and improving public health preparedness. While acknowledging limitations such as sample size disparities across age groups and the cross-sectional study design, this research contributes valuable insights to the understanding of COVID-19 immune dynamics. Future studies should aim to validate these findings in more diverse populations and explore the longitudinal relationship between IgG levels and long COVID symptoms. Besides, from the computational view, exploring advanced variations, like integrating Apriori with other algorithms or constraint-based mining, can uncover deeper insights. Post-mining analysis, including rule pruning and clustering, simplifies interpretation, and tools for visualization, like graphs or heatmaps, help to present patterns effectively. For scalability, leveraging parallel or distributed systems improves efficiency when dealing with large datasets, ensuring both performance and practical impact.

Acknowledgment
The dataset analyzed in this study was originally collected as part of a prior investigation conducted by our research team. Ethical approval for that project, including data collection and secondary analyses, was granted by the Kurdistan University of Medical Sciences Ethics Committee, Sanandaj, Iran (Approval No. IR.MUK.REC.1399.313). This study adheres to the ethical principles outlined in the Declaration of Helsinki and its subsequent amendments, ensuring the protection of patient privacy and data confidentiality.

Data Availability Statement: The data are not publicly available due to patient privacy and ethical restrictions (the dataset contains anonymized clinical and serological records from human participants collected under IRB approval No. IR.MUK.REC.1399.313) but may be available from the corresponding author upon reasonable request, subject to applicable restrictions.

Conflicts of Interest: The author declare no conflict of interest.

AI Use and Originality Declaration: Authors used AI and generative AI tools only for language assistance. They reviewed and edited all such content and take full responsibility for the accuracy, originality, plagiarism, integrity, and final content of the current work, regardless of whether any issue is detected by the journal.

References

AbdulLateef, O., Sanmargaraja, S., Oni, O., Anavhe, P., & Mewomo, C.M. (2024). An association rule mining model for the application of construction technologies during COVID-19. *International Journal of Construction Management*, 24(4), 443-453. https://doi.org/10.1080/15623599.2023.2212420

Abril, A.G., Alejandre, J., Mariscal, A., Alserawan, L., Rabella, N., Roman, E., Lopez-Contreras, J., Navarro, F., Serrano, E., Nomdedeu, J.F., & Vidal, S. (2024). Titers of IgG and IgA against SARS-CoV-2 proteins and their association with symptoms in mild COVID-19 infection. *Scientific reports*, 14(1), 12725. https://doi.org/10.1038/s41598-024-59634-y

Ahmed, T., Hasan, S.T., Akter, A., Tauheed, I., Akhtar, M., Rahman, S.I.A., Bhuiyan, T.R., Ahmed, T., Qadri, F., & Chowdhury, F. (2023). Determining clinical biomarkers to predict long-term SARS-CoV-2 antibody response among COVID-19 patients in Bangladesh. *Frontiers in Medicine*, 10, 1111037. https://doi.org/10.3389/fmed.2023.1111037

Alathari, M.J.A., Al Mashhadany, Y., Bakar, A.A.A., Mokhtar, M.H.H., Zan, M.S.D.B., & Arsad, N. (2024). COVID-19 IgG antibodies detection based on CNN-bilstm algorithm combined with fiber-optic dataset. *Journal of Virological Methods*, 330, 115011. https://doi.org/10.1016/j.jviromet.2024.115011

Alimohamadi, Y., Sepandi, M., Taghdir, M., & Hosamirudsari, H. (2020). Determine the most common clinical symptoms in COVID-19 patients: A systematic review and meta-analysis. *Journal of Preventive Medicine and Hygiene*, 61(3), E304-E312. https://doi.org/10.15167/2421-4248/jpmh2020.61.3.1530

Altuna, M., Sánchez-Saudinós, M.B., & Lleó, A. (2021). Cognitive symptoms after COVID-19. *Neurology perspectives*, 1, S16-S24. https://doi.org/10.1016/j.neurop.2021.10.005

Amjadi, M.F., O'Connell, S.E., Armbrust, T., Mergaert, A.M., Narpala, S.R., Halfmann, P.J., Bashar, S. J., Glover, C.R., Heffron, A.S., Taylor, A., Flach, B., O'Connor, D.H., Kawaoka, Y., McDermott, A.B., Sethi, A.K., & Shelef, M.A. (2021). Specific COVID-19 symptoms correlate with high antibody levels against SARS-CoV-2. *Immunohorizons*, 5(6), 466-476. https://doi.org/10.4049/immunohorizons.2100022

Bastos, M.L., Tavaziva, G., Abidi, S.K., Campbell, J.R., Haraoui, L.P., Johnston, J.C., Lan, Z., Law, S., MacLean, E., Trajman, A., Menzies, D., Benedetti, A., & Khan, F.A. (2020). Diagnostic accuracy

of serological tests for covid-19: systematic review and meta-analysis. *BMJ*, 370, m2516. https://doi.org/10.1136/bmj.m2516

Binswanger, I.A., Palmer-Toy, D.E., Barrow, J.C., Narwaney, K.J., Bruxvoort, K.J., Kraus, C.R., Lyons, J.A., Lam, J.A., & Glanz, J.M. (2024). Assessing the association between antibody status and symptoms of long COVID: A multisite study. *PloS One*, 19(6), e0304262. https://doi.org/10.1371/journal.pone.0304262

de Oliveira, M.I., Aciole, M.R., Neves, P.A.F., Silva, V.P.O.E., Silva, M.P.O.E., de Lorena, V.M.B., & de Araújo, P.S.R. (2024). A stronger antibody response in increased disease severity of SARS-CoV-2. *BMC Infectious Diseases*, 24, 1-17. https://doi.org/10.1186/s12879-023-08923-4

Dehghani, M., & Yazdanparast, Z. (2023). Discovering the symptom patterns of COVID-19 from recovered and deceased patients using Apriori association rule mining. *Informatics in Medicine Unlocked*, 42, 101351. https://doi.org/10.1016/j.imu.2023.101351

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., & Zupan, B. (2013). Orange: data mining toolbox in Python. *Journal of Machine Learning Research*, 14, 2349-2353.

Elkin, M.E., & Zhu, X. (2022). A machine learning study of COVID-19 serology and molecular tests and predictions. *Smart Health*, 26, 100331. https://doi.org/10.1016/j.smhl.2022.100331

Fajgenbaum, D.C., & June, C.H. (2020). Cytokine storm. *New England Journal of Medicine*, 383(23), 2255-2273. https://doi.org/10.1056/NEJMra2026131

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. https://doi.org/10.3758/BF03193146

Franchini, M., Maggi, F., & Focosi, D. (2024). ABO blood group-related mechanism of infection of SARS-CoV-2: an overview of systematic reviews. *Clinical Chemistry and Laboratory Medicine*, 62(3), 396-401. https://doi.org/10.1515/cclm-2023-0825

Ghaderi, H., Elahimanesh, F., Roshani, D., Hosseini Nasab, A., Nikkhoo, B., Mohammadi, A., Hessami, S., & Ghaznavi, H. (2024). Seroprevalence of anti-SARS-CoV-2 IgG antibodies across ABO blood groups and Rh Factor. *Microbes and Infectious Diseases*, 14, 1-10. https://doi.org/10.21608/mid.2024.303954.2075

Hormaty, S., Seiwan, A.N., Rasheed, B.H., Parvaz, H., Gharahzadeh, A., & Ghaznavi, H. (2025). A review on biomarker-enhanced machine learning for early diagnosis and outcome prediction in ovarian cancer management. *Cancer Medicine*, 14(17), e71224. https://doi.org/10.1002/cam4.71224

Hudu, S.A., Alshrari, A.S., Abu-Shoura, E.A.J.I., Osman, A., & Jimoh, A.O. (2025). A critical review of the prospect of integrating artificial intelligence in infectious disease diagnosis and prognosis. *Interdisciplinary Perspectives on Infectious Diseases*, 2025, 6816002. https://doi.org/10.1155/ipid/6816002

Jia, X., Zhang, P., Tian, Y., Wang, J., Zeng, H., Wang, J., Liu, J., Chen, Z., Zhang, L., He, H., He, K., & Liu, Y. (2021). Clinical significance of an IgM and IgG test for diagnosis of highly suspected COVID-19. *Frontiers in Medicine*, 8, 569266. https://doi.org/10.3389/fmed.2021.569266

Khudhur, S.D., & Khudhur, D.D. (2022). IgG-IgM antibodies-based infection time detection of COVID-19 using machine learning models. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 20(2), 340-347. https://doi.org/10.12928/TELKOMNIKA.v20i2.21649

Klein, S.L., & Flanagan, K.L. (2016). Sex differences in immune responses. *Nature Reviews Immunology*, 16(10), 626-638. https://doi.org/10.1038/nri.2016.90

Kurano, M., Ohmiya, H., Kishi, Y., Okada, J., Nakano, Y., Yokoyama, R., Qian, C., Xia, F., He, F., Zheng, L., Yu, Y., Jubishi, D., Okamoto, K., Moriya, K., Kodama, T., & Yatomi, Y. (2022). Measurement of SARS-CoV-2 antibody titers improves the prediction accuracy of COVID-19 maximum severity by machine learning in non-vaccinated patients. *Frontiers in Immunology*, 13, 811952. https://doi.org/10.3389/fimmu.2022.811952

Lakshmi, K.S., & Vadivu, G. (2017). Extracting association rules from medical health records using multi-criteria decision analysis. *Procedia Computer Science*, 115, 290-295. https://doi.org/10.1016/j.procs.2017.09.137

Lucas, C., Wong, P., Klein, J., Castro, T.B.R., Silva, J., Sundaram, M., Ellingson, M.K., Mao, T., Oh, J.E., Israelow, B., Takahashi, T., Tokuyama, M., Lu, P., Venkataraman, A., Park, A., Mohanty, S., Wang, H., Wyllie, A.L., Vogels, C.B.F., . . . Iwasaki, A. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*, 584(7821), 463-469. https://doi.org/10.1038/s41586-020-2588-y

Marklund, E., Leach, S., Axelsson, H., Nyström, K., Norder, H., Bemark, M., Angeletti, D., Lundgren, A., Nilsson, S., Andersson, L.M., Yilmaz, A., Lindh, M., Liljeqvist, J.Å., & Gisslén, M. (2020). Serum-IgG responses to SARS-CoV-2 after mild and severe COVID-19 infection and analysis of IgG non-responders. *PloS One*, 15(10), e0241104. https://doi.org/10.1371/journal.pone.0241104

Matharaarachchi, S., Domaratzki, M., Katz, A., & Muthukumarana, S. (2022). Discovering long COVID symptom patterns: association rule mining and sentiment analysis in social media tweets. *JMIR Formative Research*, 6(9), e37984. https://doi.org/10.2196/37984

Mutter, S., Hall, M., & Frank, E. (2004). Using classification to evaluate the output of confidence-based association rule mining. In G.I. Webb & X. Yu (Eds.), *Proceedings of the 17th Australian Joint Conference on Artificial Intelligence,* 538-549. Springer. https://doi.org/10.1007/978-3-540-30549-1-47

Nikolich-Žugich, J. (2018). The twilight of immunity: emerging concepts in aging of the immune system. *Nature Immunology*, 19(1), 10-19. https://doi.org/10.1038/s41590-017-0006-x

Park, S.E. (2020). Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clinical and Experimental Pediatrics*, 63(4), 119-124. https://doi.org/10.3345/cep.2020.00493

Riazi, K., Ly, M., Barty, R., Callum, J., Arnold, D.M., Heddle, N.M., Down, D.G., Sidhu, D., & Li, N. (2023). An unsupervised learning approach to identify immunoglobulin utilization patterns using electronic health records. *Transfusion*, 63(12), 2234-2247. https://doi.org/10.1111/trf.17585

Silva, F.A.F.D., Brito, B.B.D., Santos, M.L.C., Marques, H.S., Silva Júnior, R.T.D., Carvalho, L.S.D., Vieira, E.S., Oliveira, M.V., & Melo, F.F.D. (2020). COVID-19 gastrointestinal manifestations: A systematic review. *Revista da Sociedade Brasileira de Medicina Tropical*, 53, e20200714. https://doi.org/10.1590/0037-8682-0714-2020

Singh, A., Singh, D., Upreti, K., Sharma, V., Rathore, B.S., & Raikwal, J. (2023). Investigating New Patterns in Symptoms of COVID-19 Patients by Association Rule Mining (ARM). *Journal of Mobile Multimedia*, 19(1), 1-28. https://doi.org/10.13052/jmm1550-4646.1911

Sobhani, S., Aryan, R., Kalantari, E., Soltani, S., Malek, N., Pirzadeh, P., Yarahmadi, A., & Aghaee, A. (2021). Association between clinical characteristics and laboratory findings with outcome of hospitalized COVID-19 patients: A report from northeast Iran. *Interdisciplinary Perspectives on Infectious Diseases*, 2021, 5552138. https://doi.org/10.1155/2021/5552138

Sogandi, F. (2024). Identifying diseases symptoms and general rules using supervised and unsupervised machine learning. *Scientific Reports*, 14, 17956. https://doi.org/10.1038/s41598-024-69029-8

Sołkiewicz, K., Kokot, I., Dymicka-Piekarska, V., Dorf, J., & Kratz, E.M. (2024). Are changes in serum IgG glycosylation related to the severe course of SARS-CoV-2 infection and recovery process? In search of new diagnostic and prognostic biomarkers. *Journal of Inflammation Research*, 17, 1413-1427. https://doi.org/10.2147/JIR.S439005

Sun, C., Xue, M., Yang, M., Zhu, L., Zhao, Y., Lv, X., Lin, Y., Ma, D., Shen, X., Cheng, Y., Xuan, H., Jia, X., Li, T., & Han, L. (2021). Early prediction of severe COVID-19 in patients by a novel immune-related predictive model. *Msphere*, 6(5), e00752-21. https://doi.org/10.1128/mSphere.00752-21

Tandan, M., Acharya, Y., Pokharel, S., & Timilsina, M. (2021). Discovering symptom patterns of COVID-19 patients using association rule mining. *Computers in Biology and Medicine*, 131, 104249. https://doi.org/10.1016/j.compbiomed.2021.104249

Vojdani, A., Almulla, A.F., Zhou, B., Al-Hakeim, H.K., & Maes, M. (2024). Reactivation of herpesvirus type 6 and IgA/IgM-mediated responses to activin-A underpin long COVID, including affective symptoms and chronic fatigue syndrome. *Acta Neuropsychiatrica*, 36(3), 172-184. https://doi.org/10.1017/neu.2024.10

Xu, J., Chen, J., Wen, F., Liu, K., & Chen, Y. (2024). Detection methods and dynamic characteristics of specific antibodies in patients with COVID-19: A review of the early literature. *Heliyon*, 10(3), e24580. https://doi.org/10.1016/j.heliyon.2024.e24580

Ye, H., Gan, X., Zhou, W., Gao, Y., Mei, Z., Zheng, Q., Luo, X., Yuan, C., & Wu, Y. (2024). Symptom clusters and symptom networks of symptom experiences in patients with SARS-CoV-2 infection. *Heliyon*, 10(22), e40497. https://doi.org/10.1016/j.heliyon.2024.e40497

Zakur, Y., & Flaih, L. (2023). Apriori algorithm and hybrid Apriori algorithm in the data mining: A comprehensive review. *E3S Web of Conferences*, 448, 02021. https://doi.org/10.1051/e3sconf/202344802021

Zhang, C., & Zhang, S. (2002). *Association rule mining: models and algorithms*. *Springer*. https://doi.org/10.1007/3-540-46027-6

Zhuang, Z., Qi, Y., Yao, Y., & Yu, Y. (2023). A predictive model for disease severity among COVID-19 elderly patients based on IgG subtypes and machine learning. *Frontiers in Immunology*, 14, 1286380. https://doi.org/10.3389/fimmu.2023.1286380