Article

Rapid Triage of Children with Suspected COVID-19 Using Laboratory-Based Machine-Learning Algorithms

Dejan Dobrijević [1,2,*], Gordana Vilotijević-Dautović [1,2], Jasmina Katanić [1,2], Mirjana Horvat [3], Zoltan Horvat [3] and Kristian Pastor [4]

1 Faculty of Medicine, University of Novi Sad, 21000 Novi Sad, Serbia; gordana.vilotijevic-dautovic@mf.uns.ac.rs (G.V.-D.); jasmina.katanic@mf.uns.ac.rs (J.K.)
2 Institute for Child and Youth Health Care of Vojvodina, 21000 Novi Sad, Serbia
3 Faculty of Civil Engineering Subotica, University of Novi Sad, 24000 Subotica, Serbia; isicm@gf.uns.ac.rs (M.H.); horvatz@gf.uns.ac.rs (Z.H.)
4 Faculty of Technology, University of Novi Sad, 21000 Novi Sad, Serbia; kristian.pastor@uns.ac.rs
* Correspondence: dejan.dobrijevic@mf.uns.ac.rs; Tel.: +381-21-48-80-444

Abstract: In order to limit the spread of the novel betacoronavirus (SARS-CoV-2), it is necessary to detect positive cases as soon as possible and isolate them. For this purpose, machine-learning algorithms, as a field of artificial intelligence, have been recognized as a promising tool. The aim of this study was to assess the utility of the most common machine-learning algorithms in the rapid triage of children with suspected COVID-19 using easily accessible and inexpensive laboratory parameters. A cross-sectional study was conducted on 566 children treated for respiratory diseases: 280 children with PCR-confirmed SARS-CoV-2 infection and 286 children with respiratory symptoms who were SARS-CoV-2 PCR-negative (control group). Six machine-learning algorithms, based on the blood laboratory data, were tested: random forest, support vector machine, linear discriminant analysis, artificial neural network, k-nearest neighbors, and decision tree. The training set was validated through stratified cross-validation, while the performance of each algorithm was confirmed by an independent test set. Random forest and support vector machine models demonstrated the highest accuracy of 85% and 82.1%, respectively. The models demonstrated better sensitivity than specificity and better negative predictive value than positive predictive value. The F1 score was higher for the random forest than for the support vector machine model, 85.2% and 82.3%, respectively. This study might have significant clinical applications, helping healthcare providers identify children with COVID-19 in the early stage, prior to PCR and/or antigen testing. Additionally, machine-learning algorithms could improve overall testing efficiency with no extra costs for the healthcare facility.

Keywords: children; infection; COVID-19; machine learning; laboratory

1. Introduction

The coronavirus disease (COVID-19), caused by a novel strain of betacoronavirus (SARS-CoV-2), has marked the past two years with over 6.7 million deaths worldwide. New cases are still being diagnosed but with predominantly mild clinical manifestations. COVID-19 primarily occurs in the adult population, but children play a significant role in the spread of the disease. Children with COVID-19 usually only have a few mild symptoms or no symptoms at all, which is why they remain unrecognized. Additionally, newborns, infants, and toddlers cannot wear protective masks in an appropriate way, and moreover, they cannot clearly describe their health condition. Regardless of the disease's severity and the amount of viral load, it is important to consider that pediatric patients may contribute to the transmission chain. For all of the above reasons, the pediatric population should receive special attention during the current pandemic [1–3].

Even though viral pneumonia has been recognized as the main clinical presentation of this disease, representing the main cause of its severity and mortality, SARS-CoV-2 infection

may cause several complications in other organs, such as coagulation disorders (pulmonary embolism, venous thromboembolism, hemorrhages, and acute ischemic stroke) with abdominal involvement (acute mesenteric ischemia, pancreatitis, and acute kidney injury), especially in severely ill patients and those admitted to the ICU, even in children [4,5]. According to the available data, fever occurs in almost 90% of patients and weakness in 70% of patients. A dry cough is present in more than 60% of patients. Nausea and vomiting are pronounced in 5% of patients, and diarrhea occurs in almost 4% of patients [6]. Guan et al. [7] showed that 15.74% of the patients had a severe clinical form of the disease. During hospital treatment, more than 90% of patients were diagnosed with pneumonia. Acute respiratory distress syndrome (ARDS) was confirmed in 3.4% of patients and septic shock in slightly more than 1% of patients. Comorbid diseases, such as hypertension, cardiomyopathy, coronary artery disease, chronic kidney diseases, chronic lung disease, etc., are significantly more common in patients with severe symptoms of the disease (38%) compared to those with a milder form of the disease (21%). Based on the laboratory findings, the occurrence of lymphopenia, which is commonly observed in adults with COVID-19, was found in laboratory tests in only 5.5% of children diagnosed with the disease. The estimated prevalence of leukopenia in pediatric COVID-19 patients was found to be 7.3%. The prevalence rates for high C-reactive protein (CRP) levels, high LDH levels, high creatine kinase MB (CK-MB) levels, high AST levels, and high erythrocyte sedimentation rate (ESR) were estimated to be 14.0%, 17.4%, 43%, 12.3%, and 29.7%, respectively [8]. SARS-CoV-2 infection was commonly followed by hyperinflammation due to the excessive production of proinflammatory cytokines, such as IL-1, IL-2, IL-6, IL-15, IL-18, TNF-α, IFN-γ, etc. Numerous cytokines have been tested in order to reduce mortality, especially in critically ill patients [9].

The gold standard for confirming the presence of the viral genome in a biological sample is quantitative polymerase chain reaction (qPCR). Chest X-rays and computed tomography (CT) were regarded as the main diagnostic tools for the diagnosis of COVID-19. However, given the progressively increased availability of RT-PCR, CT changed from being primarily a diagnostic tool to playing a prognostic role. In fact, evaluation of the CT score became essential for proper patient management, proving to be essential for deciding whether to hospitalize the patient in healthcare settings with limited resources and a shortage of intensive care beds [10]. Although qPCR is an irreplaceable diagnostic tool in the current pandemic, a prolonged turnaround time is often a significant issue. Additionally, molecular diagnostics is relatively expensive, especially for developing countries, and represents a significant burden for laboratory staff. Due to various factors, such as disease prevalence, an increasing population, a rise in the usage of healthcare services, etc., the diagnostic costs continue to rise [11]. Therefore, the question arises as to how to perform a rapid triage of children with suspected COVID-19 prior to qPCR. Artificial intelligence (AI) offers novel and more economical solutions. According to extensive cost–benefit research by Khanna et al. [12], AI has been recognized as a promising and tremendously cost-saving diagnostic tool.

The new technologies of Industry 4.0 have significantly influenced medicine in terms of differential diagnosis, prognosis, and treatment of diseases. Hence, this digitalization and transformation of medicine is also labeled Medicine 4.0. In this era of digitalization in medicine, artificial intelligence has been recognized and used as a promising diagnostic tool and support in the fight against COVID-19 [13,14]. In the largest number of papers published so far, machine-learning algorithms were used to recognize X-ray and/or CT abnormalities in SARS-CoV-2-positive patients [15], while fewer studies focused on clinical laboratory parameters [16]. Most studies included an adult population, while the data for the pediatric population are still insufficient. Additionally, laboratory markers have been mostly used for prognosis rather than preliminary diagnosis and triage of COVID-19 patients [17].

Therefore, the aim of this study was to assess the utility of the most common machine-learning algorithms in the rapid triage of children with suspected COVID-19 using easily accessible and less expensive laboratory parameters.

2. Materials and Methods

This cross-sectional study included 280 children with PCR-confirmed SARS-CoV-2 infection (COVID group), treated at the Institute for Children and Youth Health Care of Vojvodina, Novi Sad, Serbia, in the period from March 2020 to December 2022. The control group (non-COVID group) consisted of 286 children with respiratory symptoms who were SARS-CoV-2 PCR-negative. The exclusion criteria were chronic diseases, malignancies, hematological diseases, and missing data (Figure 1). The detection of SARS-CoV-2 antigen in nasopharyngeal swab samples of children was performed by the qPCR technique at the Institute of Public Health of Vojvodina, Novi Sad, Serbia. Children were divided into six age groups according to chronological age: newborn (0–28 days), infant (1–12 months), toddler (1–3 years), preschool (4–6 years), school (7–14 years), and adolescent (15–18 years).
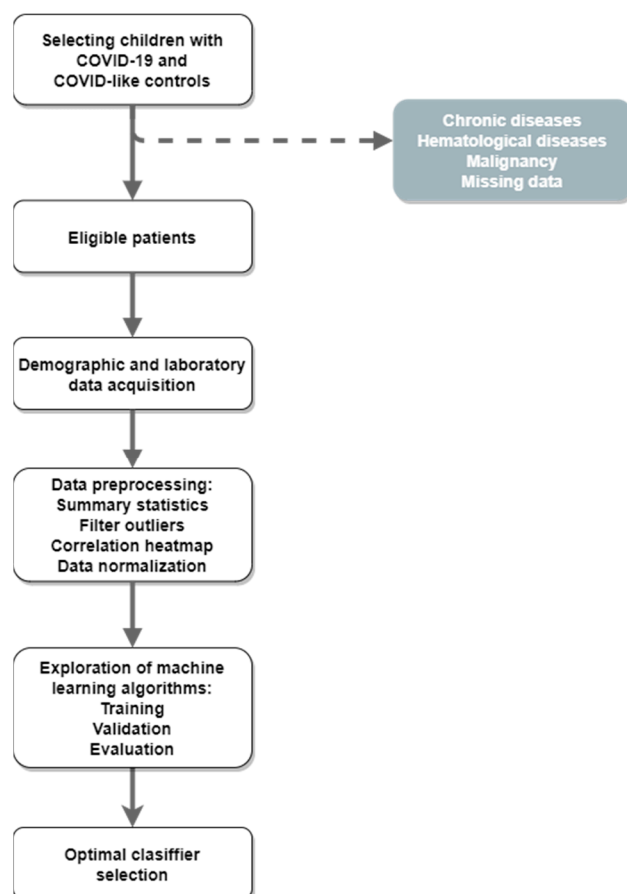

Figure 1. Workflow of predictive modeling.

2.1. Data Acquisition

Data on parameters of the complete blood count and baseline biochemical parameters, including aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl transferase (GGT), lactate dehydrogenase (LDH), and C-reactive protein (CRP), were collected on the day of admission. All data were obtained through the institute's laboratory information system using the structured query language (SQL) code as a searching tool. The blood samples were collected using 0.5 mL violet-topped microtubes with ethylenediaminetetraacetic acid dipotassium salt dehydrate (K2EDTA) as a blood clotting inhibitor (Becton Dickinson, Franklin Lakes, NJ, USA). The values were determined

from the hematology analyzer Advia 2120 (Siemens Healthcare, Erlangen, Germany) and chemistry analyzer DxC 700 AU (Beckman Coulter, Brea, CA, USA).

2.2. Data Preprocessing

First, all patients with missing data were excluded from the study, following one of the above-mentioned eligibility criteria. Second, the outliers were identified as data points located outside the whiskers of the box plot and excluded from further analysis. Third, Spearman's correlation was used to screen out highly correlated laboratory parameters in order to minimize the number of input parameters (the threshold value was set to 0.4). A correlation heatmap was used to visualize the strength of relationships between the parameters (Figure S1). Fourth, min–max normalization was applied to transform each parameter into the range [0, 1] in order to treat them with equal weight without distorting the general distribution in the source data.

2.3. Baseline Statistical Analyses

Statistical analyses (descriptive and inferential) were performed using open-source software, JASP version 0.16.4.0 (Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands). The significance level for the calculated differences was set at 0.05. For continuous random variables, the normality of distribution was estimated using the Shapiro–Wilk test. Between-group differences were analyzed using the Mann–Whitney U-test. Univariate logistic regression analysis was performed to determine the parameters, which could predict COVID-19 occurrence.

2.4. Machine-Learning Algorithms

The following machine-learning algorithms were tested in this study: random forest (RF), support vector machine (SVM), linear discriminant analysis (LDA), artificial neural network (ANN), k-nearest neighbors (KNN), and decision tree (DT). All these algorithms belong to supervised learning, and their goal is classification. The data set was divided into two subsets: the training set and the test set, with an 80:20 split. The training set was validated through stratified cross-validation, where each tuning cycle involved a different non-overlapping holdout data set. The performance of each algorithm was confirmed by additional, independent data set—the test set. The final evaluation of the model included the calculation of accuracy, sensitivity, specificity, positive predictive value, and negative predictive value from the confusion matrix, i.e., the table of predicted and actual values of a classifier. The values were expressed as percentages. Discrimination between groups (COVID and non-COVID) by machine-learning algorithms was presented using receiver operating characteristics (ROC) curves.

2.5. Ethical Approval

The study was approved by the Ethics Committee of the Institute for Children and Youth Healthcare of Vojvodina (22 July 2022; No. 3280–2).

3. Results

Following the eligibility criteria, in the period from March 2020 to December 2022, a total of 566 children treated at the Institute for Children and Youth Health Care of Vojvodina, Novi Sad, Serbia, were included in the study. The median age of the COVID-19 group was 4.2 years with a female share of 46.4%, while the median age of the non-COVID group was 3.8 years with a female share of 47.9%. Children were divided into six age groups according to chronological age: newborn (7.6%), infant (15.9%), toddler (24.9%), preschool (18.6%), school (17.8%), and adolescent (15.2%).

3.1. Clinical Laboratory Features

The initial data included 22 clinical laboratory parameters, a complete blood count, and baseline biochemical parameters. After screening out highly correlated instances

(Spearman's rank correlation coefficient over 0.4), a total of 14 parameters were included in further analysis: white blood cells (WBC), red blood cells (RBC), mean corpuscular volume (MCV), mean corpuscular hemoglobin concentration (MCHC), platelets (PLT), mean platelet volume (MPV), plateletcrit (PCT), platelet distribution width (PDW), absolute lymphocyte count (LYM#), absolute eosinophil count (EOS#), AST, GGT, LDH, and CRP (Table 1). Univariate logistic regression analysis was employed to examine the association of individual laboratory parameters with the presence of SARS-CoV-2 infection in children. The following parameters demonstrated significant diagnostic properties as independent predictors: WBC, MCHC, MPV, and PDW (Table 1). PCR-SARS-CoV-2-negative children had higher values of WBC and MCHC, while children with COVID-19 had higher values of MPV and PDW.

Table 1. Laboratory findings as diagnostic markers for children with suspected COVID-19.

Laboratory Parameter [a]	COVID Group (n = 280)	Non-COVID Group (n = 286)	Overall (n = 566)	p-Value	Univariate Analysis		Multivariate Analysis	
					OR (95% CI)	p-Value	OR (95% CI)	p-Value
WBC (10^9)	7.9 (5.6–11.8)	10.9 (7.9–15.2)	9.4 (6.6–13.9)	**<0.001**	**1.088 (1.055–1.123)**	**<0.001**	**1.052 (1.016–1.089)**	**0.004**
RBC (10^{12})	4.5 (4.1–4.9)	4.4 (4.1–4.8)	4.5 (4.1–4.8)	0.359	NA	NA	NA	NA
MCV (fL)	80.7 (77.2–85.1)	79.9 (76.1–84.9)	80.1 (76.4–85)	0.130	NA	NA	NA	NA
MCHC (g/L)	340 (331–347)	344 (332–352)	342 (331.2–350)	**0.003**	**1.039 (1.024–1.056)**	**<0.001**	**1.029 (1.014–1.044)**	**<0.001**
PLT (10^9)	300 (217–386.2)	342 (239.5–403.5)	315 (230–392.8)	**0.028**	1.001 (0.999–1.003)	0.189	NA	NA
MPV (fL)	7.8 (7.1–8.5)	7.4 (6.8–8)	7.5 (7–8.3)	**<0.001**	**1.028 (1.002–1.054)**	**0.031**	1.001 (0.996–1.007)	0.387
PCT (%)	0.24 (0.18–0.3)	0.23 (0.18–0.3)	0.23 (0.18–0.3)	0.943	NA	NA	NA	NA
PDW (%)	13.8 (11.7–16.3)	12.9 (11.7–14.6)	13.4 (11.7–15.5)	**0.010**	**1.427 (1.283–1.587)**	**<0.001**	**1.183 (1.090–1.283)**	**<0.001**
LYM# (10^9)	2.3 (1.4–3.9)	2.8 (1.9–4.9)	2.6 (1.6–4.4)	**<0.001**	1.054 (0.956–1.161)	0.291	NA	NA
EOS# (10^9)	0.07 (0.03–0.13)	0.1 (0.05–0.15)	0.09 (0.04–0.15)	0.889	NA	NA	NA	NA
AST (μkat/L)	0.62 (0.46–0.82)	0.57 (0.44–0.45)	0.58 (0.45–0.79)	**0.048**	0.821 (0.621–1.085)	0.165	NA	NA
GGT (μkat/L)	0.24 (0.18–0.47)	0.26 (0.19–0.56)	0.25 (0.18–0.52)	0.124	NA	NA	NA	NA
LDH (μkat/L)	4.39 (3.55–5.1)	4.68 (3.73–5.44)	4.4 (3.7–5.3)	**0.017**	1.107 (0.992–1.234)	0.068	NA	NA
CRP (mg/L)	5.5 (1.2–30)	13.2 (2.7–71.6)	9.7 (1.6–53.7)	**<0.001**	1.002 (0.999–1.005)	0.171	NA	NA

[a] Values are median (interquartile range: Q1–Q3); Mann–Whitney U-test. WBC—White blood cells. RBC—Red blood cells. MCV—Mean corpuscular volume. MCHC—Mean corpuscular hemoglobin concentration. PLT—Platelet. MPV—Mean platelet volume. PCT—Plateletcrit. PDW—Platelet distribution width. LYM#—Absolute lymphocyte count. EOS#—Absolute eosinophil count. AST—Aspartate aminotransferase. GGT—Gamma-glutamyl transferase. LDH—Lactate dehydrogenase. CRP—C-reactive protein. Values in bold are statistically significant.

3.2. Machine-Learning Algorithm Performances

A comparison of six investigated machine-learning algorithms, based on the standard evaluation metrics, was carried out with a reduced number of instances (Table 2).

Table 2. Machine-learning classifiers for the rapid triage of children with suspected COVID-19.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	F1 Score (%)
Random forest	85.0	86.0	83.9	84.5	85.5	85.2
Support vector machine	82.1	84.8	79.4	80.0	84.4	82.3
Linear discriminant analysis	78.8	81.1	76.7	75.4	82.1	78.1
Neural network	76.1	72.6	80.4	81.8	70.7	76.9
k-nearest neighbors	73.5	71.0	76.5	78.6	68.4	74.6
Decision tree	68.1	65.5	70.7	67.9	68.3	66.7

The RF and SVM models demonstrated the highest accuracy of 85% and 82.1%, respectively, while all the other algorithms classified instances with an accuracy lower than 80%. The RF and SVM models demonstrated better sensitivity than specificity and better negative predictive value than positive predictive value. The F1 score, which combines positive predictive value (precision) and sensitivity (recall) using their harmonic means, was higher for the RF than for the SVM model, 85.2% and 82.3%, respectively. After evaluating the performance of the best model (in our study, this was the RF model), the feature importance was compared based on its increase in node purity, i.e., its mean decrease in accuracy. The most prominent instances (node purity over 0.01) were shown to be MPV, WBC, MCHC, PDW, and LYM#. Discrimination between groups (COVID and non-COVID) by machine-learning algorithms was presented using receiver operating characteristics (ROC) curves (Figure 2).
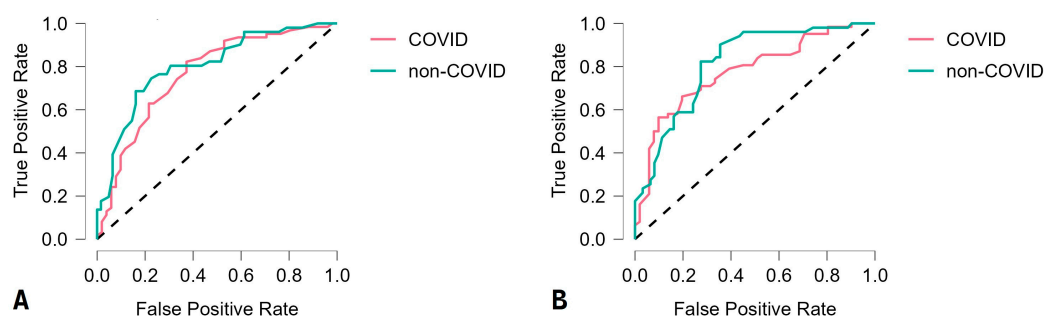
Figure 2. Random forest (**A**) and support vector machine (**B**) receiver operating characteristic curve for the rapid triage of children with suspected COVID-19.

4. Discussion

In order to limit the spread of the SARS-CoV-2 virus, it is necessary to detect positive cases as soon as possible and isolate them. However, the small number of available qPCR tests, their high price, and the relatively high percentage of false negative results of these tests brought about the need for additional diagnostic tools [18]. Machine-learning algorithms for automatic disease detection have been increasingly applied in different areas of medicine. Machine learning is a field of artificial intelligence, which provides systems with the ability to automatically learn from experience. The main purpose of these models is to find appropriate patterns in the data, i.e., to produce statistically reliable and reproducible results [19,20].

In this study, the authors evaluated the performance of the six most common machine-learning algorithms: RF, SVM, LDA, ANN, KNN, and DT. The RF and SVM models outperformed the others with an accuracy of 85% and 82.1%, respectively. Both algorithms fall under supervised machine learning. The task of supervised machine-learning algorithms is to "learn" the prediction function h(x) based on a given training data set,

so that h(x) is an optimal approximation of the target classes, in this case COVID and non-COVID [20,21].

The SVM machine-learning technique solves the problems of non-linear classification and regression using convex quadratic programming methods. This model only uses instances from the training set that contribute most to the optimal solution of the quadratic programming problem, forming the so-called support vectors. SVM is a very popular and reliable prediction method. During the COVID-19 pandemic, its application has been confirmed for diagnostic purposes [22–24], mortality risk assessment [25,26], detecting undertriage in telephone triage [27], etc.

Unlike SVM, which belongs to the category of "individual" algorithms in supervised machine learning, RF belongs to the class of ensemble methods, which combine the results of several individual methods in a certain way. This approach aims to obtain better prediction results than any of the individual methods. For RF construction, an ensemble is formed consisting of several hundred to several thousand DTs. The advantages of DTs, compared to other machine-learning methods, are their simplicity of implementation and the comprehensibility of the procedure. There are rules by which trees are quickly formed, and the output can be easily interpreted. In addition, DTs allow attributes to have missing values, which is not the case with SVM. However, one of the disadvantages of the DT method is its instability. A small change in the input training data can lead to a significant change in the topology of the tree. Instability occurs due to many possible splits, which often have approximately the same importance (competitor splits). Therefore, a small change in the data can lead to a completely different partition, which further introduces changes to all the branches of the tree below it. RF overcomes these limitations by aggregating the prediction results of hundreds of individual trees [21]. Therefore, the RF model has been widely used during the COVID-19 pandemic for disease diagnosis [28,29], predicting patient outcomes [30–32], recommending hospitalization [33], processing of healthcare and travel data to identify COVID-infected people [34], etc.

The evaluation metrics for the RF and SVM models based on clinical laboratory data reported in other studies were similar to this study. Our RF model demonstrated an accuracy of 85%, while Çubukçu et al. [35] reported an accuracy of 85.2% in their RF model using complete blood count parameters and clinical chemistry parameters as input variables, but in an adult population. Our SVM model demonstrated an accuracy of 82.1%, thus outperforming a model proposed by Thimoteo et al. [36], who reported an accuracy of 73.7% in their SVM model, including complete blood count parameters only, as well as an adult population.

The model presented in this study cannot outperform the PCR method, which is considered the diagnostic gold standard, with an average efficiency of over 96% [37]. Conversely, the suggested model outperformed the immunochromatography method used in the rapid SARS-CoV-2 antigen tests. In the beginning of the pandemic, only a few antigen tests received emergency use authorization (EUA) from regulatory authorities, indicating their acceptable performance [38]. Subsequently, some of them were reported to have a sensitivity of no more than 30% [39]. Since these tests were widely used for diagnosing active infections, their performance has significantly improved over time. One of the most commonly used rapid antigen tests at our institute has an overall sensitivity of 79.6% [40]. Our RF and SVM models demonstrated a sensitivity of 86% and 84.8%, respectively.

Including additional clinical biochemical analyses in a machine-learning algorithm, such as ferritin [41], fibrinogen, D-dimer [42], procalcitonin, interleukin-6 [43], etc., may increase the accuracy, sensitivity, and specificity of the model. However, minimal blood sample volume is imperative in pediatric health research. For example, a complete blood count analysis can be performed using a total blood volume of 25 μL. Blood volume overdraws in pediatric laboratory medicine should always be taken into consideration from a legal and ethical perspective [44].

Children experience milder symptoms during the course of the SARS-CoV-2 infection in comparison to adult individuals. Age-related differences may reflect disease severity.

These age-related differences include differences in immunity, differences in binding affinity of the SARS-CoV-2 target receptors, etc. [45,46]. In our study, children were divided into six age groups according to chronological age: newborn (0–28 days), infant (1–12 months), toddler (1–3 years), preschool (4–6 years), school (7–14 years), and adolescent (15–18 years). According to the systematic review conducted by Carobene et al. [47], only half of the PubMed and Scopus publications on the application of artificial intelligence in COVID-19 diagnostics take demographic data, such as gender and age, into consideration. The reference ranges in pediatric laboratory medicine are strictly defined and age-dependent. Therefore, it is mandatory to compare individuals within the same age group. This is often overlooked by many researchers, which produces misleading conclusions. Another important consideration and strength of the study is the period in which the children were included, i.e., March 2020–December 2022. This period covers several waves of the pandemic, both before and after the vaccination program had started.

Data on laboratory-based machine-learning approaches for the detection and triage of children with COVID-19 are scarce. Previous studies were mainly focused either on the adult population or, in the case of pediatric population, on radiological rather than laboratory findings, and on outcome rather than detection. With that being considered, this study is a unique contribution to pediatric laboratory medicine.

Multiplex PCR and rapid panel antigen tests (lateral immunochromatography) are used in the diagnosis of viral infections in pediatrics. However, the healthcare system faces a constant challenge in finding an additional diagnostic modality, which is fast, reliable, and cheap, such as AI algorithms. Nevertheless, the potential drawbacks of machine learning in personalized laboratory medicine should be taken into consideration. Diagnostics should always strongly rely on human skills, including physical examination, critical perception of medical history, etc. Machine-learning algorithms provide multi-dimensional biomedical data, which should not be overrated and should only be observed as supporting information for making a final diagnosis [48].

Analyzing the cost–benefit properties of the algorithms proposed in this study, it can be concluded that the healthcare system can benefit from these algorithms in terms of both time and money. The turnaround time for SARS-CoV-2 qPCR tests can vary depending on several factors, including testing capacity and demand, laboratory workload and staffing, supply chain issues, transportation, logistics, etc. The average turnaround time for SARS-CoV-2 qPCR tests ranges from a few hours to a couple of days [49]. On the contrary, laboratory tests used as input data in the proposed algorithm can be performed within a few minutes, and the algorithm can deliver results within a few seconds. Taking finances into consideration, the proposed algorithms have an advantage over the SARS-CoV-2 qPCR test. The price for a single SARS-CoV-2 qPCR test may vary depending on several factors, such as the country and healthcare system. However, the average price ranges from around USD 50 to USD 200 per test. On the other hand, the baseline laboratory parameters used in this study can be obtained from automated clinical chemistry analyzers for no more than USD 50 [50,51]. Moreover, all the parameters used in this study are part of routine testing upon admission to our institute. Therefore, the proposed algorithm entails no additional laboratory-related cost. Furthermore, the statistical package used in this study is an open-source program without the need for license-related costs. Based on the above-mentioned claims, it can be inferred that the implementation of the proposed machine-learning algorithms could improve overall testing efficiency with no extra costs for the healthcare facility.

The use of machine learning in COVID diagnostics has several clinical implications, which can greatly impact the detection and management of the disease, such as early detection and diagnosis, improved accuracy and efficiency, risk stratification and prognosis, personalized treatment plans, monitoring disease trends and outbreaks, etc. It is important to note that while machine learning holds great promise, it should be integrated into clinical practice with caution. Rigorous validation, ethical considerations, and ongoing monitoring

are necessary to ensure the reliability, safety, and ethical use of machine-learning models in COVID diagnostics [52,53].

There are certain limitations to the approach proposed in this study and some practical considerations for future research to be considered. First, this is a single-center study, which only includes a limited number of children. Second, the pediatric patients in this study are all European. Performing multi-institutional and multi-national studies could evaluate whether the proposed models could perform well in other human races. Third, all children with underlying conditions were excluded from the proposed study, making its clinical applicability for children with co-infections limited. Fourth, only blood counts and baseline biochemical parameters were included in the study, as they are easily and immediately obtained at admission. However, other non-laboratory parameters could also be included, resulting in a potentially better diagnostic model. Fifth, the control group was heterogeneous, without a specific pathogen classification. They were all labeled only as non-COVID, i.e., SARS-CoV-2-negative.

5. Conclusions

Herein, the six most common machine-learning algorithms for the rapid triage of children with suspected COVID-19 were presented and validated. The RF and SVM models outperformed the others with fairly high accuracy. Thereby, a set of clinical laboratory features with markedly high prediction potential was identified, including MPV, WBC, MCHC, PDW, and LYM#. The results of this study might have significant clinical applications, helping healthcare providers identify children with COVID-19 in the early stages, prior to PCR and/or antigen testing. Additionally, machine-learning algorithms could improve overall testing efficiency with no extra costs for the healthcare facility. However, the potential drawbacks of machine learning in personalized laboratory medicine should be taken into consideration.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v15071522/s1, Figure S1: Correlation heatmap.

References

1. Nunziata, F.; Salomone, S.; Catzola, A.; Poeta, M.; Pagano, F.; Punzi, L.; Lo Vecchio, A.; Guarino, A.; Bruzzese, E. Clinical Presentation and Severity of SARS-CoV-2 Infection Compared to Respiratory Syncytial Virus and Other Viral Respiratory Infections in Children Less than Two Years of Age. *Viruses* **2023**, *15*, 717.
2. De Souza, T.H.; Nadal, J.A.; Nogueira, R.J.N.; Pereira, R.M.; Brandão, M.B. Clinical manifestations of children with COVID-19: A systematic review. *Pediatr. Pulmonol.* **2020**, *55*, 1892–1899.

3.	Aykac, K.; Cura Yayla, B.C.; Ozsurekci, Y.; Evren, K.; Oygar, P.D.; Gurlevik, S.L.; Coskun, T.; Tasci, O.; Kaya, F.D.; Fidanci, I.; et al. The association of viral load and disease severity in children with COVID-19. *J. Med. Virol.* **2021**, *93*, 3077–3083. [PubMed]

4.	Brandi, N.; Ciccarese, F.; Rimondi, M.R.; Balacchi, C.; Modolon, C.; Sportoletti, C.; Renzulli, M.; Coppola, F.; Golfieri, R. An Imaging Overview of COVID-19 ARDS in ICU Patients and Its Complications: A Pictorial Review. *Diagnostics* **2022**, *12*, 846. [PubMed]

5.	Pousa, P.A.; Mendonça, T.S.C.; Oliveira, E.A.; Simões-E-Silva, A.C. Extrapulmonary manifestations of COVID-19 in children: A comprehensive review and pathophysiological considerations. *J. Pediatr.* **2021**, *97*, 116–139.

6.	Umakanthan, S.; Sahu, P.; Ranade, A.V.; Bukelo, M.M.; Rao, J.S.; Abrahao-Machado, L.F.; Dahal, S.; Kumar, H.; Kv, D. Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). *Postgrad Med. J.* **2020**, *96*, 753–758. [PubMed]

7.	Guan, W.J.; Ni, Z.Y.; Hu, Y.; Liang, W.H.; Ou, C.Q.; He, J.X.; Liu, L.; Shan, H.; Lei, C.L.; Hui, D.S.; et al. Clinical characteristic of Coronavisus disease 2019 China. *N. Engl. J. Med.* **2020**, *382*, 1708–1720. [PubMed]

8.	Qi, K.; Zeng, W.; Ye, M.; Zheng, L.; Song, C.; Hu, S.; Duan, C.; Wei, Y.; Peng, J.; Zhang, W.; et al. Clinical, laboratory, and imaging features of pediatric COVID-19: A systematic review and meta-analysis. *Medicine* **2021**, *100*, e25230.

9.	Zanza, C.; Romenskaya, T.; Manetti, A.C.; Franceschi, F.; La Russa, R.; Bertozzi, G.; Maiese, A.; Savioli, G.; Volonnino, G.; Longhitano, Y. Cytokine Storm in COVID-19: Immunopathogenesis and Therapy. *Medicina* **2022**, *58*, 144.

10.	Balacchi, C.; Brandi, N.; Ciccarese, F.; Coppola, F.; Lucidi, V.; Bartalena, L.; Parmeggiani, A.; Paccapelo, A.; Golfieri, R. Comparing the first and the second waves of COVID-19 in Italy: Differences in epidemiological features and CT findings using a semi-quantitative score. *Emerg. Radiol.* **2021**, *28*, 1055–1061.

11.	Dobrijević, D.; Katanić, J.; Todorović, M.; Vučković, B. Baseline laboratory parameters for preliminary diagnosis of COVID-19 among children: A cross-sectional study. *Sao Paulo Med. J.* **2022**, *140*, 691–696. [PubMed]

12.	Khanna, N.N.; Maindarkar, M.A.; Viswanathan, V.; Fernandes, J.F.E.; Paul, S.; Bhagawati, M.; Ahluwalia, P.; Ruzsa, Z.; Sharma, A.; Kolluri, R.; et al. Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment. *Healthcare* **2022**, *10*, 2493. [PubMed]

13.	Roy, S.; Meena, T.; Lim, S.J. Demystifying Supervised Learning in Healthcare 4.0: A New Reality of Transforming Diagnostic Medicine. *Diagnostics* **2022**, *12*, 2549.

14.	Paul, S.; Riffat, M.; Yasir, A.; Mahim, M.N.; Sharnali, B.Y.; Naheen, I.T.; Rahman, A.; Kulkarni, A. Industry 4.0 Applications for Medical/Healthcare Services. *J. Sens. Actuator Netw.* **2021**, *10*, 43.

15.	Dobrijević, D.; Antić, J.; Rakić, G.; Andrijević, L.; Katanić, J.; Pastor, K. Could platelet indices have diagnostic properties in children with COVID-19? *J. Clin. Lab. Anal.* **2022**, *36*, e24749. [PubMed]

16.	Kassania, S.H.; Kassanib, P.H.; Wesolowskic, M.J.; Schneidera, K.A.; Detersa, R. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. *Biocybern. Biomed. Eng.* **2021**, *41*, 867–879.

17.	Chadaga, K.; Chakraborty, C.; Prabhu, S.; Umakanth, S.; Bhat, V.; Sampathila, N. Clinical and Laboratory Approach to Diagnose COVID-19 Using Machine Learning. *Interdiscip. Sci. Comput. Life Sci.* **2022**, *14*, 452–470.

18.	Horvat, M.; Horvat, Z.; Pastor, K. Multivariate analysis of water quality parameters in Lake Palic, Serbia. *Env. Monit. Assess* **2021**, *193*, 410.

19.	Syeda, H.B.; Syed, M.; Sexton, K.W.; Syed, S.; Begum, S.; Syed, F.; Prior, F.; Jr, F.Y. Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review. *JMIR Med. Inform.* **2021**, *9*, e23811.

20.	Isaacs, D. Artificial intelligence in health care. *J. Paediatr. Child. Health* **2020**, *56*, 1493–1495.

21.	Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.H. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed.; Elsevier: Amsterdam, The Netherlands, 2016.

22.	Le, D.N.; Parvathy, V.S.; Gupta, D.; Khanna, A.; Rodrigues, J.; Shankar, K. IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification. *Int. J. Mach. Learn. Cyber* **2021**, *12*, 3235–3248.

23.	Maia, M.; Pimentel, J.S.; Pereira, I.S.; Gondim, J.; Barreto, M.E.; Ara, A. Convolutional Support Vector Models: Prediction of Coronavirus Disease Using Chest X-rays. *Information* **2020**, *11*, 548.

24.	Guhathakurata, S.; Kundu, S.; Chakraborty, A.; Banerjee, J.S. A novel approach to predict COVID-19 using support vector machine. In *Data Science for COVID-19*; Academic Press: Cambridge, MA, USA, 2021; pp. 351–364.

25.	Gao, Y.; Cai, G.Y.; Fang, W.; Li, H.Y.; Wang, S.Y.; Chen, L.; Yu, Y.; Liu, D.; Xu, S.; Cui, P.-F.; et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat. Commun.* **2020**, *11*, 5033. [PubMed]

26.	Wan, T.K.; Huang, R.X.; Tulu, T.W.; Liu, J.D.; Vodencarevic, A.; Wong, C.W.; Chan, K.-H.K. Identifying Predictors of COVID-19 Mortality Using Machine Learning. *Life* **2022**, *12*, 547. [PubMed]

27.	Inokuchi, R.; Iwagami, M.; Sun, Y.; Sakamoto, A.; Tamiya, N. Machine learning models predicting undertriage in telephone triage. *Ann. Med.* **2022**, *54*, 2990–2997. [PubMed]

28.	Dobrijević, D.; Andrijević, L.; Antić, J.; Rakić, G.; Pastor, K. Hemogram-based decision tree models for discriminating COVID-19 from RSV in infants. *J. Clin. Lab. Anal.* **2023**, *37*, e24862.

29.	Gupta, V.K.; Gupta, A.; Kumar, D.; Sardana, A. Prediction of COVID-19 Confirmed, Death, and Cured Cases in India Using Random Forest Model. *Big Data Min. Anal.* **2021**, *4*, 116–123.

30.	Dobrijević, D.; Antić, J.; Rakić, G.; Katanić, J.; Andrijević, L.; Pastor, K. Clinical hematochemical parameters in differential diagnosis between pediatric SARS-CoV-2 and Influenza virus infection: An automated machine learning approach. *Children* **2023**, *10*, 761.

31. Wang, J.; Yu, H.; Hua, Q.; Jing, S.; Liu, Z.; Peng, X.; Cao, C.; Luo, Y. A descriptive study of random forest algorithm for predicting COVID-19 patients outcome. *PeerJ* **2020**, *8*, e9945.

32. Cornelius, E.; Akman, O.; Hrozencik, D. COVID-19 Mortality Prediction Using Machine Learning-Integrated Random Forest Algorithm under Varying Patient Frailty. *Mathematics* **2021**, *9*, 2043.

33. Barbosa, V.A.F.; Gomes, J.C.; de Santana, M.A.; de Lima, C.L.; Calado, R.B.; Bertoldo Júnior, C.R.; Albuquerque, J.E.D.A.; de Souza, R.G.; de Araújo, R.J.E.; Júnior, L.A.R.M.; et al. Covid-19 rapid test by combining a Random Forest-based web system and blood tests. *J. Biomol. Struct. Dyn.* **2022**, *40*, 11948–11967. [PubMed]

34. Iwendi, C.; Bashir, A.K.; Peshkar, A.; Sujatha, R.; Chatterjee, J.M.; Pasupuleti, S.; Mishra, R.; Pillai, S.; Jo, O. COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm. *Front. Public Health* **2020**, *8*, 357.

35. Çubukçu, H.C.; Topcu, D.İ.; Bayraktar, N.; Gülşen, M.; Sarı, N.; Arslan, A.H. Detection of COVID-19 by Machine Learning Using Routine Laboratory Tests. *Am. J. Clin. Pathol.* **2022**, *157*, 758–766. [PubMed]

36. Thimoteo, L.M.; Vellasco, M.M.; Amaral, J.; Figueiredo, K.; Yokoyama, C.L.; Marques, E. Explainable Artificial Intelligence for COVID-19 Diagnosis Through Blood Test Variables. *J. Control Autom. Electr. Syst.* **2022**, *33*, 625–644.

37. Van Kasteren, P.B.; van der Veer, B.; van den Brink, S.; Wijsman, L.; de Jonge, J.; van den Brandt, A.; Molenkamp, R.; Reusken, C.B.; Meijer, A. Comparison of seven commercial RT-PCR diagnostic kits for COVID-19. *J. Clin. Virol.* **2020**, *128*, 104412.

38. Yamayoshi, S.; Sakai-Tagawa, Y.; Koga, M.; Akasaka, O.; Nakachi, I.; Koh, H.; Maeda, K.; Adachi, E.; Saito, M.; Nagai, H.; et al. Comparison of Rapid Antigen Tests for COVID-19. *Viruses* **2020**, *12*, 1420.

39. Scohy, A.; Anantharajah, A.; Bodéus, M.; Kabamba-Mukadi, B.; Verroken, A.; Rodriguez-Villalobos, H. Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis. *J. Clin. Virol.* **2020**, *129*, 104455. [PubMed]

40. Albert, E.; Torres, I.; Bueno, F.; Huntley, D.; Molla, E.; Fernández-Fuentes, M.Á.; Martínez, M.; Poujois, S.; Forqué, L.; Valdivia, A.; et al. Field evaluation of a rapid antigen test (Panbio™ COVID-19 Ag Rapid Test Device) for COVID-19 diagnosis in primary healthcare centres. *Clin. Microbiol. Infect.* **2021**, *27*, e7–e472.

41. Goodman-Meza, D.; Rudas, A.; Chiang, J.N.; Adamson, P.C.; Ebinger, J.; Sun, N.; Botting, P.; Fulcher, J.A.; Saab, F.G.; Brook, R.; et al. A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity. *PLoS ONE* **2020**, *15*, e0239474.

42. Assaf, D.; Gutman, Y.; Neuman, Y.; Segal, G.; Amit, S.; Gefen-Halevi, S.; Shilo, N.; Epstein, A.; Mor-Cohen, R.; Biber, A.; et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern. Emerg. Med.* **2020**, *15*, 1435–1443.

43. Zhang, R.K.; Xiao, Q.; Zhu, S.L.; Lin, H.Y.; Tang, M. Using different machine learning models to classify patients into mild and severe cases of COVID-19 based on multivariate blood testing. *J. Med. Virol.* **2022**, *94*, 357–365. [PubMed]

44. Sztefko, K.; Beba, J.; Mamica, K.; Tomasik, P. Blood loss from laboratory diagnostic tests in children. *Clin. Chem. Lab. Med.* **2013**, *51*, 1623–1626. [PubMed]

45. Tajbakhsh, A.; Jaberi, K.R.; Hayat, S.M.G.; Sharifi, M.; Johnston, T.P.; Guest, P.C.; Jafari, M.; Sahebkar, A. Age-Specific Differences in the Severity of COVID-19 Between Children and Adults: Reality and Reasons. *Adv. Exp. Med. Biol.* **2021**, *1327*, 63–78.

46. Wong, L.S.Y.; Loo, E.X.L.; Kang, A.Y.H.; Lau, H.X.; Tambyah, P.A.; Tham, E.H. Age-Related Differences in Immunological Responses to SARS-CoV-2. *J. Allergy Clin. Immunol. Pract.* **2020**, *8*, 3251–3258.

47. Carobene, A.; Milella, F.; Famiglini, L.; Cabitza, F. How is test laboratory data used and characterised by machine learning models? A systematic review of diagnostic and prognostic models developed for COVID-19 patients using only laboratory data. *Clin. Chem. Lab. Med.* **2022**, *60*, 1887–1901.

48. Lippi, G. Machine learning in laboratory diagnostics: Valuable resources or a big hoax? *Diagnosis* **2019**, *8*, 133–135.

49. Núñez, I.; Belaunzarán-Zamudio, P.F.; Caro-Vega, Y. Result Turnaround Time of RT-PCR for SARS-CoV-2 is the Main Cause of COVID-19 Diagnostic Delay: A Country-Wide Observational Study of Mexico and Colombia. *Rev. Investig. Clin.* **2022**, *74*, 071–080.

50. Kogoj, R.; Korva, M.; Knap, N.; Resman Rus, K.; Pozvek, P.; Avšič-Županc, T.; Poljak, M. Comparative Evaluation of Six SARS-CoV-2 Real-Time RT-PCR Diagnostic Approaches Shows Substantial Genomic Variant-Dependent Intra- and Inter-Test Variability, Poor Interchangeability of Cycle Threshold and Complementary Turn-Around Times. *Pathogens* **2022**, *11*, 462.

51. Minhas, N.; Gurav, Y.K.; Sambhare, S.; Potdar, V.; Choudhary, M.L.; Bhardwaj, S.D.; Abraham, P. Cost-analysis of real time RT-PCR test performed for COVID-19 diagnosis at India's national reference laboratory during the early stages of pandemic mitigation. *PLoS ONE* **2023**, *18*, e0277867.

52. Alyasseri, Z.A.A.; Al-Betar, M.A.; Doush, I.A.; Awadallah, M.A.; Abasi, A.K.; Makhadmeh, S.N.; Alomari, O.A.; Abdulkareem, K.H.; Adam, A.; Damasevicius, R. Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. *Expert Syst.* **2022**, *39*, e12759. [PubMed]

53. Heidari, A.; Jafari Navimipour, N.; Unal, M.; Toumaj, S. The COVID-19 epidemic analysis and diagnosis using deep learning: A systematic literature review and future directions. *Comput. Biol. Med.* **2022**, *141*, 105141. [PubMed]