Research on Biomedical Engineering (2023) 39:509–539
https://doi.org/10.1007/s42600-023-00286-8

ORIGINAL ARTICLE

Rapid protocols to support COVID-19 clinical diagnosis based on hematological parameters

Juliana Carneiro Gomes[1] · Valter Augusto de Freitas Barbosa[2,3] · Maíra Araújo de Santana[1] · Clarisse Lins de Lima[1] · Raquel Bezerra Calado[3] · Cláudio Roberto Bertoldo Júnior[3] · Jeniffer Emidio de Almeida Albuquerque[3] · Rodrigo Gomes de Souza[3] · Ricardo Juarez Escorel de Araújo[3] · Giselle Machado Magalhães Moreno[3] · Luiz Alberto Lira Soares[3] · Luiz Alberto Reis Mattos Júnior[3] · Ricardo Emmanuel de Souza[3] · Wellington Pinheiro dos Santos[3]

Received: 10 January 2022 / Accepted: 22 May 2023 / Published online: 3 June 2023

Abstract

Purpose In December 2019, the Covid-19 pandemic began in the world. To reduce mortality, in addiction to mass vaccination, it is necessary to massify and accelerate clinical diagnosis, as well as creating new ways of monitoring patients that can help in the construction of specific treatments for the disease.

Objective In this work, we propose rapid protocols for clinical diagnosis of COVID-19 through the automatic analysis of hematological parameters using evolutionary computing and machine learning. These hematological parameters are obtained from blood tests common in clinical practice.

Method We investigated the best classifier architectures. Then, we applied the particle swarm optimization algorithm (PSO) to select the most relevant attributes: serum glucose, troponin, partial thromboplastin time, ferritin, D-dimer, lactic dehydrogenase, and indirect bilirubin. Then, we assessed again the best classifier architectures, but now using the reduced set of features. Finally, we used decision trees to build four rapid protocols for Covid-19 clinical diagnosis by assessing the impact of each selected feature. The proposed system was used to support clinical diagnosis and assessment of disease severity in patients admitted to intensive and semi-intensive care units as a case study in the city of Paudalho, Brazil.

Results We developed a web system for Covid-19 diagnosis support. Using a 100-tree random forest, we obtained results for accuracy, sensitivity, and specificity superior to 99%. After feature selection, results were similar. The four empirical clinical protocols returned accuracies, sensitivities and specificities superior to 98%.

Conclusion By using a reduced set of hematological parameters common in clinical practice, it was possible to achieve results of accuracy, sensitivity, and specificity comparable to those obtained with RT-PCR. It was also possible to automatically generate clinical decision protocols, allowing relatively accurate clinical diagnosis even without the aid of the web decision support system.

Keywords Covid-19 · Clinical diagnosis support · Covid-19 rapid protocols · Hematological parameters · Software-based rapid test · Computer-aided diagnosis

✉ Wellington Pinheiro dos Santos
wellington.santos@ufpe.br

Juliana Carneiro Gomes
jcg@ecomp.poli.br

Valter Augusto de Freitas Barbosa
valter.barbosa@ufrpe.br

Maíra Araújo de Santana
mas2@ecomp.poli.br

Clarisse Lins de Lima
cll@ecomp.poli.br

[1] Polytechnique School of the University of Pernambuco, Recife, Brazil

[2] Academic Unit of Serra Talhada, Rural Federal University of Pernambuco, Serra Talhada, Brazil

[3] Federal University of Pernambuco, Recife, Brazil

Introduction

Motivation and problem characterization

The world that emerged after the Second World War was marked by a rapid process of globalization. An interconnected world has emerged, totally connected both by advanced means of transport, such as airplanes and ships, and by means of information and communication technologies. Trade necessarily integrates nations and intensifies the movement of people across the globe. However, from the point of view of Epidemiology, a fully connected world is also a world more susceptible to several threats, including health threats, such as epidemics and pandemics (Crooks et al. 2018, Prior et al. 2019). The pathways through which international trade flows are also the pathways used by infectious disease vectors.

In December 2019, in the city of Wuhan, China, the most critical outbreak in the last hundred years began: Coronavirus Disease 2019 (Covid-19), transmitted by the SARS-CoV2 virus, a virus of zoonotic origin until then unknown, present in bats and pangolins (Barbosa et al. 2021b, Ciotti et al. 2019; de Lima et al. 2020; Rothan and Byrareddy 2020; Wu et al. 2020). SARS-CoV-2, when compared to its predecessors, proved to be much more resistant and infectious. The most common symptoms are fever, dry cough, and tiredness (Ciotti et al. 2019, de Lima et al. 2020, Gomes et al. 2020a, Rothan and Byrareddy 2020, Wu et al. 2020). Pain and discomfort, sore throat, diarrhea, conjunctivitis, headache, loss of taste or smell, rash on the skin, or discoloration of the fingers or toes may also appear (Cascella et al. 2020; Ciotti et al. 2019; Peeri et al. 2020; Rothan and Byrareddy 2020; Wang et al. 2020b; Wu et al. 2020). Its severe symptoms are difficulty in breathing or shortness of breath, pain or pressure in the chest, and loss of speech or movement (Cascella et al. 2020, Ciotti et al. 2019, de Lima et al. 2020, Gomes et al. 2020a, Peeri et al. 2020, Rothan and Byrareddy 2020, Wang et al. 2020b, Wu et al. 2020). Despite the lower lethality, the virus spreads very quickly, producing a large volume of deaths and leaving sequels that are often permanent (Cascella et al. 2020; de Lima et al. 2020; Peeri et al. 2020; Wang et al. 2020b). Due to their high rate of contagion, public health system resources are rapidly depleted (de Lima et al. 2020). The Covid-19 pandemic is one of the biggest health crises in decades. In March 2021, SARS-CoV had already infected almost 130 million people, more than seventy million of whom recovered, while almost 3 million died (Organization 2021). In this context, it is not enough to invest in the opening of new hospital beds for the treatment of patients. It is necessary to have tests that guarantee fast and reliable diagnoses; specific treatments to decrease the lethality of the disease; efficient and low-cost vaccines applied to a considerable portion of the population; and social isolation and quarantine policies to seek to control the disease vector while vaccines and specific treatments are not available for Covid-19.

Several studies have sought to highlight the nature of Covid-19 as a disease that affects the cardiovascular system (Chatterjee et al. 2020; Fan et al. 2020; Gao et al. 2020; Liu et al. 2020; Tan et al. 2020; Zheng et al. 2020). Coronaviruses, such as SARS-CoV and SARS-CoV-2, have the angiotensin-converting zinc metallopeptidase 2 (ACE2), an enzyme present in the cell membranes of the arteries, heart, lungs and other organs as a functional receptor. ACE2 is involved in cardiac function, hypertension and diabetes (Turner et al. 2004). The MERS-CoV and SARS-CoV coronaviruses can cause acute myocarditis and heart failure (Zheng et al. 2020). Some of the impacts of coronaviruses on the cardiovascular system are increased blood pressure and increased levels of troponin I (hs-cTnI) (Zheng et al. 2020). Covid-19 patients may also develop lymphopenia, i.e., low level of lymphocytes (Fan et al. 2020; Liu et al. 2020; Tan et al. 2020); and leukopenia, i.e., few leukocytes. COVID-19 patients may also experience decreased hemoglobin levels, absolute lymphocyte count (ALC), and absolute monocyte count (AMC) (Fan et al. 2020). Patients who have developed severe forms of the disease have significantly higher levels of Interleukin-6 and D-dimer than patients who have developed a moderate form of COVID-19 (Gao et al. 2020). Therefore, considering that COVID-19 is a disease that affects blood parameters, hematological tests can be used to help diagnose the disease.

Given that the result of the RT-PCR test can take hours or even days, given the pandemic situation that increases the demand for tests with the high speed of contamination, clinical diagnosis assumes a fundamental role in determining the treatment and the type of care correct for mild, moderate and severe cases. In this sense, the observation of complementary clinical parameters, such as the hematological parameters obtained from common tests in clinical practice, takes on an important role. Several works have been using machine learning techniques to diagnosis diseases by analyzing hematological parameters automatically (Barbosa et al. 2021b, Guncar et al. 2018; Luo et al. 2016, Tanner˘ et al. 2008). Blood tests are commonly used during medical screening. The most common blood tests, like complete blood count, bilirubin, serum glucose, C-reactive protein, urea, and others, are easily available at low-cost compared with other diagnosis methods. Therefore, intelligent systems can be used to automatically analyze hematological parameters and use them to support COVID-19's clinical diagnosis and to suggest appropriate patient care (Barbosa et al. 2021b).

In this work we propose two main applications to support the clinical diagnosis of COVID-19 based on hematological parameters: (1) rapid diagnosis based on automatic analysis by an intelligent web system based on particle swarm optimization and random forests; (2) a tool to build clinical analysis protocols, to be used in contexts where the intelligent rapid diagnosis system is not available. In the first approach, an intelligent web system was developed to support decision making. This system was put into operation in the city of Paudalho, Brazil, to support the clinical diagnosis and assessment of the severity of the disease in patients admitted to intensive and semi-intensive care units, in the year 2020. In the second approach, we used decision trees to infer which hematological parameters would be statistically more relevant for clinical diagnosis. In the particular case of COVID-19, the most relevant hematological parameters were ferritin and prothrombin partial time. From these results, these hematological parameters can be inserted as a new medical protocol in the investigation of suspected cases of COVID-19. Finally, the results of general accuracy, sensitivity, and specificity were quite high for both approaches, demonstrating that rapid diagnosis is possible using only well-known and low-cost clinical exams, with the potential support of an intelligent decision support system. The high correlation between ferritin, prothrombin time and a positive diagnosis for COVID-19 in symptomatic patients also points to the need for more research on treatments to combat early clotting and the main symptoms of the disease.

Related works

Several studies emphasize the importance of hematological parameters to support COVID-19 clinical diagnosis. Some of them point to the relevance of using hematological analysis as an indicative of the severity degree of COVID-19. Fan et al. (2020) analyzed hematological parameters of 69 patients with COVID -19. The study was conducted with subjects from the National Center for Infectious Diseases (NCID) in Singapore. 65 of these patients underwent complete blood count (CBC) on the day of admission. 13.4% of patients needed intensive care unit (ICU) care, especially the elderly. During the first exams, 19 patients had leukopenia and 24 had lymphopenia, with 5 cases classified as severe (Absolute Lymphocyte Count (ALC) $< 0.5 \times 10^9/L$). The study also pointed out that patients who needed to be admitted to the ICU had lower ALC and a higher rate of Lactate Dehydrogenase (LDH). These data indicated that monitoring these hematological parameters can help to identify patients who need assistance in the ICU. The authors found that the patients who were in the ICU had a significant decrease in their hemoglobin levels, ALC and Absolute Monocyte Count (AMC) levels, when compared to the non-ICU group. ICU patients also tend to develop neutrophilia. The platelet count did not prove to be a factor for discrimination between the type of hospitalization.

Gao et al. (2020) assessed hematological characteristics of 43 patients at Fuyang Second People's Hospital. The patients had diagnosis confirmed by the COVID-19 ground truth test, the fluorescent reverse transcription-polymerase chain reaction (RT-PCR). They were divided into two groups: the moderate group with 28 patients, and the severe group with 15 patients. The groups have no significant difference in age and sex, and both were evaluated by routine blood tests and blood biochemistry parameters. Thereby, from statistical tests, the study noted that the levels of glucose (GLU), C-reactive protein (CRP), interleukin-6 (IL-6), thrombin time (TT), fibrinogen (FIB), and D-dimer were significantly higher in the severe group than in the mild group. Performing this analysis with ROC curves, the authors pointed out that the best indicators for predicting severity were IL-6 and d-D combined, with AUC of 0.840, specificity of 96.4%, and sensitivity of 93.3%. These results indicate that patients with severe conditions would have abnormal coagulation.

Liu et al. (2020) reported that lymphopenia and inflammatory cytokine storm are abnormalities commonly found in other infections caused by coronavirus, such as SARS-Cov and MERS-Cov. With that in mind, they studied 40 patients with positive RT-PCR for COVID-19 at Wuhan Union Hospital. The information provided was epidemiological, demographic, clinical manifestations, and laboratory tests. Similar to the previous study, patients were divided into two groups: mild patients, with symptoms such as epidemiological history, fever or respiratory symptoms, and abnormalities in imaging tests; the second group with severe patients, patients should additionally have symptoms such as shortness of breath, oxygen saturation $< 93\%$, respiratory > 30 times/min, or $PaO_2/FiO_2 < 300$ mmHg. A total of 27 patients were classified in the first group, while 13 were classified in the second. The study reported that levels of fibrinogen, D-dimer, total bilirubin, aspartate transaminase, alanine transaminase, lactate dehydrogenase, creatine kinase, C-reactive protein (CRP), ferritin, and serum amyloid A protein were significantly higher in the severe group. Futhermore, most severe patients presented lymphopenia, while white blood cells and neutrophils counts were higher.

These studies have pointed out that hematological parameters can be indicators of the risk factors and degree of severity of COVID-19. The identification of these clinical parameters can be essential to optimize clinical care for each group of patients. In this sense, the development of intelligent systems based on blood tests is useful. Faced with the pandemic scenario, in which most hospitals are full, decision support systems can facilitate clinical management. Thus, it can increase the assertiveness in the treatment for each case and, consequently, the number of lives saved.

Regarding the use of intelligent systems to support COVID-19 diagnosis, Saba and Khan (2021) present a collection of 14 chapters focused on COVID-19 solutions. From these chapters, we highlight the following works related to our proposal: Khan et al. (2021) present a review of the state-of-the-art of the intelligent solutions to support COVID-19 image diagnosis by applying deep learning methods over x-ray thoracic images and computerized tomography x-ray images. Hassan et al. (2021) made a selective assessment of the latest research articles relevant to the applications of artificial intelligence and machine learning (ML) techniques from the databases of the Web of Science, Scopus, and Google Scholar, using keywords of coronavirus, artificial intelligence, and ML. The authors analyzed the use of computed tomography (CT) imaging, X-ray, and magnetic resonance imaging (MRI) and their applications COVID-19 pandemics. Karimi et al. (2021) and Nourbakhsh (2021) present reviews of the state-of-the-art of the use of computational intelligence, on computerized tomography x-ray imaging to support COVID-19 diagnosis and progress evaluation.

Fei et al. (2021) highlighted the applications of modern digital technology, statistical methods, data platforms and data integration systems to improve diagnosis and treatment of diseases in clinical research and novel epidemiologic tools to tackle infection source problems, such as finding Patient Zero in the spread of epidemics. The authors concluded that analyzing and interpreting COVID-19 big data is an incredibly challenging task that requires a multi-disciplinary effort to continuously create more effective methodologies and powerful tools to transfer data information into knowledge that enables informed decision making.

Pani et al. (2022) present a large collection of intelligent solutions in the fight against COVID-19 regarding disease prediction (Pandit et al. 2022; Torcate et al. 2022), temporal and spatiotemporal forecasting (Biswas and Dash 2022; da Silva et al. 2022; Pandit et al. 2022), and intelligent diagnosis on computerized tomography x-ray imaging (Adetunji et al. 2022c; de Santana et al. 2022; Majhi et al. 2022; Pandit et al. 2022), thoracic x-ray imaging (Adetunji et al. 2022c; Majhi et al. 2022; Pandit et al. 2022; Usharani 2022), electrical impedance tomography imaging (Wolff et al. 2022), signals and symptoms (Adetunji et al. 2022a,d; Pandit et al. 2022; Peter et al. 2022; Tripathi et al. 2022), molecular analysis (Adetunji et al. 2022c; Pani et al. 2022; Sharma et al. 2022), and hematological parameters (Torcate et al. 2022) based on ML (shallow and deep algorithms) Adetunji et al. (2022b, c, e) and swarm intelligence Usharani (2022).de Santana et al. (2022) proposed an automatic system for COVID-19 diagnosis using ML techniques and CT X-ray images named IKONOS-CT, an intelligent system dedicated to provide a binary classification, differentiating COVID-19 patients from non- COVID-19 ones. For classification tasks, the authors performed 25 experiments for the following classifiers: multilayer perceptron (MLP), support vector machines (SVM), random tree and random forest, and Bayesian networks. The best overall performance was reached using Haralick as feature extractor and SVM with polynomial kernel of exponent 3. The authors found the following results: accuracy of 96.994% (1.375), sensitivity/recall of 0.952 (0.024), and specificity of 0.987 (0.014). The authors claim that, by using a computationally low-cost method based on the Haralick texture features extractor, it was possible to achieve high diagnosis performance. These experimental results point out that an effective path for COVID-19 diagnosis is composed by the combination of artificial intelligence and human-based clinical analysis.

Torcate et al. (2022) wanted to make predictions regarding treatment and assessments of severity of patients with and without COVID-19 based on blood tests. The authors used hematological data from patients who attended the units of the Brazilian public healthcare system in the city of Paudalho in 2020. Their objective was to analyze intelligent classifiers able to make hospitalization predictions considering three scenarios: regular ward, semi-intensive care unit, and intensive care unit, corresponding to mild (non-critical), moderate, and serious cases. The results obtained in the experiments show that the classifiers managed to perform better with the database balanced with the SMOTE method, and the 100-tree random forest demonstrated the best potential to perform the predictions for regular ward (sensitivity of 0.730, and specificity of 0.913), as for the semi-intensive care unit (sensitivity of 0.890, and specificity of 0.875), and intensive care unit (sensitivity of 0.640, and specificity of 0.947). The authors intended to assist health professionals in decision making, aiming to streamline the process of directing patients to the most suitable care units for each patient.

Guncar et al. (2018) proposed a system based on ML for analyzing blood tests and˘ predicting hematological diseases. Their database was acquired between the years 2005 and 2015 at the University Medical Center of Ljubljana. In this case, 43 diseases and 181 clinical parameters or features were selected to generate a first model (SBA-HEM181). In addition to it, a second model with 61 parameters was also developed (SBA-HEM061). The selection criteria were based on the frequency of use. Regarding the missing values (about 75%), the authors filled in with median values for each attribute. As classification methods, the authors tested classic approaches, such as SVM, Naive Bayes, and random forest. The simulations were repeated 10 times using tenfold cross validation. Finally, the models SBA-HEM181 and SBA-HEM061 reached an accuracy of 57% considering all the diseases chosen. By restricting the prediction to five classes, the systems achieved an accuracy of 88% and 86%, respectively, when using random forest for classification. This study also pointed to the possibility of effectively

detecting diseases through blood tests using classic intelligent classifiers.

Barbosa et al. (2021b) proposed an intelligent system to support COVID-19 diagnosis using blood exams. Heg.IA was developed in a context where it was necessary and urgent to develop a diagnostic support system that would provide rapid results with high sensitivity and high specificity. In this context, blood tests have some advantages. First, they are commonly used during medical screening. In addition, blood tests are cheaper and less time consuming than other diagnostic methods, making the system more accessible. By combining these blood test results with analysis based on artificial intelligence, the authors built a relatively robust, efficient and easily available system for diagnosing COVID-19. The intelligent system was built using a 5644-subjects public database from the Hospital Israelita Albert Einstein, Brazil (Kaggle 2020). The authors also optimized the system to reduce the number of tests needed, based on its relevance to describing the diagnostic problem and its price and availability worldwide, especially in low-income communities. First, an automatic selection of exams was performed using the particle swarm optimization method. Then, the authors manually selected some exams with the objective of reaching an ideal combination of price, time and number of procedures. This procedure resulted in 24 blood tests, which can be delivered in up to 1 h. As computational classification can be performed in milliseconds, with the 24 blood test results, a technician can obtain diagnostic results relatively quickly. The user just needs to fill the electronic form with these 24 blood test results. Using classic Bayesian networks, the system returned high diagnostic performance: 95.159% (0.693) overall accuracy, kappa index of 0.903 (0.014), sensitivity of 0.968 (0.007), accuracy of 0.938 (0.010), and specificity of 0.936 (0.010). If compared to methods based on deep learning, the proposed system also reduces the computational cost. Although the study achieved very high results, the set of selected tests did not considered all the exams indicated by subsequent recommendations from the Brazilian Ministry of Health when dealing with COVID-19 patients (Brazilian Ministry of Health Guidelines for the diagnosis and treatment of COVID-19 2020), since these exams were not established as mandatory for the context of clinical diagnosis of COVID-19 in the time of the publication.

Barbosa et al. (2021a) propose the Heg.IA web system, a solution to optimize the diagnosis of COVID-19 through the automatic analysis of blood tests using machine learning. The system aims to support decision-making regarding the diagnosis of COVID-19 and the indication of admission to a regular ward, semi-ICU or ICU based on the decision of a random forest architecture with 90 trees. To this end, healthcare professionals can enter 41 hematological parameters from common blood tests and arterial blood gases into the system. Then the proposed intelligent system provides a diagnostic report. The Heg.IA web system employs blood tests to support the diagnosis of COVID-19. The machine learning method was trained using the database provided by Hospital Israelita Albert Einstein located in São Paulo, Brazil. The database is formed by information from 5644 patients.

Among them, 559 patients were diagnosis with COVID-19 by RT-PCR with DNA sequencing and identification and additional laboratory tests during a visit to the hospital. For each patient, the database has more than one hundred laboratory tests like blood counts and urine test. From this database we set a new one that contains only 41 blood tests recommended by the Brazilian Ministry of Health when dealing with COVID-19 patients. The system achieved good results both for the diagnosis of COVID-19 and for the recommendation of hospitalization. To support the clinical diagnosis of COVID-19, the solution achieved an accuracy of 92.891% (0.851), kappa index of 0.858 (0.017), sensitivity of 0.936 (0.011), precision of 0.923 (0.011), specificity of 0.921 (0.012), and area under ROC curve of 0.984 (0.003). Regarding the indication for hospitalization, the system provided high performance: accuracy above 99% and kappa, sensitivity, specificity, area under ROC curve, and precision above 0.99. Using a method with low computational cost, based on classical decision trees, the authors obtained a high diagnostic performance. The authors claim that the Heg.IA system could be a way to overcome the unavailability of tests in the context of COVID-19. The work of Barbosa et al. (2021a) is a significant improvement of the proposal of Barbosa et al. (2021b).

Similarly, Soares et al. (2020) use a method based on artificial intelligence to identify COVID-19 through blood tests. As Barbosa et al. (2021b) and Barbosa et al. (2021a) did, they used the database from the Hospital Israelita Albert Einstein. However, since the database has many missing data, they chose to include only the subjects that had most of the data. This procedure reduced the dataset from 5644 to 599 samples. By using Support Vector Machines as a classifier and SMOTE Boost technique to perform oversampling, they achieved average specificity of 85.98%, negative predictive value (NPV) of 94.92%, average sensitivity of 70.25%, and positive predictive value (PPV) of 44.96%.

Martinez-Velazquez et al. (2021) present a solution for detecting COVID-19 infections exclusively on the basis of self-reported symptoms. The motivation of the authors was providing a relatively inexpensive and easy to deploy solution at either an individual or population scale. The authors trained and tested a large set of machine learning models to detect COVID-19 based exclusively on symptoms and signs, similar to the process of anamnesis assessment. Martinez-Velazquez et al. (2021) evaluated 15 different classifiers: decision tree (DT), perceptron neural network (NN), and

support vector machine (SVM), random forest (RF), and a voting classifier. The best model was a voting ensemble of a random forest and a decision tree. This model was able to reach a mean area under the ROC curve of 0.728, a sensitivity of 0.752, a specificity of 0.609, and a precision of 0.660. The best classifier among those reported in this study presents a sensitivity and precision higher than health professionals' anamnesis; however, the specificity is lower. The immediate benefit of adopting such an ML-powered assessment approach is to maximize the use of the limited available RT-PCR tests, to find more COVID-19 positive infections within the community by identifying a higher ratio of actual positive cases (higher sensitivity).

The intelligent systems based on blood tests may play a key role in the process of diagnosing COVID-19, since many studies are confirming evidences of how this disease affects hematological parameters. Positive cases can be forwarded for results confirmation through RT-PCR, computerized tomography scans and/or radiography (Gomes et al. 2020a, b).

Negri et al. (2020) have been using elevated D-dimer as a predictor of severity and mortality in COVID-19 patients. They also observed that COVID-19 patient autopsies revealed thrombi in the microvasculature, suggesting that hypercoagulability is a prominent feature of organ failure. In this context, Negri et al. (2020) performed a clinical study involving 27 COVID-19 patients admitted to Sirio-Libanes Hospital in São Paulo, Brazil. They treated these patients with heparin in therapeutic doses tailored to clinical severity.

Following the steps of (Negri et al. 2020), several studies confirmed the importance of D-dimer as severity predictor and specific treatment based on anticoagulants as capable to reduce COVID-19 mortality and improve prognosis (Chatterjee et al. 2020, Menezes-Rodrigues et al. 2020, Sahu and Agrawal 2020, Shi et al. 2020, Tang et al. 2020, Viecca et al. 2020, Wang et al. 2020c). Klok et al. (2020) findings also state that COVID-19 may predispose patients to both venous and arterial thromboembolism due to excessive inflammation, hypoxia, immobilization, and diffuse intravascular coagulation. The authors' findings are strongly suggestive of increasing the prophylaxis towards high-prophylactic doses. For Long et al. (2020) and Panigada et al. (2020), D-dimer and prothrombin time are most significant indicators of severe COVID-19 cases and poor prognosis due to hypercoagulation hypercoagulability together with severe inflammatory states. Other clinical studies confirm evidences to this finding and point out to partial thromboplastin time as other important biomarker (Connors and Levy 2020; Iba et al. 2020; Liao et al. 2020; Pavoni et al. 2020; Spiezia et al. 2020; Wright et al. 2020).

For Gómez-Pastora et al. (2020), as ferritin can be actively secreted at the site of infection, it is possible that ferritin can take on other functions in addition to its classic role as an iron storage protein. Ferritin has been shown to be a signaling molecule and direct mediator of the immune system (Rosário et al. 2013). Complex feedback mechanisms may exist between ferritin and cytokines in the control of pro-inflammatory and anti-inflammatory mediators, as cytokines can induce ferritin expression. Ferritin can also induce the expression of pro- and anti-inflammatory cytokines (Rosário et al. 2013). However, the pathogenic role of ferritin during inflammation remains unclear (Kernan and Carcillo 2017). For Gómez-Pastora et al. (2020), it is necessary to investigate the structure of plasma ferritin in patients with COVID-19.

For Vargas-Vargas and Cortés-Rojo (2020), there is a strong evidence supporting the hypothesis that ferritin levels might be a crucial factor influencing the severity of COVID-19. Lin et al. (2020) found an association between hyperferritinemia and disease severity in patients with COVID-19. Lin et al. (2020) conducted a retrospective study on 147 confirmed COVID-19 patients in Changsha, China. The overall proportion of severe disease was 16.32% (24/147). The severe patients had higher levels of serum ferritin than the non-severe patients. Multivariate logistic regression analysis indicated that the serum ferritin level on admission was an independent risk factor for disease severity in COVID-19 patients. Other studies agree on the importance of serum ferritin as a COVID-19 biomarker (Bataille et al. 2020; Cheng et al. 2020; Dahan et al. 2020; Kappert et al. 2020; Tural Onur et al. 2021).

The general objective of this work is to propose a solution to support the clinical diagnosis of COVID-19 through the introduction of rapid protocols based on the automatic analysis of hematological parameters using machine learning. Hematological parameters are obtained from common blood tests in clinical practice: the complete blood count and biochemical tests. These are low-cost tests that are already mandatory to support clinical diagnosis. The rapid protocols are based on two approaches: (a) the introduction of the Heg.IA web tool in clinical diagnostic practice, through the adoption of diagnostic suggestion reports based on the analysis of hematological parameters using machine learning and swarm intelligence; (b) the generation of rapid protocols to support clinical diagnosis by using offline decision trees. In this work, we present an improved version of the web system Heg.IA, developed by Barbosa et al. (2021a). In turn, the work of Barbosa et al. (2021a) is an improvement of the workstation solution developed by Barbosa et al. (2021b). In this newer version, we introduced the functionality of generating offline protocols based on decision trees, to support clinical diagnosis without having the Heg.IA system available. The motivation was to make clinical diagnosis fast and accurate, previously supported by machine learning, but without the need for a web application. Furthermore, the database used in this updated version was the result of a pilot

project applied in the public health system of the city of Paudalho, Pernambuco, Brazil, during the years 2020 and 2021, while using the older application version. The database is composed of 6215 records obtained from electronic medical records collected from 2019 to 2020. In Paudalho, the Heg.IA system was used not only to support the diagnosis of COVID-19, but also to monitor the progress of moderate and severe cases of COVID-19 in ICU and semi-ICU. The version of the Heg.IA web proposed in this work also has the functionality of predicting the hospitalization of symptomatic patients with COVID-19 (ICU or semi-ICU) based on hematological parameters. These results, however, were published by Torcate et al. (2022).

Materials and methods

Proposed method

In this work, we propose two approaches of rapid protocols to support the clinical diagnosis of COVID-19 based on hematological parameters: (1) rapid diagnosis based on automatic analysis by an intelligent web system based on particle swarm optimization and random forests; (2) rapid diagnosis by statistical inference by decision trees. In the first approach, we developed an intelligent web system for decision COVID-19 diagnostic support: Heg.IA web. This system was based on a standalone solution based on Windows and Linux PCs, Heg.IA (Barbosa et al. 2021b). The system started operating in the city of Paudalho, Brazil, as a prototype and has been used to support clinical diagnosis and assessment of disease severity in patients admitted to intensive and semi-intensive care units in public health units in the city since 2020.

These approaches are called clinical diagnosis support protocols because they are proposed: (a) as part of the clinical decision-making process, integrating the currently existing processes in the form of an intelligent decision support system using machine learning; or (b) as an application of decision trees to automatically build rules for evaluation of hematological parameters and clinical decision making by the human specialist, i.e., the medical professional.

In this approach, we used a knowledge base composed of 6215 patient records seen in health units in the city. The records were represented by age and up to 43 hematological parameters. We do not use biological sex because it is not relevant to the COVID-19 diagnostic problem. Using the particle swarm optimization method, with 20 individuals evolving in 50 generations, we reduced the dimension of the attribute vectors from 43 to 8 statistically more relevant attributes. Both the original knowledge base (43 features) and the simplified knowledge base (8 features) was used to build machine learning models based on the following architectures: multilayer perceptron neural networks (MLP), support vector machines (SVM), naive Bayes classifier, Bayesian network, decision trees, and random forests. The diagram in Fig. 1 shows a summary of this solution. A general method was proposed: patients with characteristic symptoms of COVID-19 are referred to a health unit and are evaluated by a medical team. This team orders blood tests. After obtaining the results, the health professional accesses Heg.IA web. On the website, he must login. Then he can enter the results of the patient's blood tests. Upon completion, the system will generate a report with a positive or negative diagnosis for COVID-19, in addition to the hospitalization forecast. This report can be printed and used by the medical team to define the final clinical conduct.

In the second approach, we use decision trees to find the most significant relationships between hematological parameters and decision making, i.e., whether COVID-19 is positive or not. This process was used to infer that hematological parameters are statistically more relevant for clinical diagnosis without the support of an intelligent system.

All experiments were performed with 30 repetitions of tenfold cross-validation, resulting in 300 computational experiments. As metrics of quality, we use accuracy, sensitivity, specificity, and area under the ROC curve.

Database

The city of Paudalho is located 38 km from the coastal capital Recife, State of Pernambuco, Brazil, has a semi-arid climate (Vilar and Medeiros 2019) and is inserted in the Atlantic Forest biome. It has a territorial area of 269,651 km^2 and its population was around 51,357 in 2010 (last census conducted in Brazil). It is estimated that in 2020 it had 56,933 inhabitants (IBGE 2010). Its demographic density of 185.06 inhabitants/km^2 places it among the 40 most populated municipalities in the state.

Some social determinants of health provide a more functional view of the municipality in question. Directly related to health care, the municipality has 21 SUS health establishments, encompassing primary, secondary and tertiary care (IBGE 2010). SUS is an acronym for "Unique Public Health System," the Brazilian public health system. It presents only 31.5% of households with adequate sanitation (position 132 in the State of Pernambuco) (IBGE 2010) and infant mortality rate of 13.55/1000 live births (IBGE 2010).

This retrospective study used the medical records of patients provided by the Health Secretariat of the city of Paudalho as a database. All procedures for this research were approved by the Research Ethics Committee at the Federal University of Pernambuco under number CAAE 34932020.3.0000.5208.

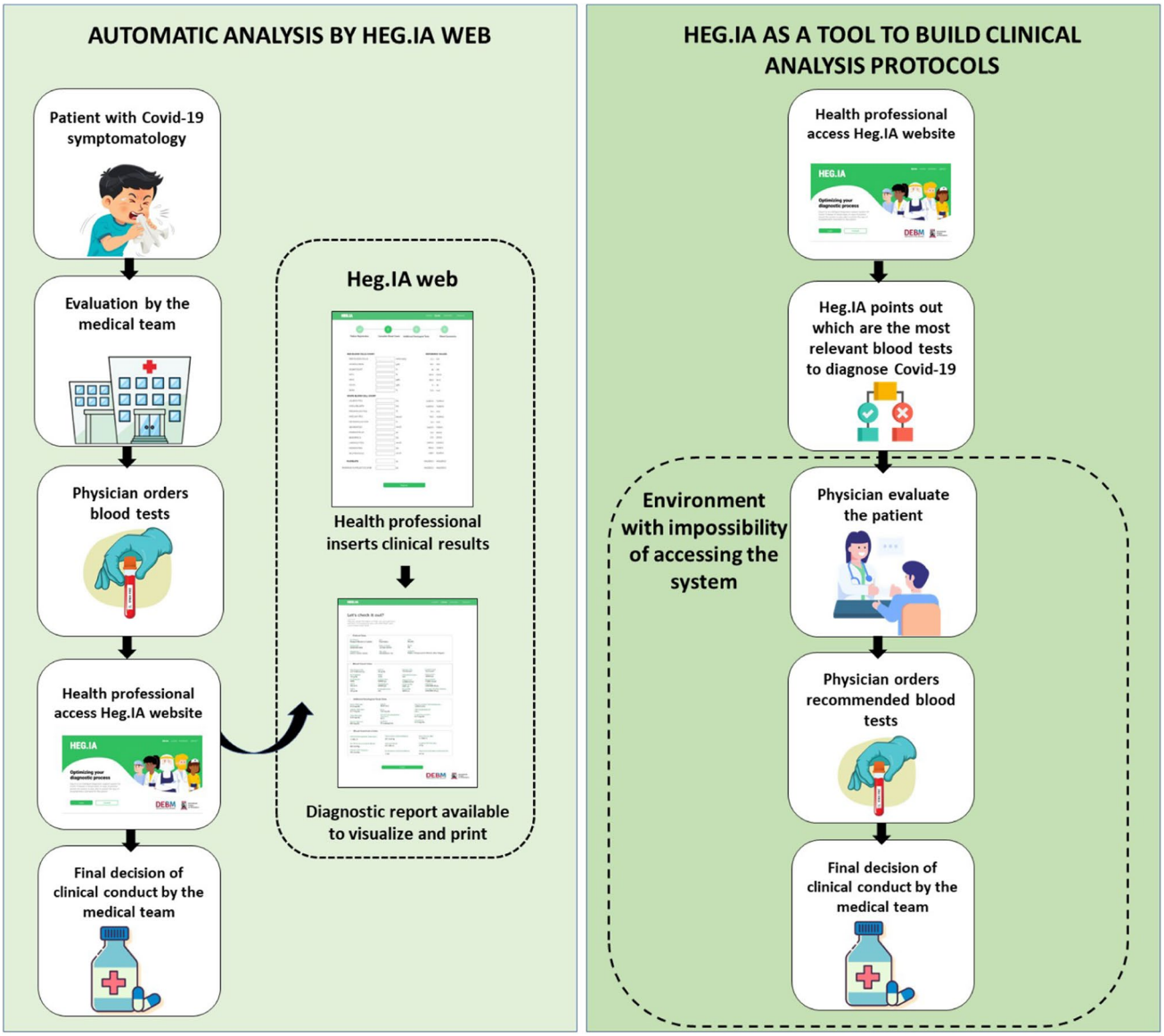

Fig. 1 General method proposed: The system can be used in two different scenarios. The first scenario consists of using the system as a diagnostic support tool. In this case, patients with characteristic symptoms of COVID-19 are referred to a health unit and are evaluated by a medical team. This team orders blood tests. After obtaining the results, the health professionals access the HegIA web. On the website, they must login. Then these health professionals can enter the results of the patient's blood tests. Upon completion, the system will generate a report with a positive or negative diagnosis for COVID-19, in addition to the hospitalization forecast. This report can be printed and used by the medical team to define the final clinical conduct. The second scenario consists of using the system to build clinical analysis protocols, in case of absence of the automatic analysis solution. In this case, the physician can generate a protocol document. In this way, the document can guide the request for blood tests, optimizing the time of collection and results. Finally, the medical team will be able to define the clinical management. Source: Authors

We used 6215 records of patients seen in outpatient clinics, emergency rooms and the emergency department of SUS in the city in question, from August 30, 2019 to August 17, 2020. Of these records, 57.61% were women (3581) and 0.27% were newborns (17) and did not have their gender identified. The mean age was 41.79 (22.94) years. Among men, 4.54% tested positive for COVID-19 and proportionally the most affected age group was over 90 years, where 20% of patients were victims of the disease. Among the women, 3.88% had the diagnosis of COVID-19 confirmed, with 11.86% of the patients aged 50 to 60 years and 11.88% of the age group 80 to 90 years tested positive.

The graph in Fig. 2 shows the demographic stratification by sex and age of all records used in this study. We highlight the positive cases of COVID-19 from male (COVID-19 M) and female (COVID-19 F) patients using

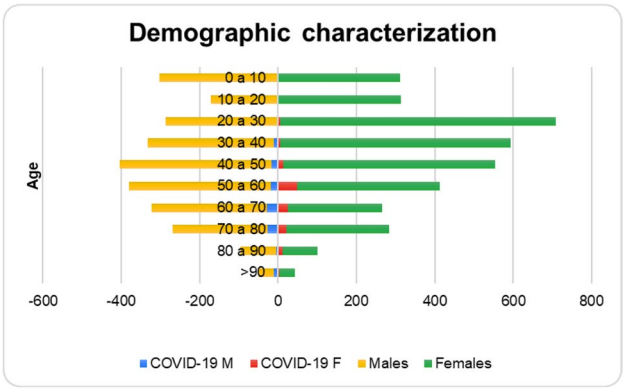

Fig. 2 Demographic stratification by sex and age of all records used in this study. We highlight the positive cases of COVID-19 from male (COVID-19 M) and female (COVID-19 F) patients using the colors blue and red, whilst the total of males and females patients are associated to orange and green, respectively. Source: Authors

the colors blue and red, whilst the total of male and female patients are associated to orange and green, respectively. Table 1 shows the demographic details of the database, in which 2617 records are related to males, whilst 3581 are females; from these records, 119 (4.54%) are males positive for COVID-19 and 139 (3.88%) are females with COVID-19 as well, in a total of 258 COVID-19 patients distributed among moderate and severe COVID-19 cases.

Since not all 6215 records correspond to medical records with tests that cover all the hematological parameters provided for in this work, the database has several missing values. To fill in these missing values, we used mean predictors. For each class (positive or negative for COVID-19), the sample mean for each attribute was calculated. This mean was used as a predictor for the missing values in the attribute with missing value in the vector of attributes of the corresponding class. The total of medical records with blood tests considered incomplete does not exceed 30% of the data set.

To perform this research, we initially selected 41 hematological parameters. These hematological parameters correspond to blood exams recommended by the Ministry of Health of Brazil as an initial clinical approach and part of the COVID-19 diagnostic process (Brazilian Ministry of Health Guidelines for the diagnosis and treatment of COVID-19 2020). Therefore, considering that health centers must already perform these tests, there is no monetary loss or time spent on additional tests. On the contrary, the diagnostic process can be optimized with the system proposed here. The list of 41 hematological parameters is shown in Fig. 3. The Complete Blood Count (CBC) with differential comprises 20 of these hematological parameters, while arterial blood gas analysis includes 9 hematological parameters. The remaining 12 exams are those of total, indirect, and direct bilirubin; serum glucose; lipase dosage; urea; D-dimer; lactic dehydrogenase; C-reactive protein (CRP); creatinine; and partial thromboplastin time (PTT); and prothrombin time activity from coagulogram. To this 41-item list, we added patient's age and a redundant lymphocytes parameter. This additional lymphocyte item corresponds to a different electronic patient record specification that used to categorize atypical lymphocytes but was abandoned since the beginning of 2020.

Feature selection

The database constructed during the pre-processing was submitted to feature selection using Particle Swarm Optimization (PSO) (Kennedy and Eberhart 1995; Poli et al. 2007; Wang et al. 2007). As objective function, we employed a simple decision tree to guide the optimization process. We set the algorithm to 20 individuals and 500 iterations. The goal of the attribute selection is to find the most significant exams for classification tasks and to reduce the number of required exams for diagnostic support.

We chose the PSO algorithm because it is a well-established search and optimization algorithm with few optimization parameters to be defined (Kennedy and Eberhart 1995;

Table 1 Demographic details of the database: distribution of the number of patients by gender and by diagnosis of COVID-19, with a percentage of the population with COVID-19 by gender and by age group. Source: Authors

Age (years)	Males (total)	COVID-19 males	Females (total)	COVID-19 females
0–10	302	0	311	1 (0.32%)
10–20	171	0	313	2 (0.63%)
20–30	287	2 (0.69%)	709	5 (0.7%)
30–40	332	11 (3.31%)	594	6 (1.01%)
40–50	405	17 (4.19%)	549	14 (2.55%)
50–60	380	19 (5%)	413	49 (11.86%)
60–70	323	29 (8.97%)	265	26 (9.81%)
70–80	270	26 (9.62%)	284	21 (7.39%)
80–90	97	5 (5.15%)	101	12 (11.88%)
> 90	50	10 (20%)	42	3 (7.14%)
Total	2617	119 (4.54%)	3581	139 (3.88%)

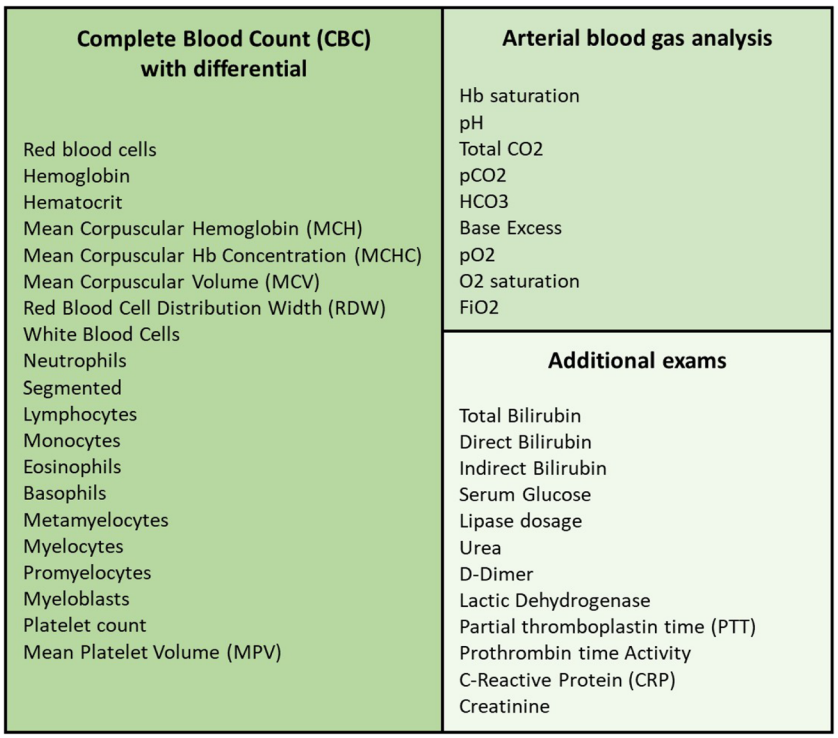

Fig. 3 The list of tests specified by the Ministry of Health of Brazil and used both in the initial assessment of symptomatic patients suspected of COVID-19 and for the control and assessment of disease progress. Source: Authors

Poli et al. 2007; Wang et al. 2007). The PSO algorithm was created taking inspiration from the behavior of flocks of birds (Kennedy and Eberhart 1995; Poli et al. 2007; Wang et al. 2007). These flocks of birds were modeled as dynamic systems where there is a global leader, who guides the flock in a direction that optimizes a given measure of performance, and local leaders. While the global leader governs the overall dynamics of the flock of birds, local leaders are defined as those who perform best among their neighbors. Each bird is modeled using pairs of spatial position and velocity vectors. Throughout the evolution of the system, it is possible for a local leader to become a global leader, as it is also possible for leaders to become simple members of the flock of birds. Each bird represents a candidate for solving the problem of maximizing or minimizing a given objective function. In the case of maximization problems, the candidates for the solution that correspond to the global maximum and local minimums are the respective global and local leaders. Figure 4 illustrates the metaphor that inspires the PSO algorithm. PSO algorithm uses a population of randomly generated particles. In this approach, each particle corresponds to randomly generated solution and have an associated velocity and position. For each particle, each position vector is binary: 1's and 0's corresponds to the presence or the absence of one of the 43 features in the process of training and testing the decision tree classifier associated to the objective function. The output of the objective function is the overall accuracy of a tenfold cross validation training process. In this objective function, a decision tree is used as classifier. We employed meta-heuristic libraries developed in Java for Weka data mining platform (Gnanambal et al. 2018). We adopted the following feature selection methods: individual weight of 0.34, inertia weight of 0.33, mutation probability of 0.01, report frequency of 20, social weight of 0.33 (Bratton and Kennedy 2007, Kennedy and Eberhart 1995, Poli et al. 2007, Van den Bergh and Engelbrecht 2004).

The feature selection implementation resulted two databases: the original database with 43 attributes, SARS-CoV-2, and the dimension-reduced database, SARS-CoV-2 (PSO). Age was not selected: it appeared as not statistically relevant according to PSO selection. The selected exams/features were the following: serum glucose, indirect bilirubin, partial thromboplastin time, lactic dehydrogenase, lipase dosage, D-dimer, ferritin, and troponin.

Classification

Multilayer perceptron

Multilayer perceptron (MLP) consists of a generalization of the Perceptron proposed by Franklin Rosenblatt in 1958. Perceptron is the model is the simplest form of a neural network, being able to deal with linearly separable problems. Multilayer perceptron networks, on the other hand, have several interconnected neurons (or nodes), arranged in layers: the input layer, the hidden layers, and the output layer. The input layer only has the network input vector, which is

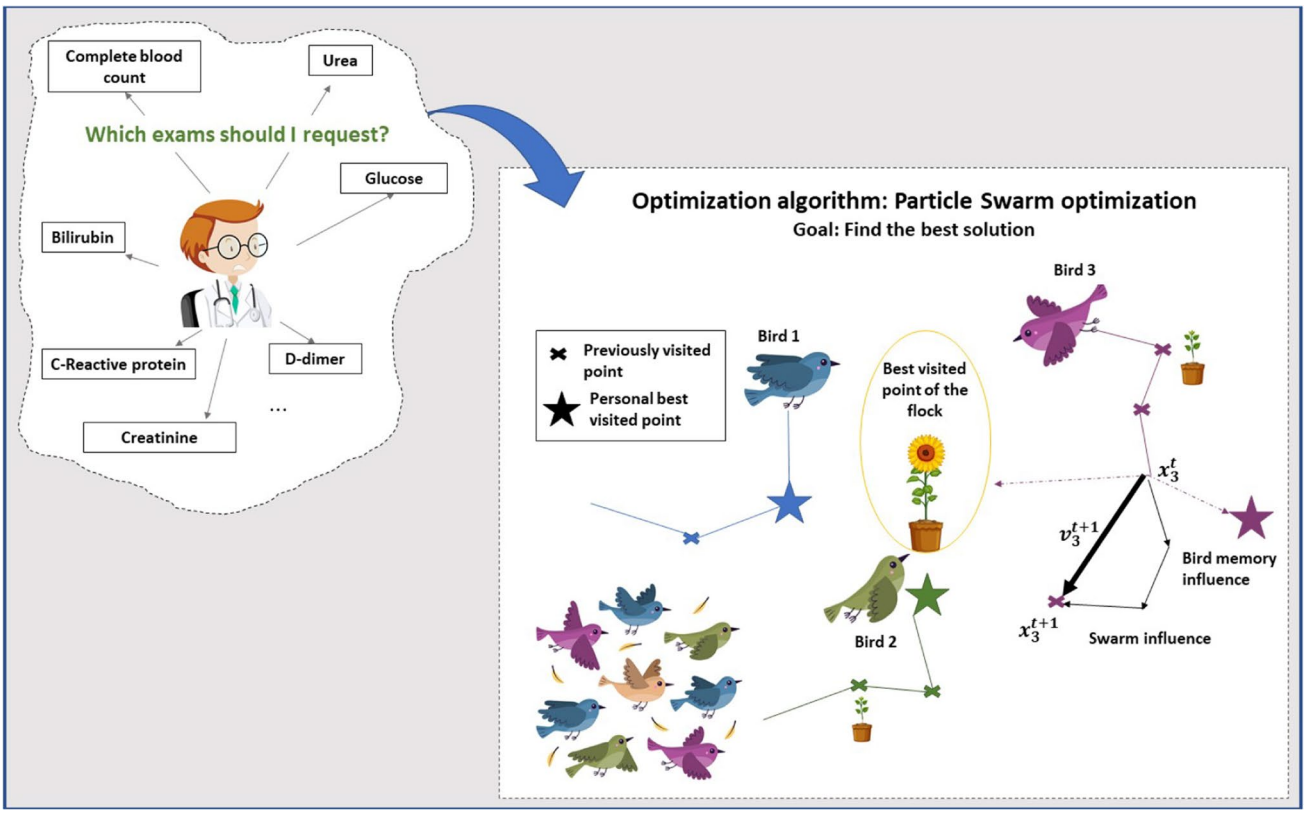

Fig. 4 The particle swarm optimization (PSO) algorithm is based on the behavior of flocks of birds. Each bird is modeled by a position vector and a velocity vector. Thus, the movement of a bird in search of resources is governed by a global leader and local leaders whose performance is defined by an objective function. In the problem of selection of the most significant exams, we use a decision tree as an objective function, with the overall accuracy of the training and testing process by tenfold cross-validation being returned as output. Each bird is a candidate for solving the problem of maximizing accuracy. Position vectors are binary, where each coordinate corresponds to the presence or absence of one of the 43 attributes. Source: Authors

passed on to the next layer. Then, each node in the next layer modifies these input values through non-linear activation functions, generating output signals. In addition, the network nodes are connected by weights, which scales these output signals. Finally, the superposition of several non-linear functions allows the mapping of the input vector to the output vector. As MLPs can have one or multiple hidden layers, this process can be repeated several times, depending on the selected architecture (Barbosa et al. 2020; Gardner and Dorling 1998; Lerner et al. 1994; Phung et al. 2005).

Through the proper selection of activation functions and synaptic weights, an MLP is able to approximate the inputs at the desired outputs. This search and adjustment of parameters is called the training process. MLPs learn in a supervised manner. During this process, errors between the actual and desired outputs are calculated. These errors are used to adjust the network (Gardner and Dorling 1998).

In order to adjust these weights, the backpropagation algorithm is the most computationally straightforward and common algorithm. It occurs in two phases: the forward and backward propagation. In the first step, the initial network weights are set to small random values. Then, this first input vector is propagated through the network to obtain an output. This actual output is compared with the desired one, and the error is calculated. In the second phase, the backward propagation, the error signal is propagated back through the network and the connection weights are updated, aiming to minimize the overall error. These steps can be repeated until the overall error is satisfactory (Haykin 2001).

MLPs and other artificial neural networks architectures are commonly used in support diagnosis applications (Naraei et al. 2016), e.g., liver disease diagnosis (Abdar et al. 2018), heart diasese diagnosis (Hasan et al. 2017), breast cancer diagnosis over breast thermography (de Vasconcelos et al. 2018; Pereira et al. 2020a,b,c; Rodrigues et al. 2019; Santana et al. 2020, 2018) and mammography images (Cordeiro et al. 2016, 2017, Cruz et al. 2018, de Lima et al. 2014, 2016, Lima et al. 2015, Silva and Santana 2020), for recognition of intracranial epileptic seizures (Raghu and Sriraam 2017), and multiple sclerosis diagnosis support (Commowick et al. 2018).

Table 2 Kernel functions of SVM. Source: Authors

Type of SVM	Kernel function
Polynomial	$K(x, y) = (x \bullet y + 1)^E$
Radila Basis Function (RBF)	$K(x, y) = \exp(-\gamma(x - y) \bullet (x - y))$

Support vector machines

Support vector machines (SVM) were created by Vladimir Vapnik and Alexey Chervonenkis (Boser et al. 1992; Cortes and Vapnik 1995) in 1963. Their main purpose is to build a linear decision surface, called a hyperplane. The idea is that this hyperplane should be able to separate classes in the best possible way. The optimal hyperplane is found when the margin of separation between it and a given nearest point is maximum (Haykin 2001).

SVM classifier are known for its good generalization performance. They are employed in several healthcare applications, such as breast cancer diagnosis using thermography and mammography (Cordeiro et al. 2016, 2017, Cruz et al. 2018, de Lima et al. 2014, 2016, de Vasconcelos et al. 2018, Lima et al. 2015, Pereira et al. 2020b, Santana et al. 2020, Silva and Santana 2020), diabetes mellitus diagnosis (Barakat et al. 2010), heart valve diseases (Çomak et al. 2007) and pulmonary infections detection (Yao et al. 2011), and also diagnosis of pulmonary cancer (Sun et al. 2013). However, its performance varies depending on the problems complexity. The type of the machine varies with the type of kernel used to build the optimal hyperplane. Table 2 shows the kernel functions used in this study: the polynomial and RBF kernels. For the first case, it was tested exponents of value 1 (linear kernel), 2, and 3.

Decision trees

Decision trees are sequential models, which combine several simple tests. They can be understood as a series of questions with "yes" and "no" answers. These tests can be the comparison of a value with a threshold or a categorical attribute compared to a set of possibilities, for instance. Thus, when analyzing the data with these tests, the decision trees will guide to a certain class in classification problems, or to a continuous value, in cases of regression problems. In this way, a decision tree is built with certain questions, called nodes. Essentially, there are four types of nodes: root, parent, child, and leaf. Starting at the root node, an instance is classified. Then, the outcome for this instance is determined ad the process continues through the tree. In addition, one node may connect to another, establishing a parent–child relationship, in which a parent node generates a child node. Finally, the terminal nodes of the tree are the leaf nodes, and they represent the final decision, that is, the predicted class or value. There are several types of decision trees, depending on the tree structure. The most popular ones are random tree and random forest. Both of them were tested in this study by using multiple configuration parameters (Kotsiantis 2013; Podgorelec et al. 2002).

Random tree uses a tree built by a stochastic process. This method considers only a few randomly selected features in each node of the tree Geurts et al. (2006).

In contrast, random forest is a model made up of many decision trees. In this case, a set of trees is built and their votes are combined to classify an instance, by means of the majority vote. Each decision tree uses a subset of attributes randomly selected from the original set of attributes (Breiman, 2001).

Bayesian network and naive Bayes classifier

Bayesian classifiers are based on Bayes' decision theory. Among the most popular Bayes' classifiers are naive Bayes and Bayes Nets, also known as Bayesian networks. Bayesian networks describe the probability distribution over a set of variables. They represent, in a simple way, the causal relationships of the variables of a system using Graph Theory, where the variables are the nodes, and the arcs identify the relationships between the variables. In the learning process, it is necessary to calculate the probability distributions and to identify the network structure. Learning the network structure can be considered an optimization problem, where the quality measure of a network structure needs to be maximized (Bouckaert 2008; Cheng and Greiner 1999).

On the other hand, the Naive Bayes classifier is a simple model that considers that the domain variables are conditionally independent, that is, one characteristic is not related to the other. Its learning is done in an inductive way, presenting a set of training data and calculating the conditional probability of each attribute, given a class. Naive Bayes needs to estimate few configuration parameters (Bouckaert 2008; Cheng and Greiner 2001).

Parameters settings of the classifiers

All experiments were performed using the Weka Java library in 30 runs. During training, we used the k-fold cross-validation method with $k = 10$ to split the set (Jung and Hu 2015). Therefore, in each run the training was performed 10 times, with 9 parts of the set being used for training and 1 part used for validation. It is worth mentioning that this method guarantees that the validation set does not participate in the training. Thus, we were able to verify the performance of the model against external training data and tenfold cross validation. The experiments were made by using the following methods:

- Naive Bayes classifier;
- Bayes net;

- Multilayer perceptron: one hidden layer, for 20, 50, and 100 hidden neurons;
- Support vector machines: we tested the following kernels, for the configuration parameter C varying for 0.01, 0.1, 1, and 10: linear kernel, polynomial kernel with degree varying for 2, 3, 4, and 5; and Radial Basis Function (RBF) kernel, with γ of 0.01, 0.25 and 0.5;
- Decision trees: J48 and random tree; random forests, for 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 trees.

Metrics

We chose the following metrics to evaluate the performance of diagnostic tests: accuracy, sensitivity, specificity, and the area under ROC curve. Accuracy is the probability that the test will provide correct results, that is, be positive in sick patients and negative in healthy patients. In other words, it is the probability of the true positives and true negatives among all the results. The sensitivity is the rate of true positives and indicates the classifier ability to detect correctly people with COVID-19. Specificity is the capacity of classifying healthy patients as negatives. It is the rate of true negatives. Finally, the area under the ROC curve is a measure of a classifier's discriminating ability. That is, given two classes — a sick individual and a non-sick individual—chosen at random, the area below the ROC curve that indicates a probability of the latter being correctly classified. If the classifier cannot discriminate between these two separately, an area under a curve is equal to 0.5. When this value is the next 1, it indicates that the classifier is able to discriminate these two cases (Hand 2009).

These metrics allow to discriminate between the target condition and health, in addition to quantifying the diagnostic exactitude (Borges 2016). The accuracy, sensitivity, and specificity can be calculated as following:

$$\text{Accuracy} = \rho_0 = \frac{TP + TN}{TP + TN + FP + FN'} \tag{1}$$

$$\text{Sensitivity} = TPR = \frac{TP}{TP + FN'} \tag{2}$$

$$\text{Specificity} = TNR = \frac{TN}{TN + FP'} \tag{3}$$

where TP is the true positives, TN is the true negatives, FP is the false positives, and FN, the false negatives. TPR and TNR are the true positive and the true negative rates, respectively (Fawcett 2006).

An area under the ROC curve (AUC) is a measure of a classifier's discriminative ability. The ROC curve is produced by calculating and plotting the true positive rate (TPR) against the false-positive rate (FPR) for a single classifier at a variety of thresholds. Given two classes chosen at random, the area below the ROC curve indicates a probability of the latter being correctly classified. If the classifier cannot discriminate between these two separately, an area under a curve is equal to 0.5. When this value is the next 1, it indicates that the classifier is able to discriminate these two cases. Mathematically (Fawcett 2006):

$$\begin{aligned} \text{AUC} &= \int_0^1 TPR dFPR \approx \frac{1}{2}(1 - FPR + TPR) \\ &= \frac{\text{Sensitivity} + \text{Specificity}}{2} \end{aligned} \tag{4}$$

The classification methods used in the COVID-19 detection task considering vectors of hematological parameters were compared using the following metrics: accuracy, sensitivity, specificity, and area under the ROC curve. Metrics are compared using the sample average and standard error (format: average (standard error)) and boxplots. Two classification methods are considered equivalent according to a given metric if their box plots match, i.e., similarity between median, lower and upper quartiles, and outliers. By means of the qualitative and quantitative comparison of the box plots, it is not necessary to assume a certain probability distribution and, therefore, it is not necessary to use tests of normality or tests of hypotheses that assume normality of the data. The classifiers are first compared considering the most clinically relevant criteria: sensitivity, specificity and area under the ROC curve. Then, they are compared considering box plots of sensitivity and specificity. Furthermore, for the final decision, we selected two of them with best performances according to boxplots and mean values, and performed the paired t-test. We computed p-values and considered a p-value less than or to 0.05 as significant.

Protocol proposals

In this subsection, we present protocol proposals to support the clinical diagnosis of COVID-19 from the hematological parameters obtained clinically from the complete blood count and biochemical tests. Clinical diagnosis support protocols are composed of empirical rules applied to clinical examinations. Our hypothesis is that decision trees, as they are techniques to support decision-making based on rules, can be used both to explain the functioning of intelligent systems to support clinical diagnosis, and to automatically build protocols for clinical diagnosis.

In decision trees, it is natural that one or more attributes dominate the decision process, creating dominant branches in the trees. Seeking to detail the role of each attribute, that is, of each hematological parameter in the construction of the clinical diagnosis, we subsequently decided to build new decision trees by removing the dominant attribute from the

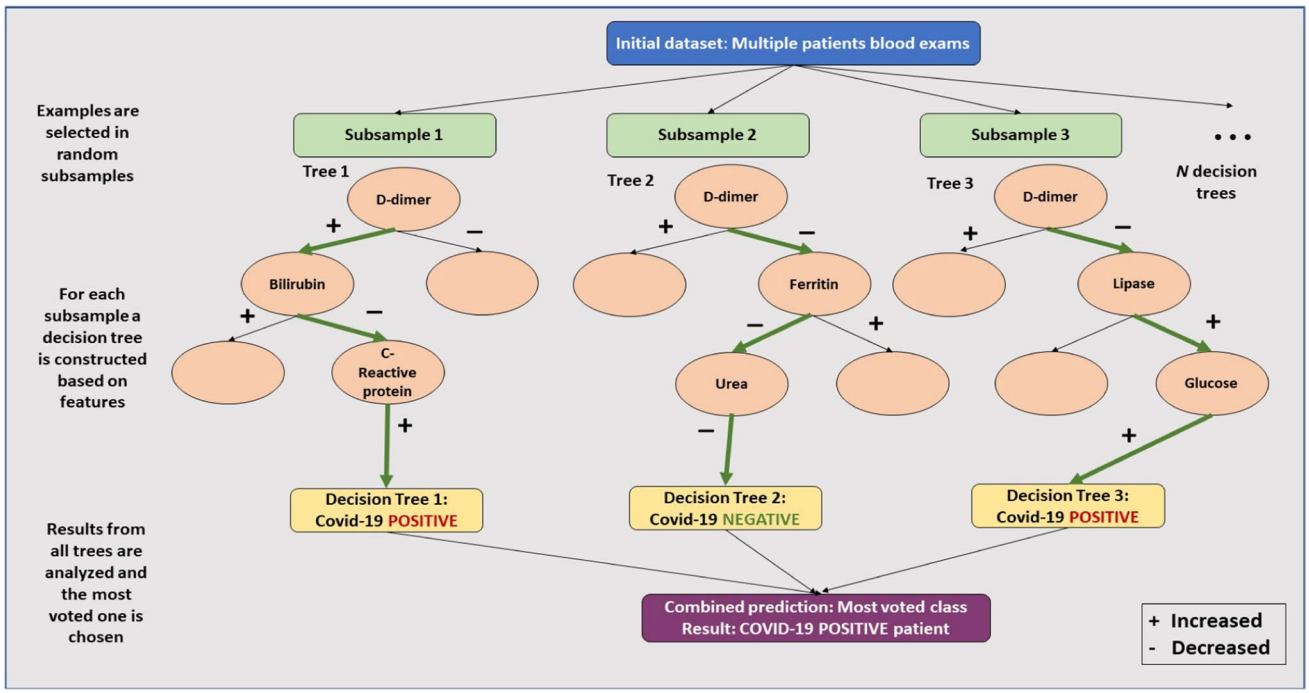

Fig. 5 Random Forest classifier architecture adapted to COVID-19 rapid diagnosis as implemented in Heg.IA web proposal. Source: Authors

previous tree. The process of building new decision trees is interrupted when the diagnostic accuracy drops below an acceptable value. In this way, each decision tree expresses a rapid clinical decision protocol where a hematological parameter is dominant.

Heg.IA web application

After selecting the best classifier, the Heg.IA web system was developed. It can be accessed through the link: https://hegia.ufpe.br/welcome. Its front-end was developed using the library React.js. This library is based on pure JavaScript. It is open source and used to create user interfaces, more specifically, single page application (SPA) web platforms. As for data access and manipulation of application state, we used the Redux-Saga structure, a powerful tool that allows us to manage masterfully asynchronous queries, receiving API data, and trigger actions to the application of state safely and easily to maintain. Furthermore, our back-end was developed in Python (version 3.7.7). Only the random forest classifier was implemented in this final solution. Figure 5 illustrates how Random Forest was be applied as a classifier for the COVID-19 rapid diagnosis proposal we suggest in this work.

On the initial screen, it is possible to visualize a brief description of the intelligent system, as well as the supporters of this initiative: The Federal University of Pernambuco (UFPE) and the Department of Biomedical Engineering at UFPE. To get to know the members of the project's development team and their respective functions, it is possible to access the "About" option on the top menu of the screen. The options "Login" and "Consult" are also available. For the "Login" option, health professionals, especially medical laboratory professionals and nurses, will be able to access their private account or register a new account, in cases of first access. In the Consult option, it is possible to view the report with the diagnosis for a specific patient, as long as the user has the patient's personal locator.

After logging into the system, the user can register a patient or view the complete history of registered patients. In the case of a new registration, personal information such as full name, ID, date of birth, telephone, sex, and full home address will be requested. In the following, the user will be directed to the screen shown in Fig. 6. In this screen, the results of the complete blood count (CBC) with differential must be entered. The units and reference values are available next to each of the hematological parameters. After filling in the CBC, the user will be directed to the screens for the other blood tests and arterial blood gas tests, as shown in Figs. 7 and 8. Thus, the list of tests required to make the predictions will be complete. The user can then check the hematological parameters entered to be sure that everything is correct. If he realizes that he made a typo, he can go back to the previous steps and correct it.

Finally, the report will be available immediately, similarly to that shown in Fig. 9. The diagnostic report will indicate the positive or negative diagnosis for COVID-19. Hospitalization

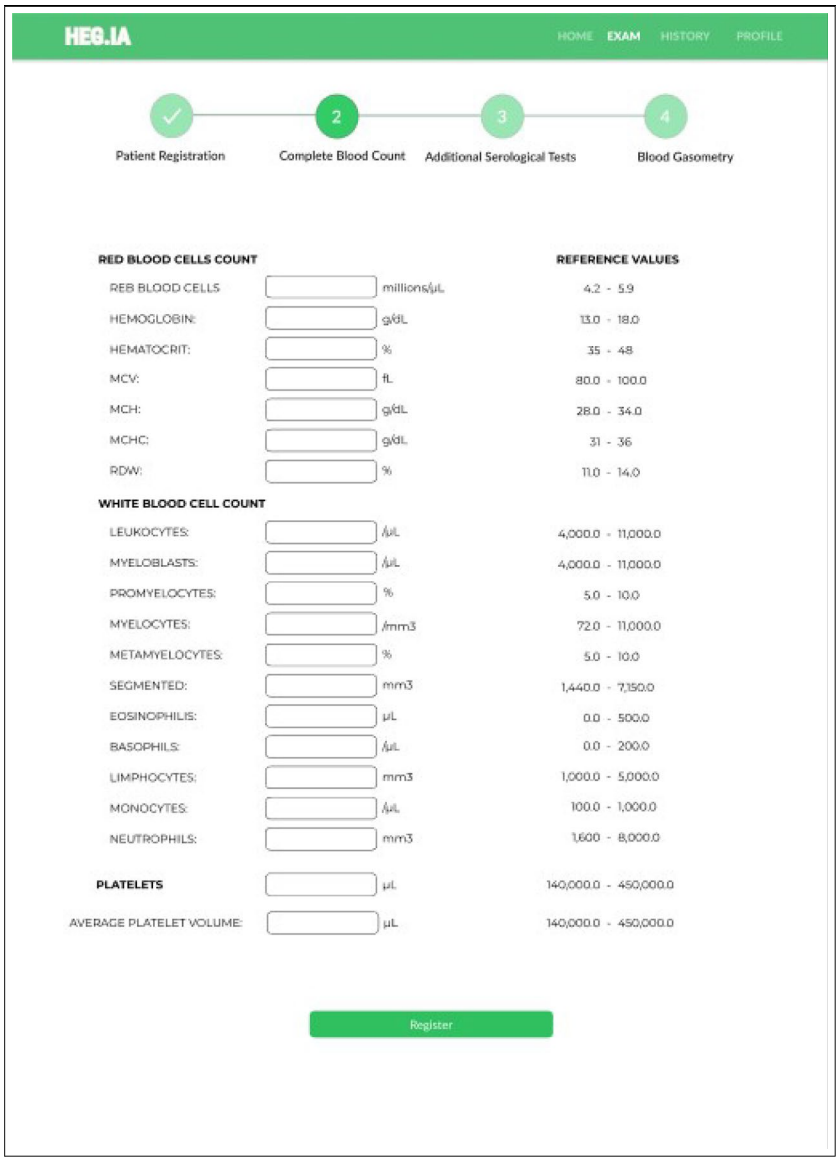

Fig. 6 Complete blood count screen: after the patient's registration, the results of the patient's complete blood count with differential can be inserted. The units and reference values can be viewed next to each hematological parameter. Source: Authors

predictions are also reported, indicating the best type of hospitalization for the patient: regular ward, semi-ICU or ICU. Information on accuracy, kappa index, sensitivity, and specificity of the determination of each of these scenarios are also available, in order to assist the physician's decision making. In addition to viewing the report, it is also possible to print it.

Results

The results of this research are organized in three parts: in the first part, we present the investigation of the best classification architecture for the original dataset and for the version with the reduced number of attributes through feature selection based on the PSO algorithm. The most suitable classifier was used in the implementation of the COVID-19 diagnostic support web system, Heg.IA web. In the second part, we present the use of decision trees as an alternative to build humanly intelligible models to support clinical diagnosis. These models, although less sophisticated, are important for the clinician to have a better understanding of the hematological parameters that are not only more prevalent in the diagnosis, but which can also be important for monitoring the clinical status of patients with COVID-19. Finally, in the third part, we present the prototype of the Heg.IA web system, in operation in the city of Paudalho, Brazil, since June 2020.

Evaluation of classifiers to support the diagnosis of COVID-19

We investigated the best classifier architectures for classifying patterns of hematological parameters. We investigate

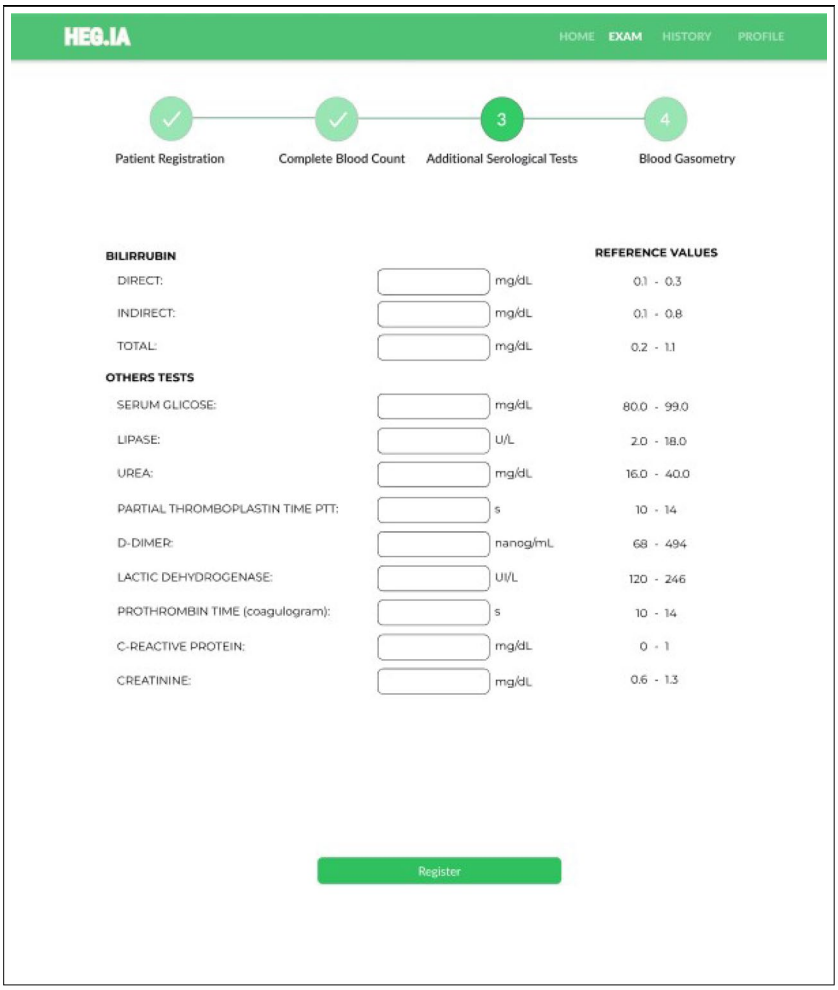

Fig. 7 Additional serological tests screen: additional tests, such as total, direct, and indirect bilirubin, can be inserted in this screen. In addition, serum glicose, dosage lipase, urea, PTT, D-dimer, lactic dehydrogenase, prothrombin time, CRP, and creatinine results can also be included here. The units and reference values can be viewed next to each hematological parameter. Source: Authors

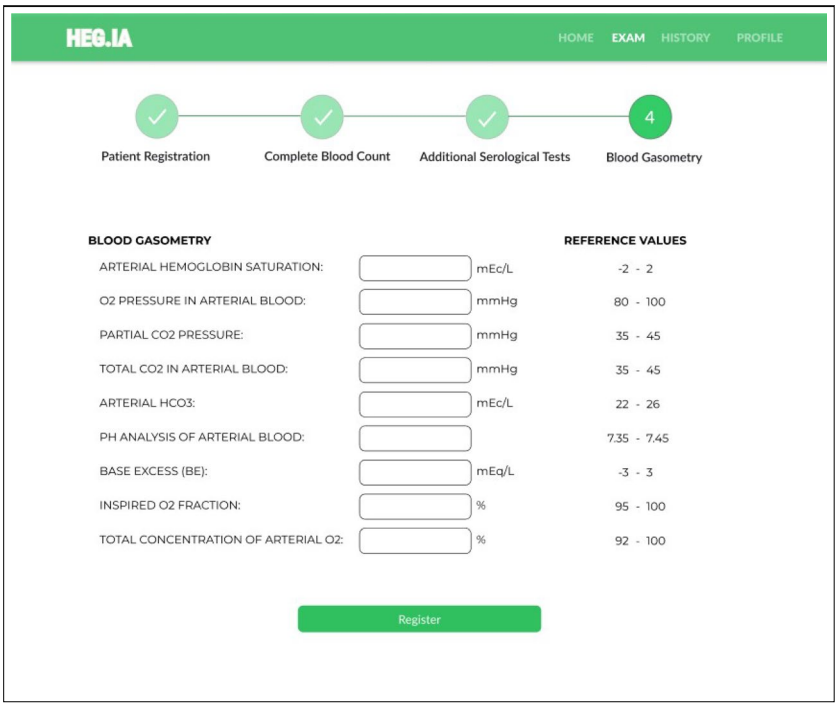

Fig. 8 Blood gasometry screen: arterial gasometry can be inserted in this screen, finalizing the list of necessary exams. The units and reference values for each hematological parameter are available. Source: Authors

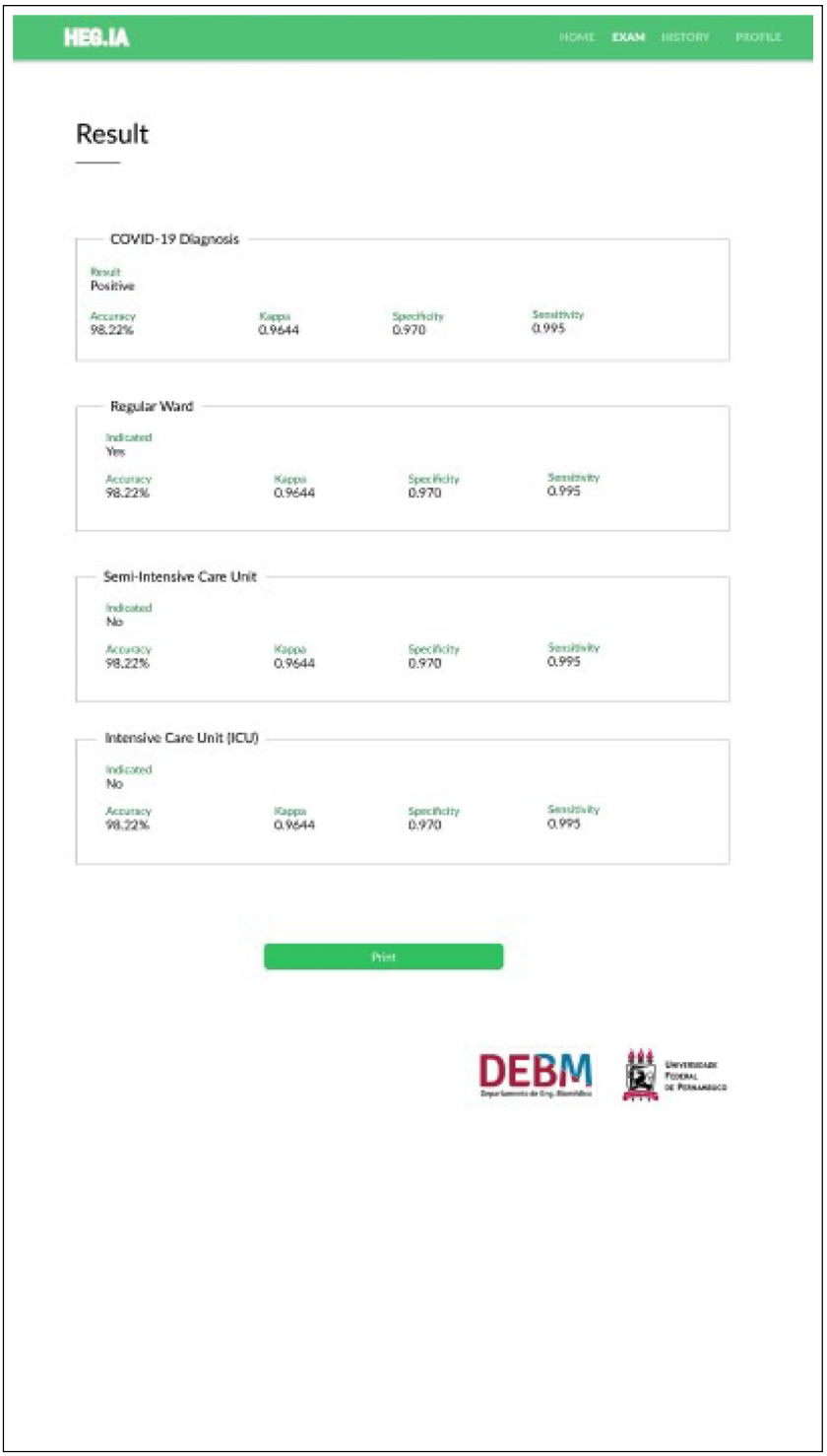

Fig. 9 Results screen: In this screen it is possible to view the patient's diagnostic report. In the report, the diagnosis for COVID-19 is available, as well as the hospitalization predictions, indicating whether the patient should be admitted to the regular ward, semi-intensive care unit, or to the ICU. Information on accuracy, kappa index, sensitivity, and specificity of the determination of each of these scenarios are also available, in order to assist the physician's decision making. In addition to viewing the report, it is also possible to print it. Source: Authors

multilayer perceptron artificial neural networks, support vector machines, Bayesian networks, Naive Bayes classifiers, simple decision trees, and random forests. We studied the behavior of these classification architectures for the 6215 patient records, 258 of which were positive for COVID-19, mostly moderate and severe cases of COVID-19.

Table 3 shows the average and standard error for accuracy, sensitivity, specificity, and area under the ROC curve, taking into account the original dataset, with all 43 attributes. Looking at Table 3, we can see that Bayes Net's performance is optimal (nearly 1.0000 or 100.00%) for all metrics, with standard error less than 0.0001. The results for Naive Bayes

Table 3 Sample mean and standard error for accuracy, sensitivity, specificity, and area under ROC curve (AUC) for all classifiers, considering the original 43-feature dataset. Source: Authors

Classifier	SARS-CoV-2											
	Accuracy		Kappa		Sensitivity		Specificity		AUC		F1 score	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Naïve Bayes	**99.8890**	**0.2149**	**0.9868**	**0.0230**	**0.9989**	**0.0022**	**0.9997**	**0.0031**	**0.9999**	**0.0011**	**0.9994**	**0.0011**
Bayes network	**99.9716**	**0.0629**	**0.9965**	**0.0077**	**0.9997**	**0.0007**	**1.0000**	**0.0000**	**1.0000**	**0.0000**	**0.9999**	**0.0003**
J48 tree	**99.9839**	**0.0483**	**0.9980**	**0.0059**	**0.9998**	**0.0005**	**1.0000**	**0.0000**	**0.9999**	**0.0003**	**0.9999**	**0.0003**
Random tree	**99.8220**	**0.2005**	**0.9775**	**0.0251**	**0.9992**	**0.0014**	**0.9755**	**0.0332**	**0.9873**	**0.0167**	**0.9991**	**0.0010**
Random forest (10 trees)	99.9619	0.0780	0.9951	0.0101	1.0000	0.0002	0.9913	0.0185	1.0000	0.0000	0.9998	0.0004
Random forest (20 trees)	99.9850	0.0469	0.9981	0.0060	1.0000	0.0000	0.9964	0.0113	1.0000	0.0000	0.9999	0.0002
Random forest (30 trees)	99.9882	0.0420	0.9985	0.0054	1.0000	0.0000	0.9972	0.0101	1.0000	0.0000	0.9999	0.0002
Random forest (40 trees)	99.9925	0.0340	0.9990	0.0043	1.0000	0.0000	0.9982	0.0082	1.0000	0.0000	1.0000	0.0002
Random forest (50 trees)	99.9925	0.0340	0.9990	0.0043	1.0000	0.0000	0.9982	0.0082	1.0000	0.0000	1.0000	0.0002
Random forest (60 trees)	99.9941	0.0303	0.9992	0.0039	1.0000	0.0000	0.9986	0.0073	1.0000	0.0000	1.0000	0.0002
Random forest (70 trees)	99.9946	0.0289	0.9993	0.0037	1.0000	0.0000	0.9987	0.0070	1.0000	0.0000	1.0000	0.0002
Random forest (80 trees)	99.9941	0.0303	0.9992	0.0039	1.0000	0.0000	0.9986	0.0073	1.0000	0.0000	1.0000	0.0002
Random forest (90 trees)	99.9946	0.0289	0.9993	0.0037	1.0000	0.0000	0.9987	0.0070	1.0000	0.0000	1.0000	0.0002
Random forest (100 trees)	**99.9930**	**0.0328**	**0.9991**	**0.0042**	**1.0000**	**0.0000**	**0.9983**	**0.0079**	**1.0000**	**0.0000**	**1.0000**	**0.0002**
MLP (20 neurons)	**99.9415**	**0.0870**	**0.9928**	**0.0108**	**0.9995**	**0.0008**	**0.9973**	**0.0099**	**0.9997**	**0.0007**	**0.9997**	**0.0005**
MLP (50 neurons)	99.9372	0.0909	0.9922	0.0113	0.9995	0.0008	0.9963	0.0114	0.9997	0.0007	0.9997	0.0005
MLP (100 neurons)	99.9372	0.0889	0.9922	0.0110	0.9995	0.0008	0.9961	0.0116	0.9997	0.0007	0.9997	0.0005
SVM polynomial E1; C = 0.01	96.3368	0.5989	0.1894	0.1581	0.9998	0.0005	0.1210	0.1454	0.5604	0.0726	0.9813	0.0031
SVM polynomial E2; C = 0.01	99.9560	0.0776	0.9945	0.0096	0.9997	0.0007	0.9973	0.0099	0.9985	0.0050	0.9998	0.0004
SVM polynomial E3; C = 0.01	99.9394	0.0951	0.9925	0.0118	0.9995	0.0008	0.9961	0.0116	0.9978	0.0058	0.9997	0.0005
SVM polynomial E4; C = 0.01	99.9512	0.0870	0.9939	0.0108	0.9997	0.0007	0.9961	0.0116	0.9979	0.0058	0.9997	0.0005
SVM polynomial E5; C = 0.01	99.9517	0.0878	0.9940	0.0109	0.9997	0.0007	0.9957	0.0129	0.9977	0.0065	0.9997	0.0005
SVM RBF G = 0.01; C = 0.01	95.8488	0.0629	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.5000	0.0000	0.9788	0.0003
SVM RBF G = 0.25; C = 0.01	95.8488	0.0629	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.5000	0.0000	0.9788	0.0003
SVM RBF G = 0.5; C = 0.01	95.9013	0.4584	0.0130	0.1120	1.0000	0.0001	0.0129	0.1115	0.5065	0.0557	0.9791	0.0023
SVM polynomial E1; C = 0.1	99.9705	0.0637	0.9964	0.0078	0.9997	0.0007	1.0000	0.0000	0.9998	0.0003	0.9998	0.0003
SVM polynomial E2; C = 0.1	99.9657	0.0723	0.9958	0.0088	0.9996	0.0008	1.0000	0.0000	0.9998	0.0004	0.9998	0.0004
SVM polynomial E3; cC = 0.1	99.9469	0.0875	0.9934	0.0109	0.9996	0.0007	0.9961	0.0116	0.9979	0.0058	0.9997	0.0005
SVM polynomial E4; C = 0.1	99.9474	0.0940	0.9935	0.0116	0.9996	0.0008	0.9961	0.0116	0.9979	0.0059	0.9997	0.0005
SVM polynomial E5; C = 0.1	99.9383	0.1065	0.9923	0.0132	0.9996	0.0009	0.9955	0.0124	0.9975	0.0063	0.9997	0.0006
SVM RBF G = 0.01; C = 0.1	95.8488	0.0629	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.5000	0.0000	0.9788	0.0003
SVM RBF G = 0.25; C = 0.1	**99.9839**	**0.0483**	**0.9980**	**0.0059**	**0.9998**	**0.0005**	**1.0000**	**0.0000**	**0.9999**	**0.0003**	**0.9999**	**0.0003**
SVM RBF G = 0.5; C = 0.1	99.9362	0.0975	0.9919	0.0124	0.9998	0.0005	0.9885	0.0199	0.9942	0.0099	0.9997	0.0005
SVM polynomial E1; C = 1	**99.9796**	**0.0552**	**0.9975**	**0.0068**	**0.9998**	**0.0006**	**1.0000**	**0.0000**	**0.9999**	**0.0003**	**0.9999**	**0.0003**
SVM polynomial E2; C = 1	99.9635	0.0712	0.9955	0.0087	0.9996	0.0007	1.0000	0.0000	0.9998	0.0004	0.9998	0.0004
SVM polynomial E3; C = 1	99.9480	0.0871	0.9935	0.0108	0.9996	0.0008	0.9968	0.0107	0.9982	0.0054	0.9997	0.0005
SVM polynomial E4; C = 1	99.9271	0.1062	0.9909	0.0132	0.9995	0.0009	0.9932	0.0151	0.9963	0.0076	0.9996	0.0006
SVM polynomial E5; C = 1	99.9244	0.1096	0.9906	0.0136	0.9995	0.0009	0.9932	0.0154	0.9963	0.0078	0.9996	0.0006
SVM RBF G = 0.01; C = 1	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	0.9999	0.0003	0.9999	0.0003
SVM RBF G = 0.25; C = 1	99.9689	0.0663	0.9961	0.0083	0.9998	0.0005	0.9964	0.0113	0.9981	0.0056	0.9998	0.0003
SVM RBF G = 0.5; C = 1	99.9512	0.0860	0.9938	0.0109	0.9998	0.0005	0.9921	0.0165	0.9960	0.0083	0.9997	0.0004
SVM polynomial E1; C = 10	99.9421	0.0945	0.9929	0.0115	0.9994	0.0010	1.0000	0.0000	0.9997	0.0005	0.9997	0.0005
SVM polynomial E2; C = 10	99.9426	0.0971	0.9930	0.0119	0.9994	0.0010	0.9996	0.0038	0.9995	0.0021	0.9997	0.0005
SVM polynomial E3; C = 10	99.9018	0.1145	0.9877	0.0143	0.9995	0.0009	0.9890	0.0183	0.9942	0.0092	0.9995	0.0006
SVM polynomial E4; C = 10	99.9163	0.1097	0.9895	0.0137	0.9995	0.0009	0.9915	0.0164	0.9955	0.0082	0.9996	0.0006
SVM polynomial E5; C = 10	99.9244	0.1096	0.9906	0.0136	0.9995	0.0009	0.9932	0.0154	0.9963	0.0078	0.9996	0.0006

Table 3 (continued)

Classifier	SARS-CoV-2											
	Accuracy		Kappa		Sensitivity		Specificity		AUC		F1 score	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
SVM RBF G = 0.01; C = 10	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	0.9999	0.0003	0.9999	0.0003
SVM RBF G = 0.25; C = 10	99.9828	0.0531	0.9979	0.0065	0.9998	0.0006	1.0000	0.0000	0.9999	0.0003	0.9999	0.0003
SVM RBF G = 0.5; C = 10	99.9603	0.0799	0.9950	0.0101	0.9998	0.0005	0.9943	0.0144	0.9971	0.0072	0.9998	0.0004

and the three MLP configurations are also considered close to optimal, with the sample means of all metrics greater than 99%, but with sample standard errors much higher than those of Bayes net, at least 100 times greater for AUC and specificity. For the Bayes net, as described on Table 3, sensitivity, specificity, and AUC reached 0.9997 (0.0007), 1.0000 (0.0000), and 1.0000 (0.0000), respectively.

Analyzing the results for the SVMs, we notice that the accuracy values are all greater than 95%. The results for the RBF kernel are similar and stable, concentrated around 95% for C = 0.01. For RBF kernel with G = 0.25 and C = 0.1, sensibility, specificity, and AUC reached 0.9998 (0.0005), 1.000 (0.000), and 0.9999 (0.0003), respectively. For the linear kernel, the accuracy results vary around 99% with less than 1% variation, with higher C values (C = 0.1,1, and 10).

Table 3 also shows that the accuracy results for Random Tree correspond to good classification results and that there was no bias due to base imbalance: the sensitivity was 0.999 (0.001), the specificity was 0.98 (0.03), and the AUC was 0.99 (0.02). Considering the accuracy, the best results were those obtained with 40, 50, 60, 70, 80, 90, and 100 trees. For the last case, sensitivity, specificity, and AUC reached 1.000 (0.000), 0.998 (0.008), and 1.000 (0.000), respectively.

Figures 10 and 11 illustrate the sensitivity and specificity values for the database with all 43 attributes, respectively. For building the boxplots, only one example of each type of classifier that presented a good performance was used, since many of them presented similar results. The selected examples are highlighted in the Table 3.

When we analyze the results of the box plots in Figs. 10 and 11, we can see that the sensitivity and specificity results are quite accurate, with results above 0.9 in both metrics. Considering the sensitivity boxplots, we can see that the selected classifiers have similar performance, difficult to distinguish visually. On the other hand, when analyzing the specificity plots, we see the presence of outliers for the random tree and MLP classifiers.

In order to distinguish the best classifier among those tested, we performed the paired *t*-test, aiming to compare the random forest and Bayes network, which presented the highest mean values of AUC, and other metrics above 99.9%. In this way, a *p*-value of 0.000000346 was found, using the accuracy values. It can be interpreted that there is a statistically significant difference between these two classifiers, and, therefore, the random forest with 100 trees classifier was chosen, considering that it has a higher average accuracy.

The feature selection method based on the PSO algorithm returned the following eight attributes that were statistically most relevant for classification: serum glucose, lactic

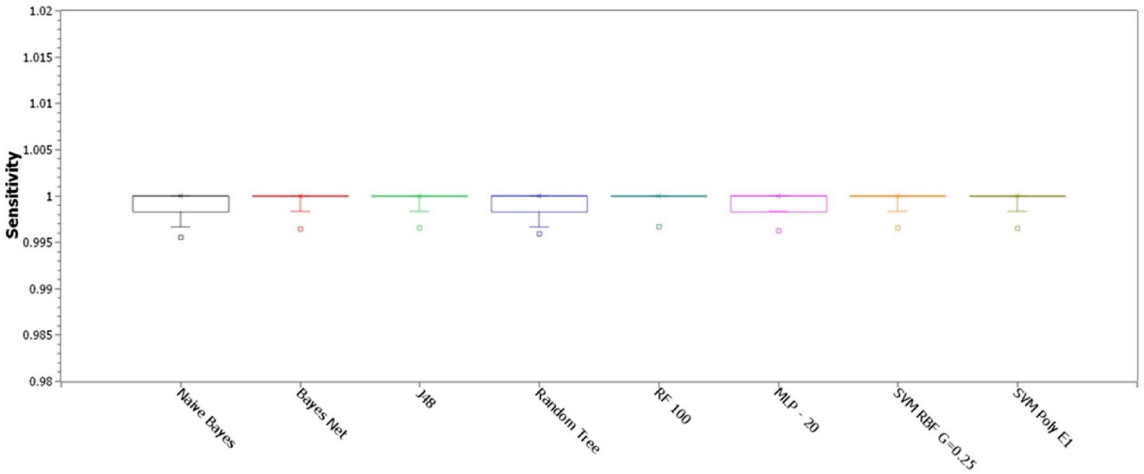

Fig. 10 Sensitivity of one of the tested configurations for each classifier type, considering the original 43-feature dataset. Source: Authors

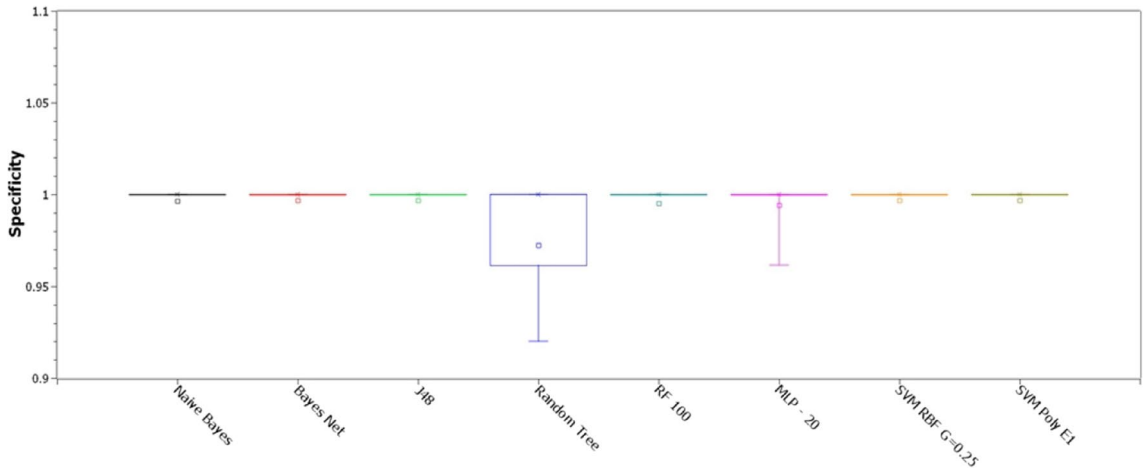

Fig. 11 Specificity of one of the tested configurations for each classifier type, considering the original 43-feature dataset. Source: Authors

dehydrogenase, lipase dosage, partial thromboplastin time, troponin, ferritin, D-dimer, and indirect bilirubin. Table 4 shows the average and standard error for accuracy, sensitivity, specificity, and area under the ROC curve results, taking into account these 8-feature PSO dimension-reduced dataset. In addition, Figs. 12 and 11 present the sensitivity and specificity behavior according to the set of classifiers we adopted herein this work.

Table 4 presents detailed results as sample average and standard error regarding accuracy, sensitivity, specificity, and AUC. In a similar way to the classification using all the extracted attributes, the sensitivity of the selected settings showed incredibly good results, as can be seen in Fig. 12. In addition, the mean specificity also remained close to 1.000. According to the Table 3, the best results were obtained with the Bayes Net classifier, with all metrics concentrated at 100% out of 1. The other classifiers also showed average accuracy above 95%. Among them, SVMs with $C = 0.001$ presented the worst performance. In contrast, for experiments with SVMs and C equal to 0.1, the results for the linear kernels, polynomial of degrees 2 and 3, and RBF with γ of 0.25 and 0.5, were significantly improved, all of them in the range of 99.9%. For C equal to 1.0, all results were considerably improved for all settings, including degrees of 4 and 5. For C equal to 10.00, there was no noticeable improvement compared to C equal to 1.00. Table 4 shows that, in this database with the 8 statistically most relevant hematological parameters, all tested SVM configurations were robust to high database imbalance only for C equal to 1.0 and 10.0, according to values of specificity and AUC, with grade 4 and 5 polynomial kernels being the least robust configurations. In addition, Table 4 details that Random Forests and the J48 decision tree are well suited to support the diagnosis of COVID-19 using the eight selected hematological parameters: accuracy of 99.98% (0.05), sensitivity of 0.998 (0.006), and maximum specificity and AUC with sample standard error less than 0.0001, for classification using random forest with 100 trees (Fig. 13).

Humanly intelligible models to support the clinical diagnosis of COVID-19

Figure 14 shows the J48 decision tree trained on the 8-feature dataset: serum glucose, partial thromboplastin time, troponin, lipase dosage, lactic dehydrogenase, ferritin, indirect bilirubin, and D-dimer. Training was done using 10- fold cross validation. The result shows that, for all 6215 records, the accuracy was 99.9839%, the sensitivity, specificity, and AUC were maximum. Only a single instance was misclassified: a negative record for COVID-19 out of the total of 5457 negatives was classified as positive. All COVID-19 positive records were classified correctly.

To assess the influence of the other 7 hematological parameters on the diagnosis of COVID-19, we retrained the J48 decision tree after removing the partial thromboplastin time. Training was done using tenfold cross validation. Figure 15 illustrates the resulting decision-making protocol, where only the following hematological parameters are statistically relevant: ferritin, troponin, lipase dosage, and serum glucose. The accuracy obtained was 99.8552%, sensitivity of 0.999, specificity of 0.985, and AUC of 0.985. We had 5 false positives and 4 false negatives.

By removing the partial thromboplastin time and lipase dosage, we can assess the importance of the other 6 hematological parameters. We train a J48 decision tree again using tenfold cross validation. The results are shown in Fig. 16. The hematological parameters considered most relevant this time were: ferritin, troponin, D-dimer, and serum glucose. The accuracy obtained was 99.8713%, with a sensitivity of

Table 4 Sample mean and standard error for accuracy, sensitivity, specificity, and area under ROC curve (AUC) for all classifiers, considering the 8-feature PSO dimension-reduced dataset. Source: Authors

Classifier	SARS-CoV-2 PSO											
	Accuracy		Kappa		Sensitivity		Specificity		AUC		F1 score	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Naïve Bayes	**99.9689**	**0.1369**	**0.9963**	**0.0149**	**0.9997**	**0.0014**	**1.0000**	**0.0000**	**1.0000**	**0.0001**	**0.9998**	**0.0007**
Bayes Network	**100.0000**	**0.0000**	**1.0000**	**0.0000**	**1.0000**	**0.0000**	**1.0000**	**0.0000**	**1.0000**	**0.0000**	**1.0000**	**0.0000**
J48 tree	**99.9839**	**0.0483**	**0.9980**	**0.0059**	**0.9998**	**0.0005**	**1.0000**	**0.0000**	**0.9999**	**0.0003**	**0.9999**	**0.0003**
Random tree	**99.9539**	**0.0849**	**0.9942**	**0.0107**	**0.9998**	**0.0006**	**0.9939**	**0.0151**	**0.9969**	**0.0076**	**0.9998**	**0.0004**
Random forest (10 trees)	99.9812	0.0517	0.9977	0.0064	0.9998	0.0005	0.9992	0.0054	1.0000	0.0001	0.9999	0.0003
Random forest (20 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (30 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (40 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (50 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (60 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (70 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (80 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (90 trees)	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	1.0000	0.0000	0.9999	0.0003
Random forest (100 trees)	**99.9839**	**0.0483**	**0.9980**	**0.0059**	**0.9998**	**0.0005**	**1.0000**	**0.0000**	**1.0000**	**0.0000**	**0.9999**	**0.0003**
MLP (20 neurons)	**99.9496**	**0.0845**	**0.9937**	**0.0106**	**0.9997**	**0.0007**	**0.9955**	**0.0132**	**0.9999**	**0.0005**	**0.9997**	**0.0004**
MLP (50 neurons)	99.9458	0.0888	0.9932	0.0111	0.9997	0.0007	0.9946	0.0145	0.9999	0.0005	0.9997	0.0005
MLP (100 neurons)	99.9458	0.0888	0.9932	0.0112	0.9997	0.0007	0.9943	0.0154	0.9999	0.0005	0.9997	0.0005
SVM polynomial E1; C = 0.01	96.3626	0.5919	0.1952	0.1561	1.0000	0.0001	0.1240	0.1440	0.5620	0.0720	0.9814	0.0030
SVM polynomial E2; C = 0.01	95.9281	0.4686	0.0245	0.1154	1.0000	0.0002	0.0195	0.1152	0.5098	0.0575	0.9792	0.0024
SVM polynomial E3; C = 0.01	95.9078	0.4678	0.0160	0.1137	1.0000	0.0002	0.0149	0.1150	0.5074	0.0574	0.9791	0.0024
SVM polynomial E4; C = 0.01	95.9024	0.4675	0.0136	0.1133	1.0000	0.0002	0.0136	0.1149	0.5068	0.0574	0.9791	0.0024
SVM polynomial E5; C = 0.01	95.8890	0.4708	0.0131	0.1133	0.9998	0.0005	0.0135	0.1149	0.5066	0.0574	0.9790	0.0024
SVM RBF G = 0.01; C = 0.01	95.8488	0.0629	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.5000	0.0000	0.9788	0.0003
SVM RBF G = 0.25; C = 0.01	95.9019	0.4631	0.0130	0.1125	1.0000	0.0000	0.0128	0.1105	0.5064	0.0552	0.9791	0.0024
SVM RBF G = 0.5; C = 0.01	95.9067	0.4728	0.0148	0.1146	1.0000	0.0001	0.0142	0.1151	0.5071	0.0575	0.9791	0.0024
SVM polynomial E1; C = 0.1	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	0.9999	0.0003	0.9999	0.0003
SVM polynomial E2; C = 0.1	99.9673	0.0687	0.9960	0.0084	0.9997	0.0007	1.0000	0.0000	0.9998	0.0004	0.9998	0.0004
SVM polynomial E3; C = 0.1	99.9673	0.0687	0.9960	0.0084	0.9997	0.0007	1.0000	0.0000	0.9998	0.0004	0.9998	0.0004
SVM polynomial E4; C = 0.1	96.1480	0.4815	0.1227	0.1305	0.9998	0.0005	0.0763	0.1190	0.5381	0.0594	0.9803	0.0025
SVM polynomial E5; C = 0.1	96.0097	0.4765	0.0659	0.1237	0.9998	0.0005	0.0430	0.1171	0.5214	0.0585	0.9796	0.0024
SVM RBF G = 0.01; C = 0.1	95.8488	0.0629	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.5000	0.0000	0.9788	0.0003
SVM RBF G = 0.25; C = 0.1	**99.9839**	**0.0483**	**0.9980**	**0.0059**	**0.9998**	**0.0005**	**1.0000**	**0.0000**	**0.9999**	**0.0003**	**0.9999**	**0.0003**
SVM RBF G = 0.5; C = 0.1	99.9689	0.0663	0.9961	0.0083	0.9998	0.0005	0.9964	0.0113	0.9981	0.0057	0.9998	0.0003
SVM polynomial E1; C = 1	**99.9839**	**0.0483**	**0.9980**	**0.0059**	**0.9998**	**0.0005**	**1.0000**	**0.0000**	**0.9999**	**0.0003**	**0.9999**	**0.0003**
SVM polynomial E2; C = 1	99.9673	0.0687	0.9960	0.0084	0.9997	0.0007	1.0000	0.0000	0.9998	0.0004	0.9998	0.0004
SVM polynomial E3; C = 1	99.9517	0.0816	0.9941	0.0100	0.9995	0.0009	1.0000	0.0000	0.9997	0.0004	0.9997	0.0004
SVM polynomial E4; C = 1	99.9517	0.0816	0.9941	0.0100	0.9995	0.0009	1.0000	0.0000	0.9997	0.0004	0.9997	0.0004
SVM polynomial E5; C = 1	99.9517	0.0816	0.9941	0.0100	0.9995	0.0009	1.0000	0.0000	0.9997	0.0004	0.9997	0.0004
SVM RBF G = 0.01; C = 1	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	0.9999	0.0003	0.9999	0.0003
SVM RBF G = 0.25; C = 1	99.9684	0.0667	0.9960	0.0084	0.9998	0.0005	0.9963	0.0115	0.9980	0.0057	0.9998	0.0003
SVM RBF G = 0.5; C = 1	99.9678	0.0671	0.9960	0.0084	0.9998	0.0005	0.9961	0.0117	0.9980	0.0058	0.9998	0.0003
SVM polynomial E1; C = 10	99.9689	0.0676	0.9962	0.0083	0.9997	0.0007	1.0000	0.0000	0.9998	0.0004	0.9998	0.0004
SVM polynomial E2; C = 10	99.9657	0.0711	0.9958	0.0087	0.9996	0.0007	1.0000	0.0000	0.9998	0.0004	0.9998	0.0004
SVM polynomial E3; C = 10	99.9517	0.0816	0.9941	0.0100	0.9995	0.0009	1.0000	0.0000	0.9997	0.0004	0.9997	0.0004
SVM polynomial E4; C = 10	99.9517	0.0816	0.9941	0.0100	0.9995	0.0009	1.0000	0.0000	0.9997	0.0004	0.9997	0.0004
SVM polynomial E5; C = 10	99.9517	0.0816	0.9941	0.0100	0.9995	0.0009	1.0000	0.0000	0.9997	0.0004	0.9997	0.0004

Table 4 (continued)

Classifier	SARS-CoV-2 PSO											
	Accuracy		Kappa		Sensitivity		Specificity		AUC		F1 score	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
SVM RBF G = 0.01; C = 10	99.9839	0.0483	0.9980	0.0059	0.9998	0.0005	1.0000	0.0000	0.9999	0.0003	0.9999	0.0003
SVM RBF G = 0.25; C = 10	99.9807	0.0540	0.9976	0.0067	0.9998	0.0005	0.9992	0.0054	0.9995	0.0027	0.9999	0.0003
SVM RBF G = 0.5; C = 10	99.9684	0.0667	0.9960	0.0084	0.9998	0.0005	0.9963	0.0115	0.9980	0.0057	0.9998	0.0003

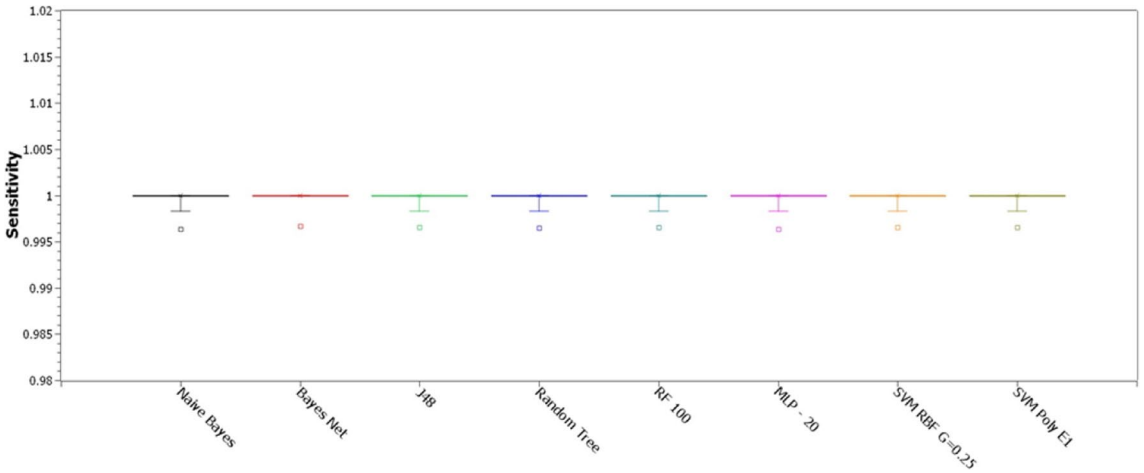

Fig. 12 Sensitivity of one of the tested configurations for each classifier type, considering the 8-feature PSO dimension-reduced dataset. Source: Authors

0.999, specificity of 0.981, and AUC of 0.990. We get 3 false positives and 5 false negatives.

We retrained a J48 decision tree removing the partial thromboplastin time, ferritin, and lipase dosage from the database. The training was done using tenfold cross validation. The decision protocol is shown in Fig. 17. All 5 remaining hematological parameters were considered relevant in the diagnostic decision: serum glucose, D-dimer, troponin, lactic dehydrogenase, and indirect bilirubin. The accuracy obtained was 99.7908%, with a sensitivity of 0.998, specificity of 0.955, and AUC of 0.982. We got 1 false positive and 12 false negatives.

Discussion

Evaluation of classifiers to support the diagnosis of COVID-19

Given that the best results were obtained with random forests and these classifiers are sufficiently robust to noisy data (possible typing errors of hematological parameters and missing data that need to be statistically estimated), we decided to adopt a Random Forest with 100 trees to implement the Heg.IA web for COVID-19 diagnosis support. Another advantage of Random Forests and decision trees in general is their low computational cost of processing, although the memory consumption can be considerable. Also, sorting with decision trees does not require input attributes to be rescheduled through normalization or standardization, a necessary preprocessing stage when we have to work with neural networks, support vector machines or Bayesian methods. Considering the 100-tree random forest, specificity, sensitivity, and AUC reached 1.0000 (0.0000), 0.9983 (0.0079), and 1.0000 (0.0000), respectively.

These results show that the 43 attributes (age and 42 hematological attributes) adopted in this work can be used to diagnose patients with symptoms of COVID-19 with a precision close to that obtained with RT-PCR, as shown in Table 3, for the results with Random Forests. Therefore, the Heg.IA web system can be an important aid for the development of a rapid protocol for the clinical diagnosis of COVID-19, providing more certainty for clinical practice. These results also show that these 42 hematological parameters, given the high accuracy, sensitivity, specificity, and AUC values obtained, comparable to those obtained with RTPCR, can be very important for monitoring COVID-19-positive patients.

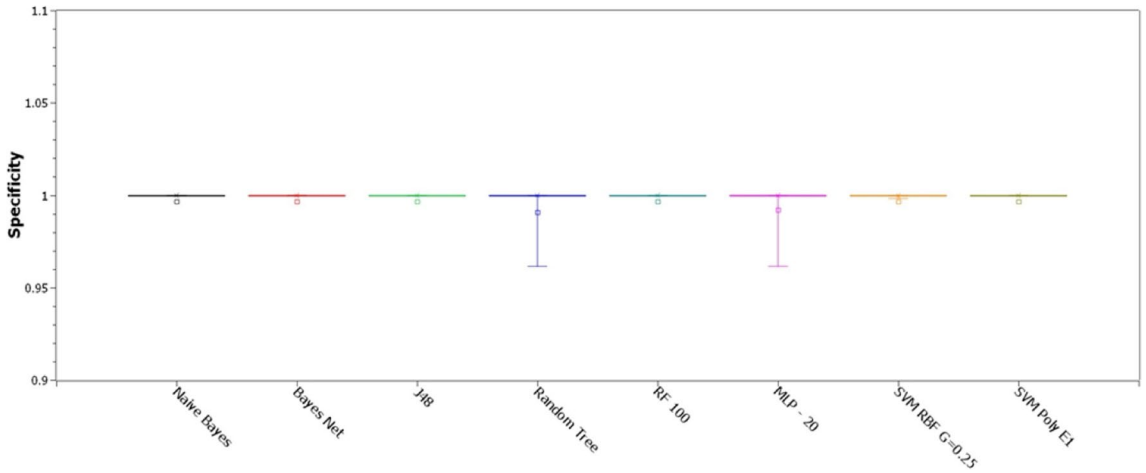

Fig. 13 Specificity of one of the tested configurations for each classifier type, considering the 8-feature PSO dimension-reduced dataset. Source: Authors

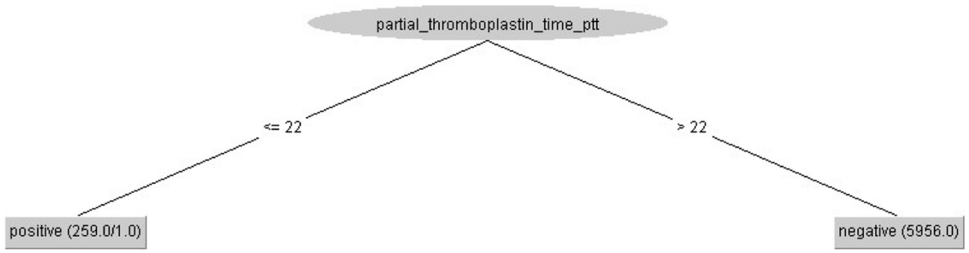

Fig. 14 Clinical decision protocol based on J48 decision tree trained on the 8-feature dataset: serum glucose, partial thromboplastin time, troponin, lipase dosage, lactic dehydrogenase, ferritin, indirect bilirubin, and D-dimer. Training was done using 10- fold cross validation. Reduced partial thromboplastin time (<22) is associated to symptomatic COVID-19 case. The accuracy was 99.9839%, the sensitivity, specificity, and AUC were maximum. We had 1 false positive. Source: Authors

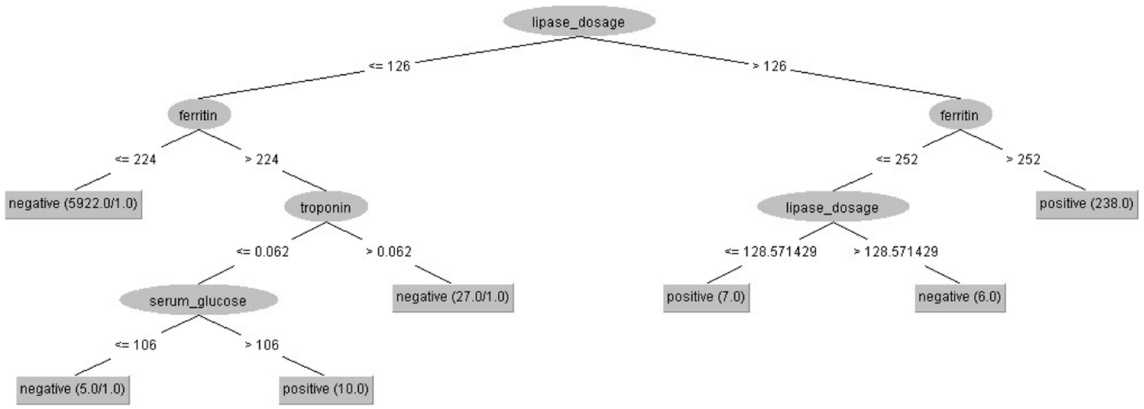

Fig. 15 Clinical decision protocol based on J48 decision tree trained on the 7-feature dataset. Only the following were statistically relevant: lipase dosage, ferritin, troponin, and serum glucose. Training was done using tenfold cross validation. Lipase dosage was dominant in this decision protocol. The accuracy was 99.8552%, sensitivity of 0.999, specificity of 0.985, and AUC of 0.985. We had 5 false positives and 4 false negatives. Source: Authors

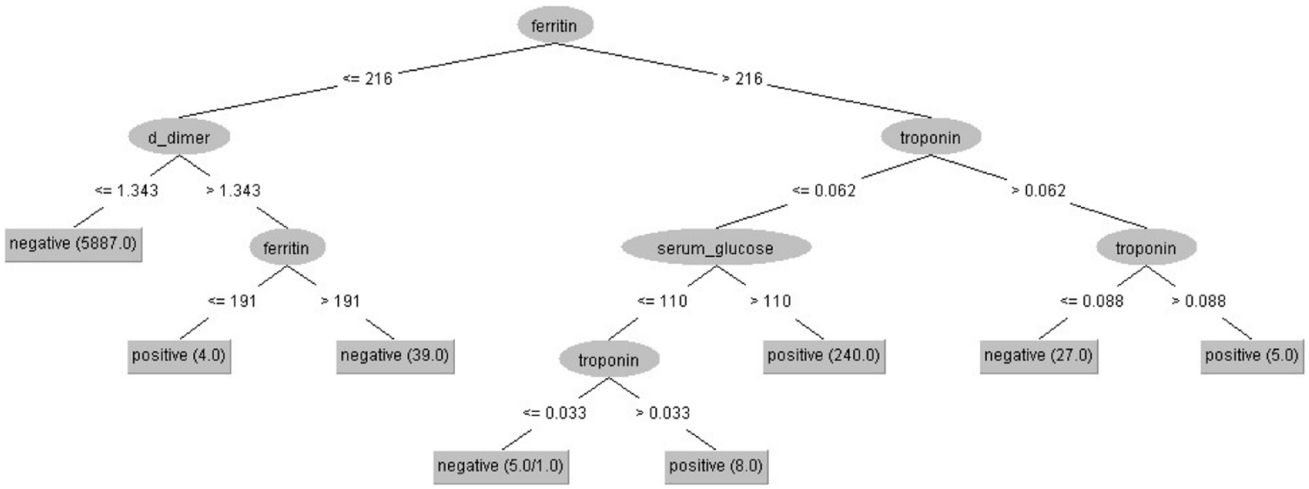

Fig. 16 Clinical decision protocol based on J48 decision tree trained on the 6-feature dataset. Only the following were statistically relevant: ferritin, troponin, D-dimer, and serum glucose. Training was done using tenfold cross validation. The accuracy obtained was 99.8713%, with a sensitivity of 0.999, specificity of 0.981, and AUC of 0.990. We get 3 false positives and 5 false negatives. Source: Authors

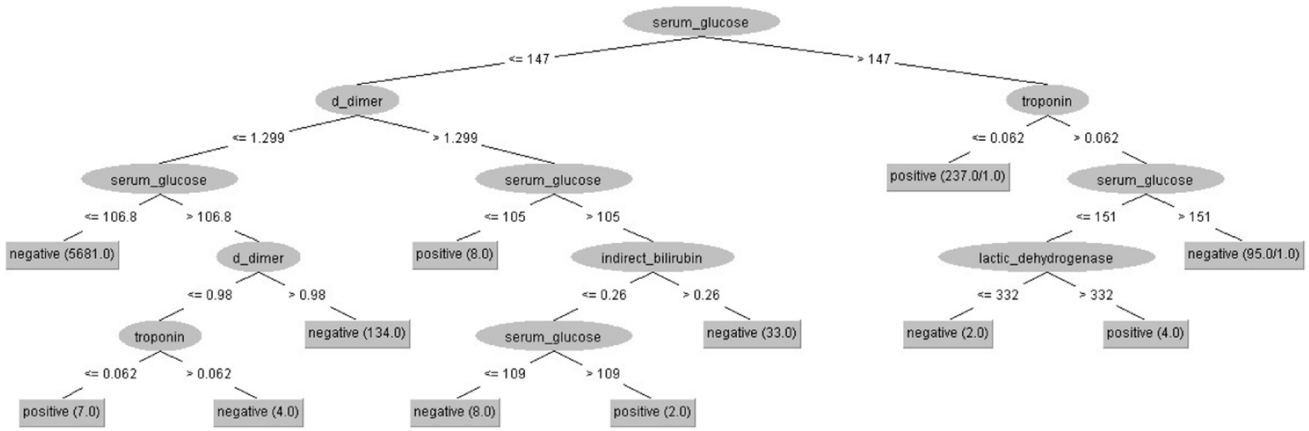

Fig. 17 Clinical decision protocol based on J48 decision tree trained on the 5-feature dataset. All remaining features were statistically relevant: serum glucose, D-dimer, troponin, lactic dehydrogenase, and indirect bilirubin. Training was done using tenfold cross validation. The accuracy obtained was 99.7908%, with a sensitivity of 0.998, specificity of 0.955, and AUC of 0.982. We got 1 false positive and 12 false negatives. Source: Authors

Furthermore, the results also show that the eight hematological parameters chosen by the PSO algorithm are very important for the clinical diagnosis of COVID-19 in symptomatic patients, reaching results comparable to those obtained with RT-PCR, the gold standard method for the diagnosis of COVID-19. Thereby, these hematological parameters are strongly related to COVID-19 itself, and should not only be part of COVID-19 clinical diagnosis protocols, but also protocols for patient follow-up and assessment of patient progress and severity of disease, as the review of the state of the art demonstrates.

The change in serum glucose by COVID-19 may explain the fact that diabetes is considered a comorbidity that can increase the chances of worsening the patient's death by COVID-19. Indirect bilirubin is formed by the breakdown of red blood cells and is then transported to the liver. Bilirubin can be related to several complications in the blood, changes in heart function and liver problems. This, in turn, may be strongly related to how the chances of aggravation and death in patients with heart disease and liver disease may increase. Troponin alterations, also considered statistically relevant in this research, also show how patients with heart disease can have their chances of worsening the disease. D-dimer is one of the breakdown products of fibrin, a protein that is involved in clot formation. Therefore, when there are changes in the coagulation process, it is possible

that there is a greater amount of circulating D-dimer. Lactic dehydrogenase is an enzyme present in all cells, but in greater concentration in the lungs, heart, skeletal muscle and liver. High levels of lactic dehydrogenase in patients with COVID-19 increase the chances of aggravation of the infection and, therefore, the probability of death. Lipase is an enzyme that is part of the digestive process. It works by breaking down fat molecules to be more easily absorbed by the intestine. The lipase test is a clinical analysis that helps identify changes in the levels of this enzyme in the individual's body. High levels of lipase are associated with acute pancreatitis. Several studies show that COVID-19 also affects the pancreas in moderate and severe cases. State-of-the-art investigations show that the occurrence of pancreatitis at different levels is reasonably common in patients positive for COVID-19 and may be a highly important clinical occurrence in intubated patients.

Reduced partial thromboplastin time indicates the possibility of early blood clotting. It is strongly related to the occurrence of thrombosis, thrombopenia and other problems related to early coagulation, common to moderate and severe cases of COVID-19 and in post- COVID-19 complicators. Ferritin is a protein produced mainly by the liver, whose basic functions are: transporting iron and mediating the inflammation process. High ferritin can be a symptom of inflammation or infections, as it is an acute phase protein, which may be increased in the following situations: COVID-19, hemolytic anemia, megaloblastic anemia, alcoholic liver disease, Hodgkin's lymphoma, myocardial infarction, leukemia, and hemochromatosis. Symptoms of excess ferritin are joint pain, tiredness, shortness of breath or abdominal pain, i.e., symptoms common to COVID-19. This information, combined with the results obtained with decision trees, suggest the clinical use of partial time thromboplastin and ferritin both for the rapid clinical diagnosis of symptomatic cases of COVID-19 without the aid of diagnostic support systems and for the assessment of severity of the disease and the follow-up of the patient's clinical situation.

These results may also indicate the possibility of constructing specific treatments for moderate and severe cases of COVID-19, which can be based on therapies with anticoagulants, IL-6 blocking drugs. Another possibility may be to adapt therapies aimed at reducing ferritin for COVID-19 cases, respecting the degree of disease severity and patient limitations. The proposed system has been used to support clinical diagnosis and assessment of disease severity in patients admitted to intensive and semi-intensive care units as a case study in the city of Paudalho, Brazil.

Furthermore, it is known that reference data for hematological parameters vary according to the age and gender of patients. Since the absolute values of hematological parameters were used for the development of models to support the diagnosis of COVID-19, these reference values were not investigated. Discussions about reference values were therefore outside the scope of this work. However, in future work, we can investigate the influence of binary attributes defined from the reference values by gender and age on the diagnosis of COVID-19 and how these results can dialogue with the clinical diagnosis.

Humanly intelligible models to support the clinical diagnosis of COVID-19

Regarding protocol proposals and human assessment of the features, Fig. 14 shows that, for the symptomatic case records of the public health system in the municipality of Paudalho, Brazil, the most important factor for a positive diagnosis for COVID-19 is the low partial thromboplastin time (less than or equal to 22). This is indicative of the possibility of early blood clotting, a clinical occurrence strongly associated with thrombopenia, thrombosis and blood clot formation. Since all patient records used in this research correspond to symptomatic patients who sought medical care, it is reasonable to consider that all positive cases of COVID-19 are moderate or severe. This raises the hypothesis that, for symptomatic COVID-19 patients in moderate or severe condition, treatments with anticoagulants such as low molecular weight heparin may be successful. In fact, the literature and clinical practice report that these treatments are already being used with relative success, but they are restricted to very severe cases, often in intubated patients. For this protocol, the accuracy was 99.98%, whilst the sensitivity, specificity, and AUC were maximum (approximately 1.00).

Figure 15 shows the decision protocol for clinical diagnosis generated automatically by decision tree J48 after excluding partial thromboplastin time. The protocol shows the importance of lipase dosing, ferritin, troponin, and serum glucose for the clinical diagnosis, monitoring, and evaluation of COVID-19. Most patients with COVID-19 (238 out of 258) had high levels of lipase dosage (greater than 126) and ferritin (greater than 252). These appear to be critically ill patients, as ferritin indicates that the patient is highly symptomatic, whereas high-dose lipase may indicate cases of pancreatitis. A few patients were positive for COVID-19 for low ferritin (less than 252) and lipase dosage between 126 and 129 (7 patients). Other 10 patients had low lipase dosage (less than 126), but high ferritin (greater than 224), low troponin (less than 0.062), and high serum glucose (greater than 106). This may indicate the influence of high glucose levels in the worsening of COVID-19 symptoms, but it may be associated with milder or moderate cases. However, to be more certain, it would be necessary to have access to the clinical histories of these patients, which was not possible in this study. For this protocol, accuracy was 99.8552%, and we got sensitivity of 0.999, specificity of 0.985, and AUC of 0.985.

Figure 16 shows the decision protocol obtained by removing the partial thromboplastin time and lipase dosage. In this case, we can observe that most patients with COVID-19 had high ferritin (greater than 216), low troponin (less than 0.062), and high serum glucose (greater than 110). A few patients had different situations: 5 patients with COVID-19 had high ferritin (greater than 216) and high troponin (greater than 0.088); 8 patients had high ferritin (greater than 216), low serum glucose (less than 110), and troponin between 0.033 and 0.062. However, for 4 patients, D-dimer was high (greater than 1.343) and ferritin was low (lower than 191), which may indicate complicators due to previous heart disease, but these conditions would need to be verified in the patients' clinical histories. For this protocol, we got accuracy of 99.8713%, sensitivity of 0.999, specificity of 0.981, and AUC of 0.990.

Figure 17 shows the decision protocol obtained automatically by the J48 decision tree after removing the partial thromboplastin time, lipase dosage, and ferritin. For this protocol, most patients with COVID-19 (237) had high serum glucose (greater than 147) and low but positive troponin (less than 0.062). As this hematologic parameter is often negative for healthy people, this condition may indicate a potential for COVID-19 to negatively alter cardiac function, possibly inducing myocardial damage and, depending on the patient's history, myocardial infarction. The other patients did not present a specific rule for the relationship between hematological parameters. Its conditions were expressed in many rules. Some patients appear to be heart disease (4 patients) because they have high troponin. These patients had high lactic dehydrogenase (greater than 332) and serum glucose between 147 and 151. One group of patients (8 patients) had high D-dimer (greater than 1299) and low serum glucose (less than 105). Another group, with more specific conditions (7 patients), had low troponin (less than 0.062), low D-dimer (less than 0.98), and serum glucose between 106.8 and 147. Finally, only 2 patients had this condition: serum glucose between 109 and 147, low indirect bilirubin (less than 0.26), and high D-dimer (greater than 1.299). For this protocol, we have accuracy of 99.79%, sensitivity of 0.998, specificity of 0.955, and AUC of 0.982.

Taking into account the use of Heg.IA web as an online solution to support clinical diagnosis, based on the random forest classifier of 100 trees, an accuracy of 99.98% (0.05), sensitivity of 0, 9998 (0.0005), specificity of 1.0000 (0.0000), and area under the ROC curve of 1.0000 (0.0000). Although these results were obtained for a different database from the one used by Wang et al. (2020a), Barbosa et al. (2021a), Barbosa et al. (2021b), and Soares et al. (2020), the results are superior to those obtained by Wang et al. (2020a), with a linear discriminant analysis method that achieved an area under ROC curve of 0.938, sensitivity of 0.9000%. and specificity of 0.8470% when using the hematological parameters of neutrophil-to-lymphocyte ratio and red blood cell distribution width-coefficient of variation; by Barbosa et al. (2021a), with a random forest of 90 trees (accuracy of 92.89% (0.85), sensitivity of 0.9360 (0.0110), specificity of 0.9210 (0.0120), and area under the ROC curve of 0.984 (0.003); by Barbosa et al. (2021b), with a Bayesian Network (accuracy of 95.160% (0.690), sensitivity of 0.9700 (0.0100), specificity of 0.9400 (0.0100), and area under the curve ROC of 0.9550 (0.0100); and by Soares et al. (2020), with a quadratic kernel support vector machine (average sensitivity of 70.25% [95%CI: 66.57–73.12%], average specificity of 85.98% [95%CI: 84.94–86.84%], and average area under the ROC curve of 86.78% [95%CI: 85.65–87.90%]). It is important to highlight that, in this work, the results were obtained for 7 most relevant blood tests, selected by particle swarm optimization, including patient age, while Barbosa et al. (2021a) obtained 12 most relevant blood tests by selection by particle swarm optimization. particles; Barbosa et al. (2021b) found 24 most relevant exams, combining particle swarm optimization and manual selection; and Soares et al. (2020) manually selected 16 blood tests. These results show that our proposed method was not only able to achieve better diagnostic support performance, but also considers a smaller set of hematological parameters (only 7), in addition to the patient's age, totaling 8 attributes.

It is also important to mention that there are several other studies that investigate the relationship between COVID -19 and changes in hematological parameters (Jalil et al. 2022; Matin et al. 2022; Szklanna et al. 2021; Usul et al. 2020). However, it is also worth noting that these works do not deal with the analysis of these data to support the diagnosis of COVID-19.

Conclusion

In this study, we developed a web system, Heg.IA, which seeks to optimize the diagnosis of COVID-19 by combining blood tests and artificial intelligence. From the system, a healthcare professional may have a diagnostic report after providing patient's age and 42 hematological parameters from common blood tests. The system is able to indicate if the patient is infected with SARS-CoV-2 virus. The proposed system is based on decision trees and achieved great performance of accuracy, sensitivity, specificity, and area under ROC for all tested scenarios. Considering SARS-CoV-2 detection, the system may play an important role as a highly efficient rapid test to aid at the clinical diagnosis as a way to provide rapid clinical diagnosis protocols. The system is designed to be easily updated with new data from blood test results available in electronic medical records as PDF files. The proposed system has been used to support

clinical diagnosis and assessment of disease severity in patients admitted to intensive and semi-intensive care units.

The experimental results demonstrate that the eight hematological parameters selected by PSO, i.e., serum glucose, lactic dehydrogenase, lipase dosage, partial thromboplastin time, troponin, ferritin, D-dimer, and indirect bilirubin, are very significant for the clinical diagnosis of COVID-19 in symptomatic patients. Classification results are comparable to those obtained with RT-PCR. Consequently, serum glucose, lactic dehydrogenase, lipase dosage, partial thromboplastin time, troponin, ferritin, D-dimer, and indirect bilirubin are closely related to COVID-19 dynamics. The experimental results are suggestive for the use of these eight hematological parameters as part of COVID-19 clinical diagnosis protocols, as well as patient monitoring assessment of patient progress and severity of disease, confirming state-of-the-art results. These hematological parameters may even be important for the construction of new specific treatments for COVID-19, which take into account the clinical and biochemical interpretation of these variables.

Acknowledgements We thank the Mayor of the City of Paudalho, Mr. Marcelo Gouveia, for the collaboration and partnership with our research team. We are also grateful to the Health Department of the Municipality of Paudalho and all the medical, nurses and biomedical professionals of the Municipal Public Health System who collaborated with this research.

Funding The authors received partial financial support of this research from the Federal University of Pernambuco, Google Cloud COVID19 Research Grant, and the Brazilian research agencies CAPES and CNPq.

Data availability Data supporting this study are available from the Department of Biomedical Engineering at Universidade Federal de Pernambuco, Recife, Brazil, under demand. Access to the data is subject to approval and a data sharing agreement due to the presence of sensible health data.

Declarations

Ethics standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of interest All authors declare no competing interests.

References

Abdar M, Yen NY, Hung JC-S. Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision trees. J Med Biol Eng. 2018;38(6):953–65.

Adetunji CO, Olaniyan OT, Adeyomoye O, Dare A, Adeniyi MJ, Alex E, Rebezov M, Garipova L, Shariati MA. ehealth, mhealth, and telemedicine for covid-19 pandemic. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, pages 157–168. Springer, 2022a.

Adetunji CO, Olaniyan OT, Adeyomoye O, Dare A, Adeniyi MJ, Alex E, Rebezov M, Isabekova O, Shariati MA. Smart sensing for covid-19 pandemic. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, pages 145–156. Springer, 2022b.

Adetunji CO, Olaniyan OT, Adeyomoye O, Dare A, Adeniyi MJ, Alex E, Rebezov M, Koriagina N, Shariati MA. Diverse techniques applied for effective diagnosis of covid-19. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, pages 45–58. Springer, 2022c.

Adetunji CO, Olaniyan OT, Adeyomoye O, Dare A, Adeniyi MJ, Alex E, Rebezov M, Petukhova E, Shariati MA. Internet of health things (ioht) for covid-19. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, pages 75–87. Springer, 2022d.

Adetunji CO, Olaniyan OT, Adeyomoye O, Dare A, Adeniyi MJ, Alex E, Rebezov M, Petukhova E, Shariati MA. Machine learning approaches for covid-19 pandemic. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, pages 133–143. Springer, 2022e.

Barakat N, Bradley AP, Barakat MNH. Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Trans Inf Technol Biomed. 2010;14(4):1114–20.

Barbosa VAF, Santana MA, Andrade MKS, Lima RCF, Santos WP. Deepwavelet neural networks for breast cancer early diagnosis using mammary termographies. In H. Das, C. Pradhan, and N. Dey, editors, Deep Learning for Data Analytics: Foundations, Biomedical Applications, and Challenges. Academic Press, London, 1st edition, 2020.

Bataille S, Pedinielli N, Bergounioux J-P. Could ferritin help the screening for COVID-19 in hemodialysis patients? Kidney Int. 2020;98(1):235–6.

Biswas S, Dash S. LSTM-CNN deep learning–based hybrid system for real-time COVID-19 data analysis and prediction using Twitter data. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, pages 239–257. Springer, 2022.

Borges L. Medidas de acurácia diagnóstica na pesquisa cardiovascular. Int J Cardiovasc Sci. 2016;29(3):218–22.

Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory, 1992; 144–152.

Bouckaert RR. Bayesian network classifiers in weka for version 3-5-7. Artif Intell Tools. 2008;11(3):369–87.

Bratton D, Kennedy J. Defining a standard for particle swarm optimization. In 2007 IEEE Swarm Intelligence Symposium, 120–127. IEEE, 2007.

Brazilian Ministry of Health. *Guidelines for the diagnosis and treatment of COVID-19*. Brazilian Society of Clinical Analyzes, 2020. URL www.sbac.org.br/blog/2020/04/09/ diretrizes-para-diagnostico-e-tratamento-da-covid-19/. Last accessed: 2020 June. 03. L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, evaluation and treatment coronavirus (covid-19). In *StatPearls [Internet]*. Stat Pearls Publishing, 2020.

Chatterjee S, Sengupta T, Majumder S, Majumder R. COVID-19: a probable role of the anticoagulant ProteinS in managing COVID-19-associated coagulopathy. Aging (albany NY). 2020;12(16):15954.

Cheng J, Greiner R. Comparing bayesian network classifiers. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, 101–108. Morgan Kaufmann Publishers Inc., 1999.

Cheng J, Greiner R. Learning Bayesian belief network classifiers: Algorithms and System. Adv Artif Intell. 2001;2056(1):141–51.

Cheng L, Li H, Li L, Liu C, Yan S, Chen H, Li Y. Ferritin in the coronavirus disease2019 (covid-19): a systematic review and meta-analysis. J Clin Lab Anal. 2020;34(10):e23618.

Ciotti M, Angeletti S, Minieri M, Giovannetti M, Benvenuto D, Pascarella S, Sagnelli C, Bianchi M, Bernardini S, Ciccozzi M. Covid-19 outbreak: an overview. Chemotherapy. 2019;64(5–6):215–23.

Çomak E, Arslan A, Türkoglu I. A decision support system based on support vectorˇ machines for diagnosis of the heart valve diseases. Comput Biol Med. 2007;37(1):21–7.

Commowick O, Istace A, Kain M, Laurent B, Leray F, Simon M, Pop SC, Girard P, Ameli R, Ferré J-C, et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Sci Rep. 2018;8(1):1–17.

Connors JM, Levy JH. Covid-19 and its implications for thrombosis and anticoagulation. Blood J Am Soc Hematol. 2020;135(23):2033–40.

Cordeiro FR, Santos WP, Silva-Filho AG. A semi-supervised fuzzy growcut algorithm to segment and classify regions of interest of mammographic images. Expert Syst Appl. 2016;65:116–26.

Cordeiro FR, de Santos WP, Silva-Filho AG. Analysis of supervised and semisupervised growcut applied to segmentation of masses in mammography images. Comput Methods Biomech Biomed Eng: Imaging & Visualization. 2017;5(4):297–315.

Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.

Crooks VA, Andrews GJ, Pearce J. Routledge Handbook of Health Geography. Routledge, 2018.

Cruz T, Cruz T, Santos W. Detection and classification of lesions in mammographies using neural networks and morphological wavelets. IEEE Lat Am Trans. 2018;16(3):926–32.

Dahan S, Segal G, Katz I, Hellou T, Tietel M, Bryk G, Amital H, Shoenfeld Y, Dagan A. Ferritin as a marker of severity in COVID-19 patients: a fatal correlation. Israel Med Assoc J: IMAJ. 2020;22(8):494–500.

da Silva ACG, de Lima CL, da Silva CC, Moreno GMM, Silva EL, Marques GS, de Araújo LJB, Júnior LAA, de Souza SBJ, de Santana MA, Gomes JC, de Barbosa VAF, Musah A, Kostkova P, da Silva Filho AG, dos Santos WP. Machine learning approaches for temporal and spatio-temporal covid-19 forecasting: a brief review and a contribution. Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 2022; 333–357.

de Barbosa VAF, Gomes JC, de Santana MA, de Lima CL, Calado RB, Bertoldo Junior CR, de Albuquerque JEA, de Souza RG, de Araújo RJE, MattosJunior LAR, de Souza RE, dos Santos WP. Covid-19 rapid test by combining a random forest-based web system and blood tests. J Biomol Struct Dynamics. 2021a;2021:1–20.

de Barbosa VAF, Gomes JC, de Santana MA, de Jeniffer EA, de Souza RG, de Souza RE, dos Santos WP. Heg.IA: An intelligent system to support diagnosis of Covid-19 based on blood tests. Res Biomed Eng. 2021b;2021:1–18.

de Lima CL, da Silva CC, da Silva ACG, LuizSilva E, Marques GS, de Araújo LJB, Albuquerque Júnior LA, de Souza SBJ, de Santana MA, Gomes JC, de Barbosa VAF, Musah A, Kostkova P, dos Santos WP, da SilvaFilho AG. Covid-sgis: A smart tool for dynamic monitoring and temporal forecasting of covid-19. Front Public Health. 2020;8:761.

de Lima SM, da Silva-Filho AG, dos Santos W. A methodology for classification of lesions in mammographies using zernike moments, elm and svm neural networks in a multi-kernel approach. In 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 988–991. IEEE, 2014.

de Lima SM, da Silva-Filho AG, dos Santos WP. Detection and classification of masses in mammographic images in a multi-kernel approach. Comput Methods Programs Biomed. 2016;134:11–29.

de Santana MA, Pereira JMS, da Silva FL, de Lima NM, de Sousa FN, de Arruda GMS, de Lima RCF, da Silva WWA, de Santos WP. Breast cancer diagnosis based on mammary thermography and extreme learning machines. Res Biomed Eng. 2018;34:45–53 ISSN 2446-4740.

de Santana MA, Gomes JC, de Freitas Barbosa VA, de Lima CL, Bandeira J, Valença MJS, de Souza RE, Masood AI, dos Santos WP. An intelligent tool to support diagnosis of covid-19 by texture analysis of computerized tomography x-ray images and machine learning. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 259–282. Springer, 2022.

de Vasconcelos J, dos Santos W, de Lima R. Analysis of methods of classification of breast thermographic images to determine their viability in the early breast cancer detection. IEEE Lat Am Trans. 2018;16(6):1631–7.

Fan BE, Chong VCL, Chan SSW, Lim GH, Lim KGE, Tan GB, Mucheli SS, Kuperan P, Ong KH. Hematologic parameters in patients with COVID19 infection. Am J Hematol. 2020;04:2020.

Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett. 2006;27(8):861–74.

Fei Z, Ryeznik Y, Sverdlov A, Tan CW, Wong WK. An overview of healthcare data analytics with applications to the covid-19 pandemic. IEEE Trans Big Data. 2021;1–20:2021.

Gao Y, Li T, Han M, Li X, Wu D, Xu Y, Zhu Y, Liu Y, Wang X, Wang L. Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID19. J Med Virol, 2020.

Gardner MW, Dorling S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ. 1998;32(14–15):2627–36.

Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63(1):3–42.

Gnanambal S, Thangaraj M, Meenatchi V, Gayathri V. Classification algorithms with attribute selection: an evaluation study using weka. Int J Adv Netw Appl. 2018;9(6):3640–4.

Gomes JC, de Barbosa VAF, de Santana MA, Bandeira J, Valença MJS, de Souza RE, Ismael AM, dos Santos WP. Ikonos: An intelligent tool to support diagnosis of covid-19 by texture analysis of x-ray images. Res Biomed Eng. 2020a;2020:1–14.

Gomes JC, de Silva LHS, Ferreira J, Junior AAF, de Rocha ALS, Castro L, da Silva NRC, Fernandes BJT, dos Santos WP. Optimizing the molecular diagnosis of Covid-19 by combining RT-PCR and a pseudo-convolutional machine learning approach to characterize virus DNA sequences. bioRxiv, 2020b.

Gómez-Pastora J, Weigand M, Kim J, Wu X, Strayer J, Palmer AF, Zborowski M, Yazer M, Chalmers JJ. Hyperferritinemia in critically ill covid-19 patients–is ferritin the product of inflammation or a pathogenic mediator? *Clinica Chimica Acta*. Int J Clin Chem. 2020;2020(509):249–51.

Guncar G, Kukar M, Notar M, Brvar M, Cernelc P, Notar M, Notar M. An appli-ˇ cation of machine learning to haematological diagnosis. Sci Rep. 2018;8(1):1–12.

Hand DJ. Measuring classifier performance: a coherent alternative to the area under the roc curve. Mach Learn. 2009;77(1):103–23.

Hasan TT, Jasim MH, Hashim IA. Heart disease diagnosis system based on multilayer perceptron neural network and support vector machine. Int J Curr Eng Technol. 2017;77(55):2277–4106.

Hassan SA, Gull S, Akbar S, Hanif I, Iqbal S, Aziz MW. Artificial intelligence in coronavirus detection: Recent findings and future perspectives. In Intelligent Computing Applications for COVID-19, 23–48. CRC Press, 2021.

Haykin S. Neural networks: principles and practice. Bookman. 2001;11:900.
Iba T, Levy JH, Levi M, Thachil J. Coagulopathy in covid-19. J Thromb Haemost. 2020;18(9):2103–9.
I. B. d. G. e. E. IBGE. Censo brasileiro de 2010, 2010.
Jalil AT, Shanshool MT, Dilfy SH, Saleh MM, Suleiman AA. Hematological and serological parameters for detection of covid-19. J Microbiol Biotechnol Food Sci. 2022;11(4):e4229–e4229.
Jung Y, Hu J. A k-fold averaging cross-validation procedure. J Nonparametric Stat. 2015;27(2):167–79.
Kaggle. Diagnosis of COVID-19 and its clinical spectrum. Kaggle, 2020. URL www. kaggle.com/einsteindata4u/covid19. Last accessed: 2020 Apr. 07.
Kappert K, Jahic A, Tauber R.´ Assessment of serum ferritin as a biomarker in covid-19: bystander or participant? insights by comparison with other infectious and noninfectious diseases. Biomarkers, 2020; 1–10
Karimi M, Harouni M, Nasr A, Tavakoli N. Automatic lung infection segmentation of covid-19 in ct scan images. In Intelligent Computing Applications for COVID-19, 235–253. CRC Press, 2021.
Kennedy J, Eberhart R. Particle swarm optimization. In *Proceedings of ICNN'95*-International Conference on Neural Networks, volume 4, pages 1942–1948. IEEE, 1995.
Kernan KF, Carcillo JA. Hyperferritinemia and inflammation. Int Immunol. 2017;29(9):401–9.
Khan AR, Mehmood K, Ayesha N. Deep learning for covid-19 infection's diagnosis, prevention, and treatment. In Intelligent Computing Applications for COVID-19, 1–22. CRC Press, 2021.
Klok F, Kruip M, Van der Meer N, Arbous M, Gommers D, Kant K, Kaptein F, van Paassen J, Stals M, Huisman M, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res. 2020;191:145–7.
Kotsiantis SB. Decision trees: a recent overview. Artif Intell Rev. 2013;39(4):261–83.
Lerner B, Levinstein M, Rosenberg B, Guterman H, Dinstein L, Romem Y. Feature selection and chromosome classification using a multilayer perceptron neural network. In Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), volume 6, pages 3540–3545. IEEE, 1994.
Liao D, Zhou F, Luo L, Xu M, Wang H, Xia J, Gao Y, Cai L, Wang Z, Yin P, et al. Haematological characteristics and risk factors in the classification and prognosis evaluation of covid-19: a retrospective cohort study. Lancet Haematol. 2020;7(9):e671–8.
Lima S, Azevedo W, Cordeiro F, Silva-Filho A, Santos W. Feature extraction employing fuzzy-morphological decomposition for detection and classification of mass on mammograms. In Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Ann Conf. 2015;2015:801–4.
Lin Z, Long F, Yang Y, Chen X, Xu L, Yang M. Serum ferritin as an independent risk factor for severity in covid-19 patients. J Infect. 2020;81(4):647–79.
Liu J, Li S, Liu J, Liang B, Wang X, Wang H, Li W, Tong Q, Yi J, Zhao L, et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. EBioMedicine. 2020; 102763.
Long H, Nie L, Xiang X, Li H, Zhang X, Fu X, Ren H, Liu W, Wang Q, Wu Q. Ddimer and prothrombin time are the significant indicators of severe COVID-19 and poor prognosis. BioMed Res Int. 2020; 2020.
Luo Y, Szolovits P, Dighe AS, Baron JM. Using machine learning to predict laboratory test results. Am J Clin Pathol. 2016;145(6):778–88.
Majhi B, Thangeda R, Majhi R. A review on detection of covid-19 patients using deep learning techniques. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 59–74. Springer, 2022.
Martinez-Velazquez R, Tobón DPV, Sanchez A, El Saddik A, Petriu E, et al. A machinelearning approach as an aid for early covid-19 detection. Sensors. 2021;21(12):4202.
Matin S, Safarzadeh E, Rezaei N, Negaresh M, Salehzadeh H, Matin S, Shari-fiazar AH, Abazari M, Dadkhah M. Hematological parameters as diagnostic factors: Correlation with severity of covid-19. Acta Bio Medica: Atenei Parmensis. 93(2), 2022.
Menezes-Rodrigues FS, PadrãoTavares JG, Pires M, de Oliveira R, Guzella de Carvalho P, Ruggero Errante P, OmarTaha M, José Fagundes D, Caricati-Neto A. Anticoagulant and antiarrhythmic effects of heparin in the treatment of COVID-19 patients. J Thromb Haemost. 2020;18(8):2073–5.
Naraei P, Abhari A, Sadeghian A. Application of multilayer perceptron neural networks and support vector machines in classification of healthcare data. In *2016* Future Technologies Conference (FTC), 848–852. IEEE, 2016.
Negri EM, Piloto BM, Morinaga LK, Jardim CVP, Lamy SAE-D, Ferreira MA, D'Amico EA, Deheinzelin D. Heparin therapy improving hypoxia in COVID-19 patients: a case series. Front Physiol. 2020; 11.
Nourbakhsh Z. Covid-19 ct image segmentation and detection. In Intelligent Computing Applications for COVID-19, pages 97–122. CRC Press, 2021.
Organization WH. Weekly epidemiological update on COVID-19 – 16 March 2021, 2021. URL https://www.who.int/publications/m/item/weekly-epidemiological-update---16march-2021. Last accessed: March 20, 2021.
Pandit P, Krishnamurthy K, Bakshi B. Artificial intelligence (ai) and big data analytics for the covid-19 pandemic. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 1–17. Springer, 2022.
Pani SK, Dash S, dos Santos WP, Bukhari SAC, Flammini F. Assessing COVID19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis. Springer Cham, Cham, 2022.
Panigada M, Bottino N, Tagliabue P, Grasselli G, Novembrino C, Chantarangkul V, Pesenti A, Peyvandi F, Tripodi A. Hypercoagulability of COVID-19 patients in intensive care unit: a report of thromboelastography findings and other parameters of hemostasis. J Thromb Haemost. 2020;18(7):1738–42.
Pavoni V, Gianesello L, Pazzi M, Stera C, Meconi T, Frigieri FC. Evaluation of coagulation function by rotation thromboelastometry in critically ill patients with severe covid-19 pneumonia. J Thromb Thrombolysis. 2020;50:281–6.
Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, Baghbanzadeh M, Aghamohammadi N, Zhang W, Haque U. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? Int J Epidemiol 2020; 2020.
Pereira JMS, Santana MA, Lima RCF, Lima SML, Santos WP. Method for classification of breast lesions in thermographic images using elm classifiers. In W. P. dos Santos, M. A. de Santana, and W. W. A. da Silva, editors, Understanding a Cancer Diagnosis, 117–132. Nova Science, New York, 1 edition, 2020a.
Pereira JMS, Santana MA, Lima RCF, Santos WP. Lesion detection in breast thermography using machine learning algorithms without previous segmentation. In W. P. dos Santos, M. A. de Santana, and W. W. A. da Silva, editors, Understanding a Cancer Diagnosis, pages 81–94. Nova Science, New York, 1 edition, 2020b.
Pereira JMS, Santana MA, Silva WWA, Lima RCF, Lima SML, Santos WP. Dialectical optimization method as a feature selection tool for breast cancer diagnosis using thermographic images. In W. P. dos Santos, M. A. de Santana, and W. W. A. da Silva, editors,

Understanding a Cancer Diagnosis, 95–118. Nova Science, New York, 1 edition, 2020c.

Peter O, Swain S, Muduli K, Ramasamy A. IoT in Combating COVID-19 Pandemics: Lessons for Developing Countries. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 113–131. Springer, 2022.

Phung SL, Bouzerdoum A, Chai D. Skin segmentation using color pixel classification: analysis and comparison. IEEE Trans Pattern Anal Mach Intell. 2005;27(1):148–54.

Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. J Med Syst. 2002;26(5):445–63.

Poli R, Kennedy J, Blackwell T. Particle Swarm Optimization. Swarm Intell. 2007;1(1):33–57.

Prior L, Manley D, Sabel CE. Biosocial health geography: New 'exposomic' geographies of health and place. Prog Hum Geogr. 2019;43(3):531–52.

Raghu S, Sriraam N. Optimal configuration of multilayer perceptron neural network classifier for recognition of intracranial epileptic seizures. Expert Syst Appl. 2017;89:205–21.

Rodrigues AL, de Santana MA, Azevedo WW, Bezerra RS, Barbosa VA, de Lima RC, dos Santos WP. Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. Res Biomed Eng. 2019;35(3):213–22.

Rosário C, Zandman-Goddard G, Meyron-Holtz EG, D'Cruz DP, Shoenfeld Y. The hyperferritinemic syndrome: macrophage activation syndrome, Still's disease, septic shock and catastrophic antiphospholipid syndrome. BMC Med. 2013;11(1):1–11.

Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (covid-19) outbreak. J Autoimmun. 2020;109:102433.

Saba T, Khan AR. Intelligent Computing Applications for COVID-19: Predictions, Diagnosis, and Prevention. Boca Raton: CRC Press; 2021.

Sahu D, Agrawal T. Is it the Covid-19 happy hypoxia syndrome or the Covid-19 infodemic syndrome? Diabetes Metab Syndr. 2020;14(5):1399.

Santana MA, Pereira JMS, Lima RCF, Santos WP. Breast lesions classification in frontal thermographic images using intelligent systems and moments of haralick and zernike. In W. P. dos Santos, M. A. de Santana, and W. W. A. da Silva, editors, Understanding a Cancer Diagnosis, pages 65–80. Nova Science, New York, 1 edition, 2020.

Sharma S, Aroura S, Gupta A, Priyadarshini A. Bioinformatics in diagnosis of covid19. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 197–219. Springer, 2022.

Shi C, Tingting W, Li J-P, Sullivan MA, Wang C, Wang H, Deng B, Zhang Y. Comprehensive Landscape of Heparin Therapy for Covid-19. Carbohydr Polym. 2020; 117232.

Silva WWA, Santana MA, Silva Filho AG, Lima SML, Santos WP. Morphological extreme learning machines applied to the detection and classification of mammary lesions. In T. K. Gandhi, S. Bhattacharyya, S. De, D. Konar, and S. Dey, editors, Advanced Machine Vision Paradigms for Medical Image Analysis. Elsevier, London, 2020.

Soares F, Villavicencio A, Fogliatto FS, Rigatto MHP, Anzanello MJ, Idiart MA, Stevenson M. A novel specific artificial intelligence-based method to identify COVID-19 cases using simple blood exams. medRxiv. 2020.

Spiezia L, Boscolo A, Poletto F, Cerruti L, Tiberio I, Campello E, Navalesi P, Simioni P. COVID-19-related severe hypercoagulability in patients admitted to intensive care unit for acute respiratory failure. Thromb Haemost. 2020;120(6):998.

Sun T, Wang J, Li X, Lv P, Liu F, Luo Y, Gao Q, Zhu H, Guo X. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in ct based on a multi-dimensional data set. Comput Methods Programs Biomed. 2013;111(2):519–24.

Szklanna PB, Altaie H, Comer SP, Cullivan S, Kelliher S, Weiss L, Curran J, Dowling E, O'Reilly KM, Cotter AG, et al. Routine hematological parameters may be predictors of covid-19 severity. Front Med. 2021;8:682843.

Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang Y-Q, Wang Q, Miao H. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct Target Ther. 2020;5(1):1–3.

Tang N, Bai H, Chen X, Gong J, Li D, Sun Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. J Thromb Haemost. 2020;18(5):1094–9.

Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, Lai YL, Ng LC, Leo YS, Puong LT, Vasudevan SG, et al. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS Negl Trop Dis. 2008; 2(3)

Torcate AS, Fonseca FS, Lima ART, Santos FP, Oliveira TDMS, de Santana MA, Gomes JC, de Lima CL, de Freitas Barbosa VA, da Souza RE, dos Santos WP. Prediction of care for patients in a covid-19 pandemic situation based on hematological parameters. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, pages 169–196. Springer, 2022.

Tripathi A, Pandey AB, Singh AK, Jain A, Tyagi V, Vashist PC. Diagnosis for covid-19. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 89–111. Springer, 2022.

Tural Onur S, Altın S, Sokucu SN, Fikri BI, Barça T, Bolat E, Toptas M. Could ferritin level be an indicator of COVID-19 disease mortality? J Med Virol. 2021;93(3):1672–7.

Turner AJ, Hiscox JA, Hooper NM. Ace2: from vasopeptidase to sars virus receptor. Trends Pharmacol Sci. 2004;25(6):291–4.

Usharani B. Covid-19 detection using discrete particle swarm optimization clustering withimage processing. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 221–238. Springer, 2022.

Usul E, San I, Bekgöz B, Sahin A. Role of hematological parameters in covid-19 patients in the emergency room. Biomarkers Med. 2020;14(13):1207–15.

Van den Bergh F, Engelbrecht AP. A cooperative approach to particle swarm opti-mization. IEEE Trans Evol Comput. 2004;8(3):225–39.

Vargas-Vargas M, Cortés-Rojo C. Ferritin levels and covid-19. Rev Panam Salud Publica. 2020;44:e72.

Viecca M, Radovanovic D, Forleo GB, Santus P. Enhanced platelet inhibition treatment improves hypoxemia in patients with severe Covid-19 and hypercoagulability .A case control, proof of concept study. Pharmacol Res. 2020;158:104950.

Vilar HN, de Medeiros RM. Índice de aridez na Zona da Mata no Estado de Pernambuco-Brasil. J Environ Anal Progress. 2019;4(1):14–20.

Wang C, Deng R, Gou L, Fu Z, Zhang X, Shao F, Wang G, Fu W, Xiao J, Ding X, et al.Preliminary study to identify severe from moderate cases of covid-19 using combined hematology parameters. Annal Transl Med. 2020a; 8(9).

Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus– infected pneumonia in Wuhan. China Jama. 2020b;323(11):1061–9.

Wang J, Hajizadeh N, Moore EE, McIntyre RC, Moore PK, Veress LA, Yaffe MB, Moore HB, Barrett CD. Tissue plasminogen activator (tPA) treatment for Covid-19 associated acute respiratory distress syndrome (ARDS): a case series. J Thromb Haemost. 2020c;18(7):1752–5.

Wang X, Yang J, Teng X, Xia W, Jensen R. Feature selection based on rough sets and particle swarm optimization. Pattern Recogn Lett. 2007;28(4):459–71.

Wolff JGB, Marcondes DWC, de Santos WP, Bertemes-Filho P. Image reconstruction for covid-19 using multifrequency electrical impedance tomography. In Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis, 359–405. Springer, 2022.

Wright FL, Vogler TO, Moore EE, Moore HB, Wohlauer MV, Urban S, Nydam TL, Moore PK, McIntyre RC Jr. Fibrinolysis shutdown correlation with thromboembolic events in severe COVID-19 infection. J Am Coll Surg. 2020;231(2):193–203.

Wu Y-C, Chen C-S, Chan Y-J. The outbreak of covid-19: an overview. J Chin Med Assoc. 2020;83(3):217.

Yao J, Dwyer A, Summers RM, Mollura DJ. Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. Acad Radiol. 2011;18(3):306–14.

Zheng Y-Y, Ma Y-T, Zhang J-Y, Xie X. COVID-19 and the cardiovascular system. Nat Rev Cardiol. 2020;17(5):259–60.