




Heg.IA: an intelligent system to support diagnosis of Covid-19 based on blood tests

Valter Augusto de Freitas Barbosa¹ · Juliana Carneiro Gomes² · Máira Araújo de Santana² · Jeniffer E. de A. Albuquerque¹ · Rodrigo Gomes de Souza¹ · Ricardo Emmanuel de Souza¹ · Wellington Pinheiro dos Santos¹ 

Received: 27 May 2020 / Accepted: 17 November 2020 / Published online: 7 January 2021

© Sociedade Brasileira de Engenharia Biomedica 2021

Abstract

Purpose A new kind of coronavirus, the SARS-CoV-2, started the biggest pandemic of the century. More than a million people have been killed by Covid-19. Because of this, quick and precise diagnosis test is necessary. The current gold standard is the RT-PCR with DNA sequencing and identification, but its results take too long to be available. Tests base on IgM/IgG antibodies have been used, but their sensitivity and specificity may be very low. Many studies have been demonstrating the Covid-19 impact on hematological parameters.

Method This work proposes an intelligent system to support Covid-19 diagnosis based on blood testing. Laboratory parameters obtained from the hemogram and biochemical tests defined as standards to support clinical diagnosis were used as input features. Afterward, we used particle swarm optimization, evolutionary algorithms, and manual selection based on cost minimization to select the most significant features.

Results We tested several machine learning methods, and we achieved high classification performance: overall accuracy of $95.159\% \pm 0.693$, kappa index of 0.903 ± 0.014 , sensitivity of 0.968 ± 0.007 , precision of 0.938 ± 0.010 , and specificity of 0.936 ± 0.011 . These results were achieved using classical and low computational cost classifiers, with Bayes Network being the best of them. In addition, only 24 blood tests were needed.

Conclusion This points to the possibility of a new rapid test with low cost. The desktop version of the system is fully functional and available for free use.

Keywords Covid-19 · Blood tests · Software-based rapid test · Computer-aided diagnosis · Machine learning for diagnosis · Covid-19 rapid test

Introduction

A new kind of coronavirus started the biggest pandemic of the century. The virus SARSCoV-2 is related to severe acute

respiratory syndrome coronavirus (Lin et al. 2020; Okba et al. 2020), one of the most dangerous consequences of Coronavirus Disease 2019 (Covid-19). The clinical symptoms of Covid-19 patients include fever, cough, sore throat, and even gastrointestinal infection symptoms in some cases (Guo et al. 2020b). In the most severe cases, it can lead to shortness of breath and death.

SARS-Cov2 probably originated from a pangolin virus called Pangolin-CoV. They are 91.02% identical in the whole-genome level (Zhang et al. 2020b). The first record of the SARS-CoV-2 is from Wuhan, China, in later December 2019. In May 2020, it has already spread to 215 countries in the world (WHO 2020). The virus has a stronger human-to-human transmission capacity. Until May 2020, more than 3 million people were infected. Until September 2020, the virus has infected more than 30 million people, causing almost one

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

Juliana Carneiro Gomes
jcg@ecom.poli.br

Máira Araújo de Santana
mas2@ecom.poli.br

¹ Federal University of Pernambuco, Recife, Brazil

² Polytechnique School of the University of Pernambuco, Recife, Brazil

million deaths. Covid-19 is present in almost all countries in the world. In this sense, quick and precise diagnosis tests are necessary (Beeching et al. 2020). The ground-truth test in Covid-19 diagnosis is the reverse transcription polymerase chain reaction (RT-PCR) with DNA sequencing and identification (Döhla et al. 2020). However, its result needs several hours to be available (Döhla et al. 2020).

Tests based on IgM/IgG antibodies but nonspecific for Covid-19 may have very low sensitivity and specificity, depending on viral charge (Burog et al. 2020; Döhla et al. 2020; Egner et al. 2020; Patel et al. 2020; Tang et al. 2020b). IgM/IgG tests do not directly detect the SARS-Cov2 presence; indeed, they detect the serological evidence of recent viral infection. According to Li et al. (2020), positive response could come from antibodies of other coronaviruses and common flu viruses. The use of IgM/IgG rapid test kits as a definitive diagnosis of Covid-19 in currently symptomatic patients is not recommended (Burog et al. 2020) because tests in samples collected in the first week of illness have only 18.8% sensitivity and 77.8% of (Liu et al. 2020b). Döhla et al. (2020) compared the results of IgM/IgG with reverse transcription polymerase chain reaction (RT-PCR) in 59 patients and IgM/IgG tests achieved 36.4% sensitivity and 88.9% specificity.

IgG/IgM tests performed during the second week of the disease have 100% sensitivity and 50% specificity (Liu et al. 2020b). That means, when the viral charge is high, IgG/IgM tests reach high sensitivities and specificities. However, in those cases, the disease is considerably severe (Guo et al. 2020a; Hoffman et al. 2020).

Many works have been demonstrating the Covid-19 impact on the cardiovascular system and hematological characteristics (Fan et al. 2020; Gao et al. 2020; Guncar et al. 2018; Liu et al. 2020a; Tan et al. 2020; Zheng et al. 2020). According to Turner et al. (Turner et al. 2004), the coronaviruses, like SARS-CoV and SARS-Cov2, have as functional receptor the zinc metalloproteinase angiotensin-converting enzyme 2 (ACE2). This enzyme is present in cell membranes of arteries, heart, lungs, and other organs. ACE2 is implicated in heart function, hypertension, and diabetes. Considering this fact, Zheng et al. (Zheng et al. 2020) pointed out that Middle East respiratory syndrome-related coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV) can cause acute myocarditis and heart failure. Some of the coronaviruses' impacts on the cardiovascular system are the increase in blood pressure and augment of troponin I (hs-cTnI) levels (Zheng et al. 2020). Other works found that patients with Covid-19 can develop lymphopenia (low level of lymphocytes in the blood) (Fan et al. 2020; Liu et al. 2020a; Tan et al. 2020) and leukopenia (few white blood cells). They can also have decreased hemoglobin levels, absolute lymphocyte count (ALC), and absolute monocyte count (AMC) (Fan et al. 2020). Patients that developed severe forms of the disease have significantly higher levels of hematological

characteristics as interleukin-6 and D-dimer than patients that developed moderate forms of Covid-19 (Gao et al. 2020). Therefore, as Covid-19 is a disease that affects the blood, hematological exams can be used in diagnostic tests.

On the other hand, machine learning techniques have been applied to disease diagnosis by analyzing hematological parameters (Guncar et al. 2018; Luo et al. 2016; Tanner et al. 2008). Machine learning is a subarea of artificial intelligence in computer science. One of the main goals of machine learning techniques is to perform pattern detection in databases.

In this way, it is important to develop a diagnosis support system able to provide a quick result and with high sensibility even for early stages of the disease. In this context, the use of blood tests has some advantages. First, they are commonly used during medical screening. Second, they do not take too long to be ready. They also have low cost when compared to other diagnostic methods.

In this work, we propose an intelligent system to support Covid-19 diagnosis based on machine learning techniques. This pioneering system employs blood tests as inputs of the specialist system to aid in the diagnosis of Covid-19. We used the database provided by Hospital Israelita Albert Einstein, São Paulo, Brazil. This database incorporates data from patients in hospitalization condition. All subjects were experiencing at least two characteristic symptoms of Covid-19. Thus, patients are not healthy or in normal health conditions. For this reason, Heg.IA classification seeks to distinguish between patients with and without Covid-19, although they may be affected by other respiratory and general viral diseases. Therefore, it is important to mention that this system does not focus on asymptomatic patients or those who have already recovered from the disease.

The database has data from 5644 patients. In total, 559 among these patients were Covid-19 positive. This diagnosis was performed using the gold standard method, reverse transcription polymerase chain reaction (RT-PCR) (Kaggle 2020). More than one hundred laboratory exams are available in the database, such as blood counts and urine test. Then, aiming to provide a quick diagnosis, we performed the reduction of exams by choosing which tests are most relevant to the diagnosis. First of all, we performed selection using attribute selection algorithms based on evolutionary search (ES) (Kim et al. 2000; Liang et al. 2000) and particle swarm optimization (PSO) (Kennedy and Eberhart 1995; Poli et al. 2007; Wang et al. 2007). In the following, we made a feature reduction by hand, to find a set of exams that could minimize price, turnaround time, and amount of blood samples. Finally, we obtained a database with 24 blood tests. According to empirical information,¹ these exams may have a turnaround time of 1 to

¹ According to the Brazilian Society of Clinical Analysis, there is no reference document to establish standard turnaround times for the exams described above.

2 h in an emergency context. Consequently, our proposal can be classified as a software-based rapid test.

We performed several experiments with different machine learning techniques as multilayer perceptron (Haykin 2001; Rosenblatt 1958), support vector machine (Boser et al. 1992; Cortes and Vapnik 1995), random trees, random forest (Breiman 2001; Geurts et al. 2006), Bayesian networks, and naive Bayes (Cheng and Greiner 2001; Haykin 2001) and evaluated their results. Experiments show that the Bayesian networks have superior performance with respect to other techniques. In order to choose this method, we considered six parameters: accuracy, sensibility, specificity, precision, recall, and kappa index.

This paper is organized as follows: the “Related works” section discusses some related works that found that Covid-19 can affect hematological characteristics. The “Materials and methods” section presents the Kaggle database and reviews the theoretical machine learning method concepts necessary to understand this work. It is also described how we manipulated the Kaggle database and how we performed the experiments. The “results” section shows and analyzes the experimental results and the resulting desktop application we developed. We finalized the paper in the “Conclusion” section.

Related works

Several studies have shown the importance of blood tests for diagnosis and indicative of the degree of severity of Covid-19. Fan et al. (2020) analyzed hematological indexes of 69 patients with Covid-19. All of them were treated at the National Center for Infectious Diseases (NCID), located in Singapore. Among these patients, 65 underwent a complete blood count (CBC) on the day of admission. In addition, demographic information such as age, gender, ethnicity, and location was made available for this research. About 13.4% of patients needed intensive care unit (ICU) care, especially the elderly. During the first exams, 19 patients had leukopenia (few white blood cells) and 24 had lymphopenia (low level of lymphocytes in the blood), 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 parameters can help to identify patients who need assistance in the ICU. Also considering 9 patients who were in the ICU, the authors observed that the patients had a significant decrease in their hemoglobin levels, ALC, and absolute monocyte count (AMC) levels compared to the non-ICU group. ICU patients also tend to be neutrophilic. The platelet count did not prove to be a factor for discrimination between the types of hospitalization.

Tan et al. (2020) also looked at the complete blood count of patients (cured ones and also 12 patients who died from Covid-19). The authors sought to obtain key indicators of disease progression, as a way of supporting future clinical management decisions. In the case of patients who died, blood tests were monitored continuously throughout the treatment process. As in the previous study, the authors observed lymphopenia in this group. Based on this, the study then outlined a model (Time-LYM% model, TLM) for classifying disease severity and predicting prognosis. Thus, the blood lymphocyte percentage (LYM%) was divided into two cases, considering the first 10–12 days of symptoms: LYM% $> 20\%$ are classified as moderate cases and with a high chance of recovery. LYM% $< 20\%$ are classified as severe cases. In a second exam, 17–18 days after the first symptoms, patients with LYM% $> 20\%$ are recovering, patients with $5 < \text{LYM}\% < 20\%$ are in danger, and LYM% $< 5\%$ are in critical condition. In order to validate the model, the authors evaluated 90 patients with Covid-19. The consistency between the guideline and TLM-based disease classification was verified using the kappa statistic ($\text{kappa} = 0.48$). These results indicate that probably LYM% should be used together with other parameters for a better evaluation of Covid-19.

Gao et al. (2020) observed hematological characteristics of 43 patients at Fuyang Second People’s Hospital. The patients had a 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. The blood tests observed were: Routine blood tests (white blood cell [WBC] count, lymphocyte count [LYM], mononuclear count [MONO], neutrophils count [NEU]) were performed on the blood samples. Blood biochemistry parameters (aspartate aminotransferase [AST], alanine aminotransferase [ALT], glucose [GLU], urea, creatinine [Cr], cystatin [Cysc], uric acid [UA], and C-reactive protein [CRP]); coagulation functions (the D-dimer [dD], thrombin time [TT], prothrombin time [PT], fibrinogen [FIB], activated partial thromboplastin time [APTT]); procalcitonin (PCT); and interleukin6 (IL6). Using statistical tests, the study noted that the levels of GLU, CRP, IL-6, TT, FIB, and d-D 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 an AUC of 0.840. The combination also achieved a specificity of 96.4% and sensitivity of 93.3%, using tandem and parallel testing, respectively. These results indicate that patients with severe conditions would have abnormal coagulation.

Liu et al. (2020a) 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 diagnosed with Covid-19 confirmed by throat-swab specimens analyzed with RT-PCR. The patients were treated at Wuhan Union Hospital in January 2020. The information provided was epidemiological, demographic, and 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 $\text{PaO}_2/\text{FiO}_2 < 300$ mmHg. In the first group, 27 patients were classified, while 13 were classified in the second. The study reported that the 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. Furthermore, most severe patients presented lymphopenia that can be related to the significantly decreased absolute counts of T cells, especially CD8+ T cells, while white blood cell and neutrophil 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 parameters can be essential to facilitate 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.

Guncar et al. [24] proposed a system based on machine learning for analyzing blood tests and predicting hematological diseases. The models were developed with data from the University Medical Center of Ljubljana, which were collected between the years 2005 and 2015. In this case, 43 diseases and 181 parameters or attributes were selected to generate a first model (SBA-HEM181). In addition, a second model with 61 parameters was also developed (SBA-HEM061). The selection of the smaller group was made according to the frequency of use. Regarding the missing values (about 75%), the authors filled in with median values for each attribute. For the development of the intelligent system, classic classifiers such as support vector machines, naive Bayes, and random forest were tested, with the latter being chosen. The simulations were repeated 10 times using 10-fold 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. These results pointed to the possibility of detecting diseases through blood tests using classic classifiers.

Materials and methods

In this work, we explore, evaluate, and analyze the performance of several machine learning techniques to predict Covid-19 diagnosis based on different clinical exam configurations. In addition, we also propose a specialist system to aid Covid-19 diagnosis and present a desktop-based prototype.

Proposed method

In this work, we propose a support system for Covid-19 detection using blood exams called Heg.IA. It is a smart system pioneered in the rapid detection of Covid-19 through blood tests. The basic operation of this system is that in emergency situations, the medical team should order some blood exams of patients that have characteristic symptoms of Covid-19; with the results in hand, the healthcare professional could insert them in a desktop system. In the following, the system will inform whether the patient is infected with Covid-19 or not. The result will be presented with the respective classification probabilities, helping the experts in making their final decision. Figure 1 shows a general schematic of the solution.

Database

For the development of this project, we used the database provided by Hospital Israelita Albert Einstein located in São Paulo, Brazil. The database was made available in the competition format (Kaggle 2020). The database contains the results of more than one hundred laboratory exams, such as blood count, tests for the presence of viruses such as influenza A, and urine tests, of 5644 patients. Among these patients, 559 of them are infected with SARS-Cov2. The Covid-19 diagnoses were made with reverse transcription polymerase chain reaction (RT-PCR) with DNA sequencing and using additional laboratory tests during a visit to the hospital (Kaggle 2020). The clinical data provided in the data were normalized to have a mean value equals to zero and a standard deviation equals to one (Kaggle 2020).

Database pre-processing

In order to use the provided data as input parameters for a machine learning method, it was necessary to assign integer numeric values to the columns containing categorical data. These columns have exam results for detecting whether a pathogen is present or not from the appearance of patient's urine, for example. Besides, considering that 90.1% of data is from patients without Covid-19, it is reasonable to assume that missing values are within the normal range for a hospitalized condition.

In this way, for the parameters on the presence of pathogens, it was assigned the value zero (0) for pathogens' absence

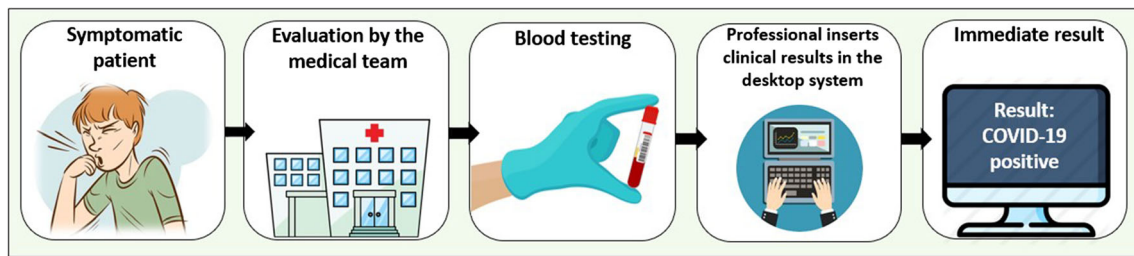


Fig. 1 General proposal to the diagnostic support system for the detection of Covid-19. Patients with symptoms of Covid-19 should go to health centers. The medical team must ask for blood tests. Then, the results can

be entered into the Heg.IA system, which will indicate if the patient is infected with SARS-Cov2

and one (1) for pathogens' presence. Thus, for the attribute with SARS-CoV-2 exam result, which gives the information whether the patient has Covid-19 or not, it was assigned the value 0 for cells marked as negative and 1 for cells with positive result. Following this procedure, the attributes respiratory syncytial virus, Influenza A, influenza B, parainfluenza 1, coronavirus NL63, rhinovirus/ enterovirus, mycoplasma pneumoniae, coronavirus HKU1, parainfluenza 3; *Chlamydomphila pneumoniae*, adenovirus, parainfluenza 4, coronavirus 229E, coronavirus OC43, influenza A H1N1 2009, *Bordetella pertussis*, metapneumovirus, parainfluenza 2, influenza B evaluated with rapid test, influenza A evaluated with rapid test, and Strepto A, it was assigned value 0 for the not detected cases and 1 for abnormalities detected, and value 2 was used for missing values.

The changes made to the attributes pertaining to urine tests followed the same pattern of the pathological examinations. Thus, for the attributes urine-esterase, urine-hemoglobin, urine-bile pigments, urine-ketone bodies, urine-protein, urine-Hyaline cylinders, urine-granular cylinders, and urine-yeasts, it was put on 0 for absent cases and missing values and value 1 for the presence. In the urine aspect column, 0 was adopted for clear and missing values, 1 for lightly cloudy class, and 2 for cloudy class. In urine-crystals, the absence category and missing values were replaced by 0, amorphous urate replaced by 1, and calcium oxalate replaced by 2. The light-yellow category and missing values from urine-color were assigned as 0, yellow as 1, and orange as 2. Whereas for "urine urobilinogen," the results classified as normal were replaced by 1 and not done or missing value by 0. Lastly, the values in urine leukocytes less than 10,000 were replaced by 0 and the values greater or equal than 10,000 by 1.

Considering that the database is normalized with a mean of 0 and unitary variance, the missing values from columns containing numeric values were filled by 0. In other words, the missing values were filled with statistical expectancy, i.e., the sample average. Thus, when filling in the missing data with zeros, in this case, we estimate that these patients have parameters under the expected values of this database. Since all patients were in a hospitalization condition and experiencing at least two characteristic symptoms of Covid-19, patients are

not healthy or in normal health condition. Furthermore, to perform the classification, we estimated the reference values of the hematological parameters considering a priori probabilities calculated from the whole database itself. Thus, given that the database was made available in its normalized version with mean 0 and variance 1, the reference values used to denormalize the data do not refer to healthy patients' conditions, but to hospitalized ones.

Besides, the low number of patients with Covid-19 in the database might bias the classification to the majority class (healthy patient class). In view of that, we performed a synthetic balancing by using Synthetic Minority Oversampling Technique (SMOTE). It is an oversampling technique for generating synthetic samples from the minority class. SMOTE uses linear combinations of two similar samples to construct new data (Blagus and Lusa 2013).

After changing categorical data and handling missing values, it was possible to build up a database used to predict Covid-19 diagnosis.

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) and evolutionary search (ES) (Kim et al. 2000; Liang et al. 2000). Both use 20 individuals and 500 iterations. The goal of the attribute selection is to find which exams are more relevant for classification tasks and also to reduce the number of required exams for diagnostic support and consequently its price. Thus, the solution can be extended to public healthcare center with low costs.

We chose to use PSO and ES algorithms, since they are well-established methods for feature selection (Eberhart and Shi 2001; Feng et al. 2017; Jóhannesson et al. 2002; Kennedy and Eberhart 1995; Poli et al. 2007; Shi and Eberhart 1999; van den Bergh and Engelbrecht 2006; Wang et al. 2007; Zhu et al. 2018). Both feature selection methods are widely used due to their global searchability. They use an objective function aiming to minimally penalize the classification performance, keeping it as high as possible with a reduced number

of features. In other words, we chose to use algorithms that work as decision trees, in which the features are ranked from the accuracy. PSO algorithm uses a population of randomly generated particles. In this approach, each particle corresponds to randomly generated solution and has an associated velocity and position. ES, on the other hand, is a version of genetic algorithms, which emulates the Darwinism principle of natural selection and genetic reproduction. Thus, this method uses the evolution principles to assess the best solution to a given problem.

The feature selection implementation resulted in three databases: the original database with 108 attributes and SARS-CoV-2 (PSO) and SARS-CoV-2 (ES) with 63 and 62 attributes, respectively. The acronyms PSO and ES in database names indicate that the database was obtained by attribute selection using PSO and ES respectively.

After selecting attributes with PSO and ES, a last selection was manually performed. We considered the availability of these tests in the Brazilian public health system. In addition, we assessed both cost and time of conducting and processing these exams. To build this fourth database, we removed the tests one by one and evaluated the classification performance at each step. Thus, it was possible to significantly decrease the number of attributes with minor impacts on the performance of the system: we removed 39 exams. Finally, a fourth database was generated with only 24 attributes. The base was named SARS-CoV-2 (Cheap Exams). This complete process is described in the diagram in Fig. 2.

With regard to turnaround time, the results of all 63 exams selected with PSO can be obtained between 8 and 24 h in the best scenario. On the other hand, the 24 manually selected exams can be obtained in 1 to 2 h under ICU conditions. Considering the costs, we were able to reduce the average total cost of taking these exams from US\$ 330.92 to about US\$ 25.22, i.e., just 7.62% of the original cost. These values are detailed in the table in Fig. 3. In the table, it is possible to observe the values of the list of exams selected by the PSO (green), and the values of the final exam list, selected manually (orange). The indicated prices are average estimate in the context of Brazilian hospitals, presented in Real (R\$) and in Dollar (US\$), considering the current quote at the time of writing this work.

Classification

Multilayer perceptron

In 1958, Rosenblatt (1958) proposed the perceptron model for supervised learning. Rosenblatt was a psychologist and one of the pioneers in the concept of artificial neural networks. His perceptron model is the simplest form of neural network used for binary classifications of linearly separable patterns. It

consists of a single neuron with adjustable synaptic weights and a bias (Haykin 2001).

Multilayer perceptron (MLP) is a generalization of the single-layer perceptron. It consists of a feed-forward network with an input layer, hidden layers, and one output layer. The addition of hidden layers allows the network the ability to classify more complex problems than single-layer perceptron such as image classification (Barbosa et al. 2020; Lerner et al. 1994; Phung et al. 2005).

MLP training is performed in a supervised way, through the error-correction learning rule. This rule adjusts the synaptic weights in a way that the network output becomes closer to the expected output. This method usually uses an error backpropagation algorithm to adjust these weights. Based on the gradient descent algorithm, backpropagation proceeds in two phases: propagation and backpropagation. In the propagation phase, an output is obtained for a given input pattern. In the backpropagation phase, an error is calculated using the desired output and the output obtained in the previous phase. Then, the error is used to update the connection weights. Thus, backpropagation aims to iteratively minimize the error between the network output obtained and the desired output (Haykin 2001).

MLPs and other artificial neural networks architectures like extreme learning machines have been commonly used in support diagnosis applications, e.g., 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; Santana et al. 2018), mammography images (Cordeiro et al. 2016; Cordeiro et al. 2017; Cruz et al. 2018; de Lima et al. 2014; de Lima et al. 2016; Lima et al. 2015; Silva et al. 2020), and multiple sclerosis diagnosis support (Commowick et al. 2018).

Support vector machines

Created by Vladimir Vapnik and Alexey Chervonenkis (Boser et al. 1992; Cortes and Vapnik 1995), in 1963, the support vector machine (SVM) performs a nonlinear mapping on the database in a space of high dimension called feature space. So the algorithm builds a linear decision surface, called hyperplane, to separate distinct classes (Cortes and Vapnik 1995). SVMs have been widely used in several medical applications, e.g., breast cancer diagnosis over breast thermography (de Vasconcelos et al. 2018; Pereira et al. 2020b; Santana et al. 2020) and mammography images (Cordeiro et al. 2016; Cordeiro et al. 2017; Cruz et al. 2018; de Lima et al. 2014; de Lima et al. 2016; Lima et al. 2015; Silva et al. 2020).

Thus, the training process of a support vector machine aims to find the hyperplane equation which maximizes the distance between it and the nearest data point. That hyperplane is called an optimal hyperplane (Haykin 2001). This machine uses

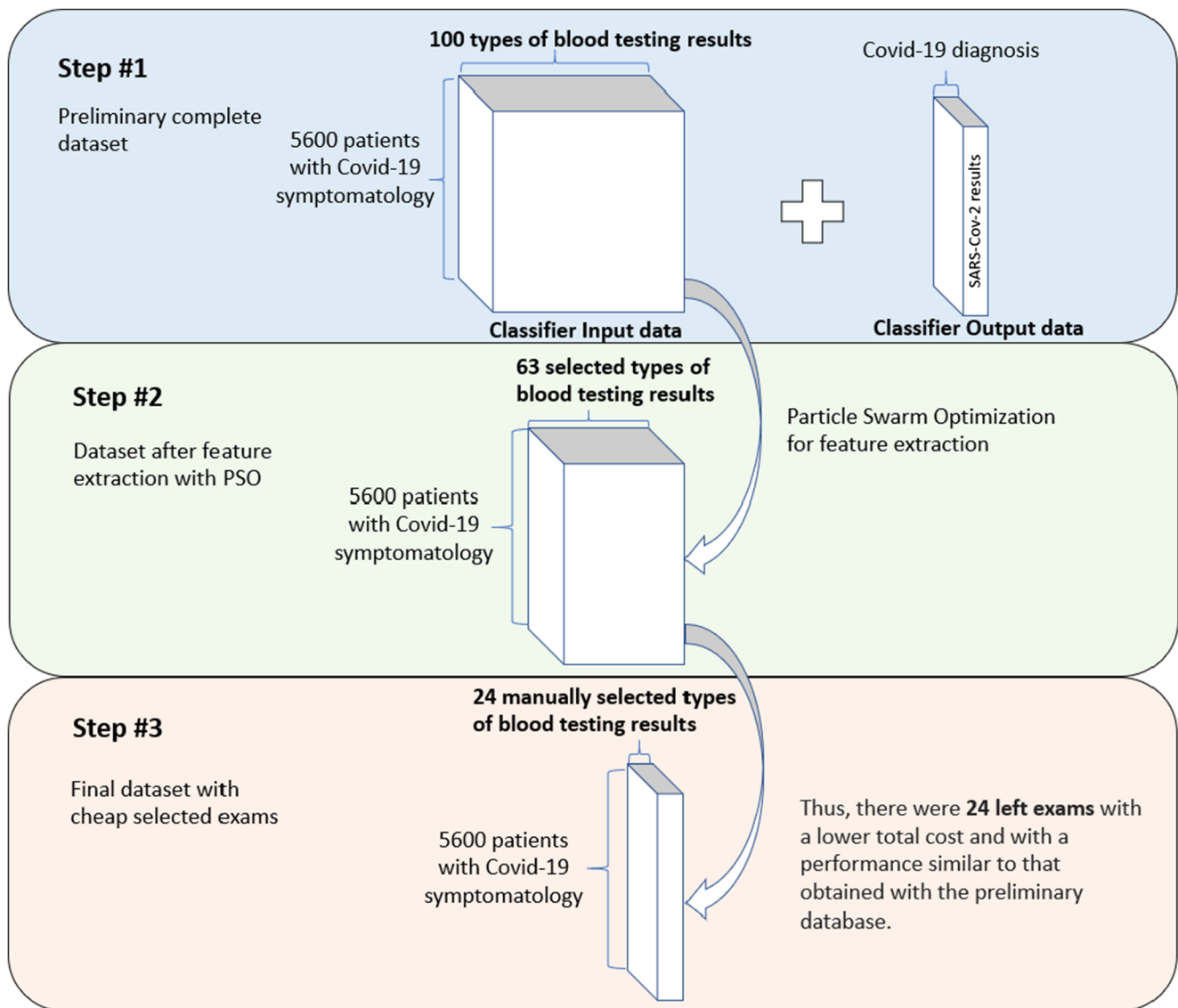


Fig. 2 Description of the feature extraction process: PSO and ES algorithms were applied to the original database, resulting in 63 attributes. In the following, 24 attributes were selected manually with a

lower total cost and with a performance similar to the preliminary database. We manually removed 39 exams

support vector learning, which is a subset of the training data. SVM algorithm is well known for its ability to provide good generalization performance (Haykin 2001). However, this performance may decrease by increasing the complexity of the hyperplane. The type of machine varies with the type of kernel used to build the optimal hyperplane. In this work, we employed two kernel configurations: polynomial and radial basis function (RBF) kernels, described as following:

$$K(x, y) = (x \cdot y + 1)^E, \quad (1)$$

$$K(x, y) = \exp(-\gamma(x-y) \cdot (x-y)), \quad (2)$$

where E is the polynomial degree; \mathbf{x} and \mathbf{y} are the input vectors, corresponding to the input feature vector and the weights vector.

Decision trees

Decision trees are a type of supervised machine learning model. They can be used in classification and regression problems. In most cases, trees have nodes responsible to store information. Essentially, in a tree, there are four types of nodes: root, leaf, parent, and child. The starting point is the root node that has the highest hierarchical level. One node may connect to another, establishing a parent-child relationship, in which a parent node generates a child node. The terminal nodes are composed of leaf nodes. Thus, they have no children and represent a decision. In this way, using such trees, the algorithm makes a decision after following a path that starts from the root node and reaches a leaf node. There are several types

Fig. 3 Details of the prices of each exam in Brazilian Reais and Dollars. The table shows a considerable reduction in costs for exams. In the green table, it is possible to observe the tests selected by the PSO method and their respective prices. The orange list shows the exams on the final list. It is important to highlight that several hematological parameters are embedded in the complete blood count, urine, and arterial and venous gasometry

Exam	Price (R\$)	Price (U\$)
Hemogram	60.00	11.52
Urine	20.00	3.84
Arterial gasometry	60.00	11.52
Venous gasometry	60.00	11.52
International normalized ratio (INR)	9.00	1.73
Lactic Dehydrogenase	8.70	1.67
Vitamin B12	24.65	4.73
Creatine phosphokinase CPK	14.50	2.78
Ferritin	36.00	6.91
Arterial Lactic Acid	15.00	2.88
Lipase dosage	5.63	1.08
Albumin	4.35	0.83
Influenza A	120.00	23.03
Influenza B rapid test	120.00	23.03
Influenza A rapid test	120.00	23.03
CoronavirusNL63	90.00	17.27
Coronavirus HKU1	90.00	17.27
Chlamydomphila pneumoniae	404.04	77.55
Adenovirus	39.00	7.49
Coronavirus229E	165.00	31.67
CoronavirusOC43	90.00	17.27
Bordetella pertussis	90.00	17.27
Urea	7.30	1.40
Proteina C reativa	19.43	3.73
Magnesium	4.64	0.89
Phosphor	6.00	1.15
Alanine transaminase	4.06	0.78
Aspartate transaminase	7.00	1.34
Gammaglutamyltransferase	5.20	1.00
Direct Bilirubin	4.35	0.83
Alkaline phosphatase	5.22	1.00
Strepto A	15.00	2.88
Total:	R\$ 1724.07	US 330.92

Exam	Price (R\$)	Price (U\$)
Hemogram	60.00	11.52
International normalized ratio (INR)	9.00	1.73
Lactic Dehydrogenase	8.70	1.67
Lipase dosage	5.63	1.08
Albumin	4.35	0.83
Urea	7.30	1.40
Alkaline phosphatase	5.22	1.00
Magnesium	4.64	0.89
Phospor	6.00	1.15
Alanine transaminase	4.06	0.78
Aspartate transaminase	7.00	1.34
Gammaglutamyltransferase	5.20	1.00
Direct Bilirubin	4.35	0.83
Total:	RS 131.45	U\$ 25.22

of decision trees. They usually differ from the way the method goes through the tree structure. Random tree and random forest methods are the most main ones.

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

On the other hand, the random forest algorithm consists of a collection of trees. These trees hierarchically divide the data, in a way that each tree votes for a class of the problem. In the

end, the algorithm chooses the most voted class as the prediction of the classifier (Breiman 2001).

Bayesian network and naive Bayes

Bayes net and naive Bayes are classifiers based on Bayes' decision theory. They are also called the test procedure by the Bayes hypothesis. Bayesian classifiers seek to find a minimum mean risk. By considering a set of correct decisions and a set of incorrect decisions, they use conditional probability to

create the data model. The product of the frequency of each decision with the cost involved in making the decision results in the weights (Haykin 2001). Bayesian networks behave like a linear classifier for a Gaussian distribution. Its behavior is comparable to that of a single layer perceptron.

The standard Bayes network algorithm assesses the probability of occurrence of a class from the values given by the others. Thus, this method assumes dependence between the features. Naive Bayes, in turn, considers that all features are independent of each other, being only connected to the class (Cheng and Greiner 2001). This method does not allow dependency between features. Considering this assumption represents an unrealistic condition, the algorithm is considered “naive.”

Parameter settings of the classifiers

The experiments were made by using the following techniques: SVM with polynomial kernel of degree (E) 1, 2, and 3 and RBF kernel with γ of 0.01; MLP with 50 and 100 neurons in hidden layer; random forest with 10, 20, 30, ..., 100 trees; random tree; naive Bayes; Bayesian network. Each configuration was executed 25 times to assess statistical information. As the evaluation method, we chose to perform a 10-fold cross-validation, essentially resulting in 90% of instances for training and 10% for validation. We used 9155 training instances and 1017 validation instances.

Metrics

We chose six metrics to evaluate the performance of diagnostic tests: accuracy, precision, sensitivity, specificity, recall, and precision. 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 recall and sensitivity metrics can be calculated mathematically in the same way. They are the rate of true positives and indicate the classifier ability to detect correctly people with Covid-19. However, they are commonly used in different contexts. In the machine learning context, the term recall is common. However, in the medical world, the use of the sensitivity metric is more frequent. Precision, on the other hand, is the fraction of the positive predictions that are actually positive. Specificity is the capacity of classifying healthy patients as negatives. It is the rate of true negatives. Finally, the Kappa index is a very good measure that can handle very well both multi-class and imbalanced class problems, like the one proposed here. These six metrics allow to discriminate between the target condition and health, in addition to quantifying the diagnostic exactitude (Borges 2016). They can be calculated according to the equations in Table 1.

In Table 1, TP is the true positives, TN is the true negatives, FP is the false positives, and FN is the false negatives, ρ_o is observed agreement or accuracy, and ρ_e is the expected agreement, defined as the following:

$$\rho_e = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + FP + FN + TN)^2} \quad (3)$$

Results

Feature selection

During the feature selection procedure, we first performed selections based on PSO and ES algorithms. The features selected by both methods resulted in equivalent classification performances, as described in Table 2. Considering the values of the mean and standard deviation of the six metrics, the two methods presented similar performance: mean accuracy of 99.0865, mean kappa index of 0.9817, sensitivity of 0.9938, mean precision of 0.9880, and mean specificity of 0.9880. In addition, most selected exams were the same. Thus, we chose to select manually the cheaper exams from the ones selected by PSO algorithm. Figure 4 details the list of exams for each situation. In the first list of Fig. 4 (in blue) are all 108 exams from the original database. The second (in green) shows the 63 most relevant exams selected by the PSO algorithm. In the third list, we present the final set of 24 exams.

Classification

This section presents the results obtained in the classification phase. Tables 3 and 4 show the results of mean and standard deviation for all studied classifiers. Table 3 presents the results for the database with the complete set of exams. Table 4 shows the results for the reduced database, with 24 cheaper exams.

Figures 5 and 6 show the performance of the different methods for diagnosing SARSCoV-2 using only the 24 cheapest exams. Figure 5 presents the results of the accuracy of these methods, while Fig. 6 shows the results of kappa index. Both graphs contain statistical information from the 25 repetitions performed for each configuration. As previously mentioned, 10-fold cross-validation was used in all tests.

From the results in Figs. 5 and 6, it is possible to see that the Bayes net method overcame the others. Therefore, in Table 2, we compare the results of this method to the complete database and the two databases with a reduced amount of exams.

Table 1 Metrics used to evaluate classifiers performance: overall accuracy, sensitivity (recall), precision, specificity, and kappa index

Metrics	Equation
Accuracy	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity and recall	$Sensitivity = Recall = \frac{TP}{TP+FN}$
Precision	$Precision = \frac{TP}{TP+FP}$
Specificity	$Specificity = \frac{TN}{TN+FP}$
Kappa index	$\kappa = \frac{p_0 - p_c}{1 - p_c}$

Heg.IA desktop application

The prototype of the developed system is fully functional in a desktop version. It works like this: The health professional will be able to type the test results in the application. In this first version of the system, the exams must be entered in the unit indicated on the screens. Figures 7 and 8 show the list of exams with their respective fields. Figure 7 shows general exams, while Fig. 8 exams belonging to the blood count. After completing the exams, the professional can select the “predict” option. Although the system works more efficiently with all fields filled in, it will also provide predictions with missing exams. Finally, it is possible to view the screen in Fig. 9, where the positive or negative diagnosis for Covid-19 is presented. In addition to the diagnosis, the screen presents the values of the main classification metrics, helping the professional in his decision-making. On the screen, it is also possible to view results to predict hospitalization. This information is nearing completion, and the full system will be presented in the near future.

Knowing that the classification method was trained with a normalized database, our desktop application rescales the data to fit with this data. The rescaling is made considering the reference values of each exam. So we could estimate each new value according to the normalization.

The code for this desktop version is freely available for non-commercial purposes on Github repository: github.com/Biomedical-Computing-UFPE/Heg.IA-Desktop.

Discussion

The initial list of blood tests is quite extensive. In the context of researching a new disease, a broader investigation is always needed. Thus, it is possible to verify the presence of other factors that may worsen the patient’s health status. Examples of this are altered glycemic indexes, chronic obstructive pulmonary diseases, systemic arterial hypertension, and infections by other groups of viruses. This investigation also makes it possible to assess the course of the disease in different groups of individuals. Despite the great relevance of this search, a group of 107 clinical examinations is very large in the context of a pandemic. They could lead to long periods of analysis. This highlights the need to apply methods to select the most important attributes.

In this way, during the feature selection phase, our method was able to reduce the amount of features (exams) in more than 20%. This was an important achievement, since it reduces the amount of required procedures. Therefore, resulting in less-expensive and less time-consuming diagnostic process.

For the classification stage, Figs. 5 and 6 show that, overall, Bayes net method overcame the others in accuracy, which was greater than 95%, and in kappa statistic, reaching more than 0.90. All tested configurations of random forest also showed good performance, with accuracies around 95% and kappa close to 0.90. Naive Bayes algorithm performed slightly worse than the others, but still achieving great results for both accuracy and kappa. SVM, MLP, and random tree classifiers, on the other hand, achieved less impressive results, with accuracies around 90% and kappa between 0.75 and 0.85. As for the data dispersion, it was slightly greater for kappa results, when compared to accuracy. However, both graphs show low dispersion, indicating good reliability of the system.

The superior results of the Bayes network may indicate or confirm the features (exams) are statistically independent or, at least, preserve a low level of interdependence. Therefore, even though all 24 selected exams showed to be relevant to diagnose SARS-CoV-2, our hypothesis is that they could exhibit some degree of independence from each other.

Table 2 Comparison of classification performance using Bayes net among the three databases: with all 107 attributes (original database), with attribute extraction using PSO, and with the 24 cheap exams. The results show that the metrics had a minimal reduction with the decrease of the attributes

Database	Accuracy (%)		Kappa statistic		Recall/Sensitivity		Precision		Specificity	
	Average	StdDev	Average	StdDev	Average	StdDev	Average	StdDev	Average	StdDev
All exams (original database)	99.0908	0.2809	0.9818	0.0056	0.9937	0.0032	0.9882	0.0049	0.9881	0.0050
Database with ES feature selection	99.0865	0.2840	0.9817	0.0057	0.9938	0.0031	0.9880	0.0050	0.9880	0.0051
Database with PSO feature selection	99.0865	0.2840	0.9817	0.0057	0.9938	0.0031	0.9880	0.0050	0.9880	0.0051
Cheap Exams (24 exams)	95.1589	0.6932	0.9032	0.0139	0.9676	0.0075	0.9377	0.0104	0.9356	0.0114

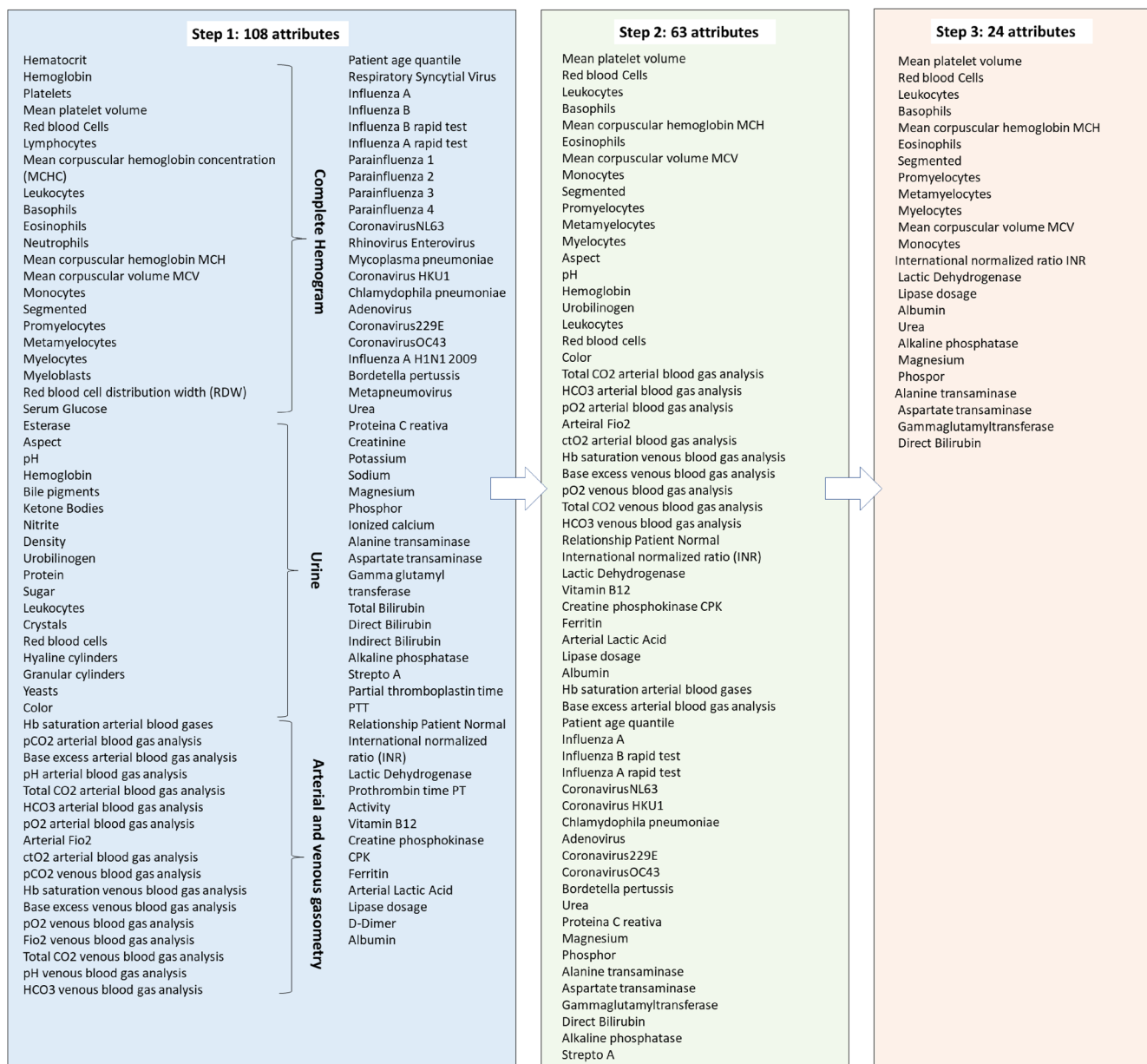


Fig. 4 Description of the attributes in each of the bases used. Initially, the original database had 108 exams or attributes. After applying the PSO method for extracting attributes, a second base was formed with 63 attributes. Finally, 24 exams were selected manually

The results presented in Table 3 show the classification performance using Bayes net among the three databases: with the complete set of exams (original database), with features selected by PSO, and the one including just the cheaper exams. The table shows the average and standard deviation values for all six metrics: accuracy, kappa statistic, recall, sensibility, precision, and specificity. By comparing the results of all databases, we observed a minimal reduction in accuracy, recall, and sensitivity with the decrease of the amount of features. All these metrics showed a decrease lower than 4%, when using only the cheap exams. For precision and specificity, the reduction for 24 exams led to a decrease of around 5%. The

selection of these exams triggered a reduction of 8% in the kappa statistic.

Thus, the 24 selected tests have been shown to be efficient in the diagnosis of Covid-19 in a fast way. Several studies have analyzed the clinical potential of these hematological parameters. Lippi and Plebani (Lippi and Plebani 2020) point out that many patients with Covid-19 do not have abnormal coagulation tests at the time of admission to the hospital. However, there is a gradual increase as the severity of the disease increases. Wan et al. (2020) also detected cases of hypercoagulable state and secondary hyperfibrinolysis. These findings confirm that parameters such as the mean volume of platelets are important in identifying severe cases and

Table 3 Results of experiments with multiple classifiers using the original database (all exams)

Classifier		SARS-Cov2 (all exams)			
		Kappa	Accuracy	Recall	Precision
SVM kernel	Mean	0.930	96.460	0.960	0.960
Polynomial E1	Std Dev	0.010	0.570	0.010	0.010
SVM kernel	Mean	0.930	96.330	0.960	0.960
Polynomial E2	Std Dev	0.010	0.580	0.010	0.010
SVM kernel	Mean	0.920	96.080	0.960	0.960
Polynomial E3	Std Dev	0.010	0.580	0.010	0.010
SVM Kernel	Mean	0.910	95.630	0.960	0.960
	Std Dev	0.010	0.620	0.010	0.010
MLP (50 neurons in the hidden layer)	Mean	0.900	95.060	0.950	0.950
	Std Dev	0.020	0.840	0.010	0.010
MLP (100 neurons in the hidden layer)	Mean	0.900	95.060	0.950	0.950
	Std Dev	0.020	0.840	0.010	0.010
Random forest (10 iterations)	Mean	0.950	97.350	0.970	0.970
	Std Dev	0.010	0.500	0.000	0.000
Random forest (20 iterations)	Mean	0.960	97.900	0.980	0.980
	Std Dev	0.010	0.430	0.000	0.000
Random forest (30 iterations)	Mean	0.960	98.060	0.980	0.980
	Std Dev	0.010	0.420	0.000	0.000
Random forest (40 iterations)	Mean	0.960	98.130	0.980	0.980
	Std Dev	0.010	0.430	0.000	0.000
Random forest (50 iterations)	Mean	0.960	98.160	0.980	0.980
	Std Dev	0.010	0.430	0.000	0.000
Random forest (60 iterations)	Mean	0.960	98.190	0.980	0.980
	Std Dev	0.010	0.430	0.000	0.000
Random forest (70 iterations)	Mean	0.960	98.200	0.980	0.980
	Std Dev	0.010	0.440	0.000	0.000
Random forest (80 iterations)	Mean	0.960	98.220	0.980	0.980
	Std Dev	0.010	0.440	0.000	0.000
Random forest (90 iterations)	Mean	0.960	98.230	0.980	0.980
	Std Dev	0.010	0.440	0.000	0.000
Random forest (100 iterations)	Mean	0.960	98.240	0.980	0.980
	Std Dev	0.010	0.450	0.000	0.000
Bayesian network	Mean	<i>0.980</i>	<i>99.090</i>	<i>0.990</i>	<i>0.990</i>
	Std Dev	<i>0.010</i>	<i>0.270</i>	<i>0.000</i>	<i>0.000</i>
Naive Bayes	Mean	0.950	97.420	0.970	0.980
	Std Dev	0.010	0.470	0.000	0.000
Random tree	Mean	0.840	91.800	0.920	0.920
	Std Dev	0.030	1.600	0.020	0.020

The emphasized texts point to the best results

predicting the risk of mortality (Lippi et al. 2020; Tang et al. 2020a).

Lippi and Plebani (Lippi and Plebani 2020) also summarized the main contributions of laboratory tests to Covid-19. Among them are those of albumin, aspartate transaminase, alanine transaminase, bilirubin, lactate dehydrogenase, and leukocytes. These same exams were

selected by the methods in our work, confirming the relevance of the results obtained. Some other studies also showed cases of leukopenia and lymphopenia (Fan et al. 2020; Huang et al. 2020; Liu et al. 2020a; Tan et al. 2020). This may indicate a decreased immunological response to the virus. On the other hand, the potential clinical significance of the increased value of

Table 4 Results of experiments with multiple classifiers using the database with 24 attributes (Cheap Exams)

Classifier		SARS-Cov2 with Cheap Exams (24 selected)			
		Kappa	Accuracy	Recall	Precision
SVM kernel	Mean	0.820	91.010	0.960	0.870
Polynomial E1	Std Dev	0.020	0.920	0.010	0.010
SVM kernel	Mean	0.820	91.120	0.970	0.870
Polynomial E2	Std Dev	0.020	0.870	0.010	0.010
SVM kernel	Mean	0.800	89.840	0.970	0.850
Polynomial E3	Std Dev	0.020	0.960	0.010	0.010
SVM kernel RBF	Mean	0.770	88.530	0.960	0.840
	Std Dev	0.020	0.950	0.010	0.010
MLP (50 neurons in the hidden layer)	Mean	0.820	91.200	0.940	0.890
	Std Dev	0.020	1.050	0.020	0.020
MLP (100 neurons in the hidden layer)	Mean	0.820	91.200	0.940	0.890
	Std Dev	0.020	1.050	0.020	0.020
Random forest (10 iterations)	Mean	0.880	93.770	0.970	0.910
	Std Dev	0.020	0.820	0.010	0.010
Random Forest (20 iterations)	Mean	0.880	94.040	0.970	0.910
	Std Dev	0.020	0.820	0.010	0.010
Random forest (30 iterations)	Mean	0.880	94.130	0.980	0.910
	Std Dev	0.020	0.800	0.010	0.010
Random forest (40 iterations)	Mean	0.880	94.200	0.980	0.910
	Std Dev	0.020	0.780	0.010	0.010
Random Forest (50 iterations)	Mean	0.880	94.250	0.980	0.910
	Std Dev	0.020	0.770	0.010	0.010
Random forest (60 iterations)	Mean	0.880	94.270	0.980	0.910
	Std Dev	0.020	0.790	0.010	0.010
Random forest (70 iterations)	Mean	0.880	94.290	0.980	0.920
	Std Dev	0.020	0.780	0.010	0.010
Random forest (80 iterations)	Mean	0.890	94.320	0.980	0.920
	Std Dev	0.020	0.790	0.010	0.010
Random forest (90 iterations)	Mean	0.890	94.350	0.980	0.920
	Std Dev	0.020	0.800	0.010	0.010
Random forest (100 iterations)	Mean	0.890	94.370	0.980	0.920
	Std Dev	0.020	0.790	0.010	0.010
Bayesian network	Mean	<i>0.900</i>	<i>95.160</i>	0.970	<i>0.940</i>
	Std Dev	<i>0.010</i>	<i>0.690</i>	0.010	<i>0.010</i>
Naive Bayes	Mean	0.860	92.940	<i>1.000</i>	0.880
	Std Dev	0.020	0.810	<i>0.000</i>	0.010
Random tree	Mean	0.810	90.330	0.900	0.900
	Std Dev	0.020	1.060	0.020	0.010

The emphasized texts point to the best results

aminotransferases and bilirubin aligned with low albumin values is a liver injury. Wan et al. (2020) suggest that these changes in severe patients were more obvious. Liver damage might be directly caused by the viral infection of liver cells (Zhang et al. 2020a). High levels of lactate dehydrogenase can be related to pulmonary injury and/or widespread organ damage (Lippi and

Plebani 2020). In addition, Lippi and Plebani (2020) believe that increased monocyte levels can show a severe viral infection.

These results support the idea that Covid-19 can cause damages beyond hematological and respiratory issues. Examples are multi-organ failure (MOF) and its complications and intravascular coagulopathy (Lippi

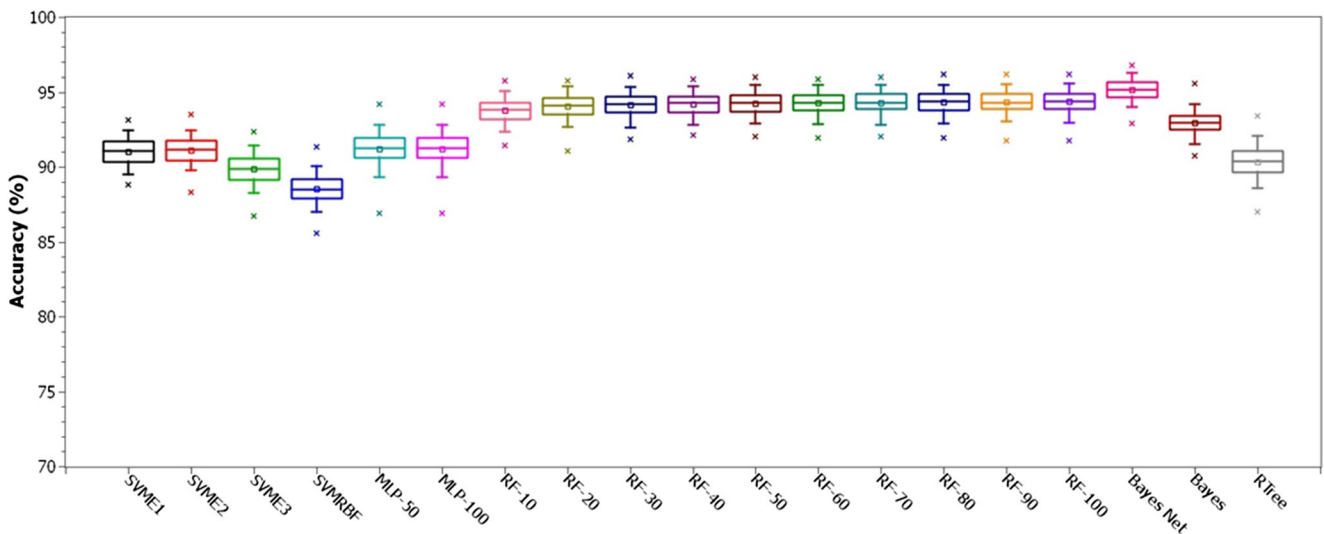


Fig. 5 Comparison of accuracy between configurations for Cheap Exams database

et al. 2020). These results highlight the importance of diagnosis based on clinical examinations.

Conclusion

A new kind of coronavirus, SARS-CoV-2, started the biggest pandemic of the century. This virus has a stronger human-to-human transmission capacity and has already led to millions of infected people and thousands of deaths. One of the main strategies to fight the pandemic is testing in a precise and quick way (WHO 2020).

The ground-truth test in Covid-19 diagnosis is the reverse transcription polymerase chain reaction (RT-PCR) with DNA sequencing and identification. RT-PCR is precise, but takes

several hours to be assessed. Another type of test, based on IgM/IgG antibodies, delivers results quickly; however, they are nonspecific for Covid-19 and may have very low sensitivity and specificity. IgM/IgG tests do not directly detect the SARS-CoV-2 presence; indeed, they detect the serological evidence of recent infection.

Considering this, the development of a diagnosis support system to provide fast results with high sensitivity and specificity is necessary and urgent. In this context, blood tests have some advantages. First, they are commonly used during medical screening. Besides that, blood tests are less expensive and less time-consuming than other diagnosis methods, thus providing a more accessible system. By combining these blood test results to analysis based on artificial intelligence, we were able to

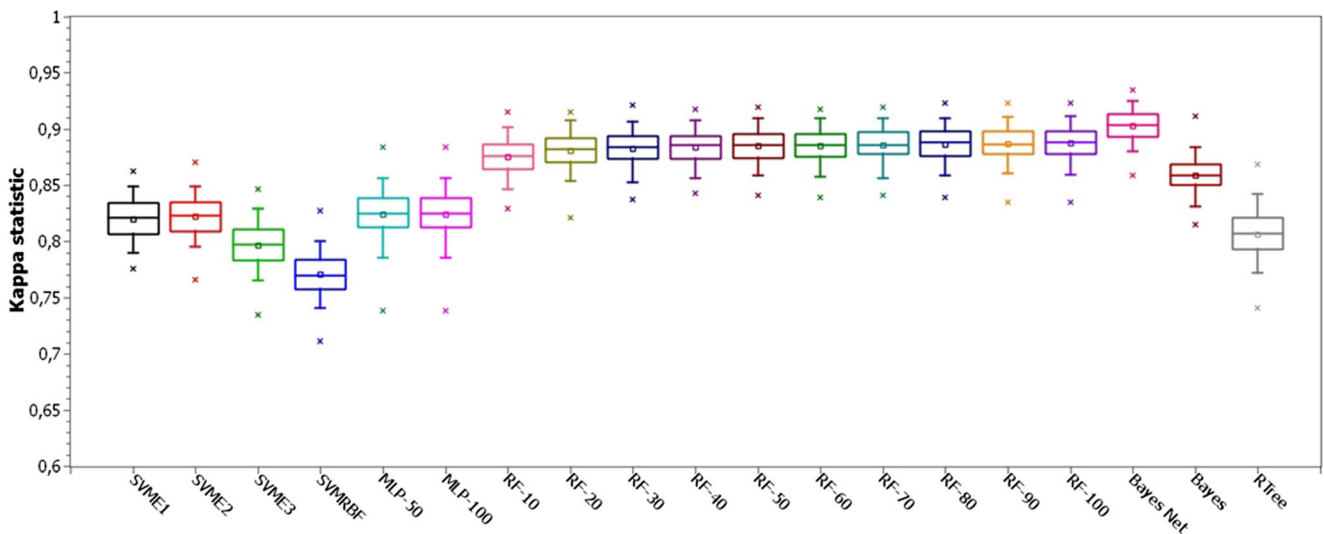
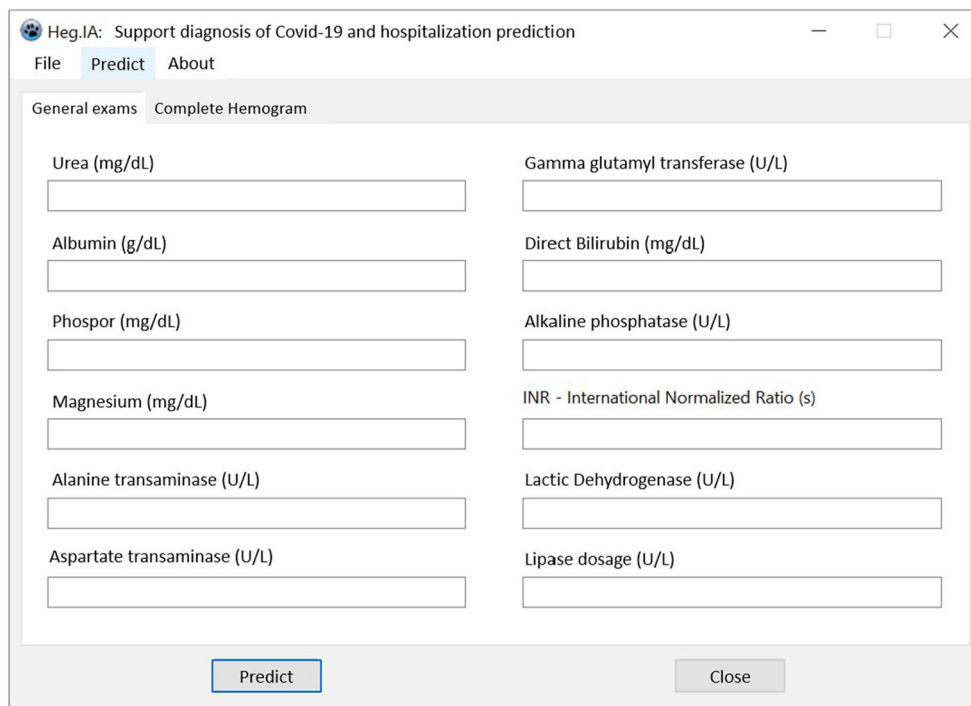


Fig. 6 Comparison of Kappa statistic between configurations for Cheap Exams database

Fig. 7 Screen 1 of the Heg.IA desktop application. General blood tests should be entered using the indicated units



provide a robust, efficient, and easily available system to diagnose Covid-19.

We optimized the system by reducing the amount of required exams, based on their relevance to describe the diagnosis problem and on their price and availability worldwide, especially in low-income communities. Firstly, we performed an automatic exam selection using

particle swarm optimization (PSO) method. Then, we manually chose some exams aiming to reach an optimal combination of price, time, and number of procedures. This procedure resulted in 24 blood tests, which can be delivered in up to 1 h. Since the computational classification can be performed in milliseconds, with the 24 blood test results, a technician can get diagnostic results relatively fast.

Fig. 8 Screen 2 of the Heg.IA desktop application. Blood tests from complete hemogram should be inserted using the indicated units

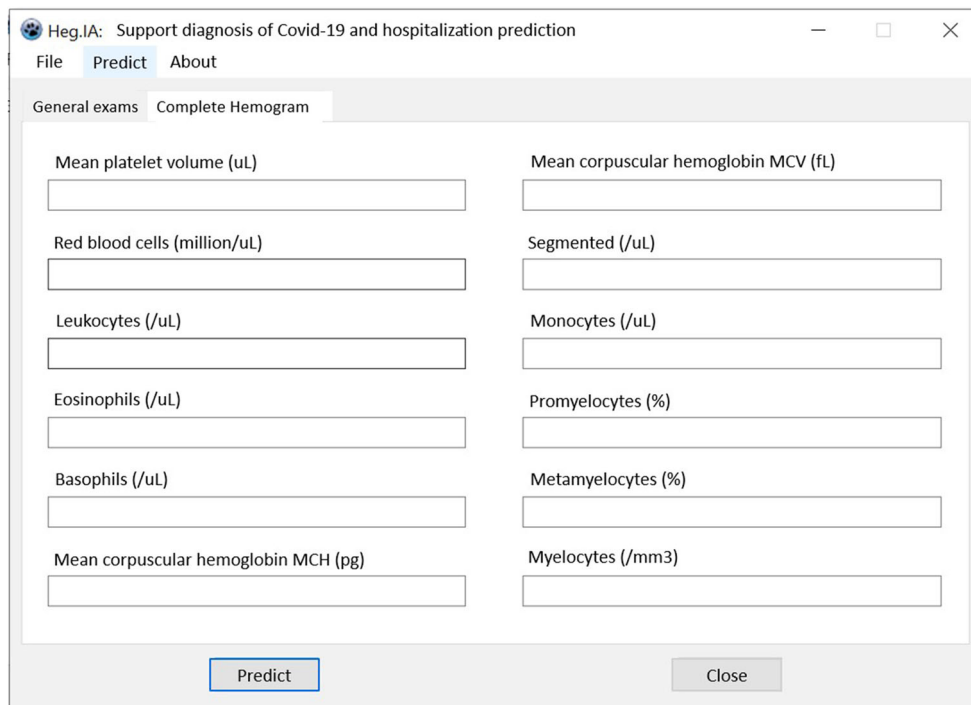


Fig. 9 Covid-19 diagnostic prediction results screen from the Heg.IA desktop application: percentage accuracy, index κ , specificity, and sensibility for Covid-19 diagnosis and hospitalization prediction

The user just has to fill in the electronic form with these 24 blood tests results.

By using a computationally simple method, based on the classical Bayesian networks, we were able to achieve high diagnosis performance: $95.159\% \pm 0.693$ of overall accuracy, kappa index of 0.903 ± 0.014 , sensitivity of 0.968 ± 0.007 , precision of 0.938 ± 0.010 , and specificity of 0.936 ± 0.011 . If compared to deep learning-based methods, the proposed system also reduces the computational cost.

More importantly, our approach provides a possible way out for the test availability issue in the context of Covid-19. The availability of this software system combined with low cost and fast tests, based on blood analysis, may be of great help to overcome the testing challenges being experienced worldwide. Less-favored countries and communities can especially benefit from this solution. With this system, we may be able to expand access to effective testing and thereby reach and save more people.

Acknowledgments The authors are grateful to the Brazilian research agencies CAPES and CNPq, for the partial financial support of this research.

Funding This study was funded by the Brazilian research agencies CAPES and CNPq.

Compliance with ethical 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 The authors declare they have no conflicts of interest.

References

- Barbosa VAF, Santana MA, Andrade MKS, Lima RCF, Santos WP. Deep-wavelet neural networks for breast cancer early diagnosis using mammary thermographies. In: Das H, Pradhan C, Dey N, editors. Deep Learning for Data Analytics: Foundations, Biomedical Applications, and Challenges. 1st ed. London: Academic Press; 2020.
- Beeching NJ, Fletcher TE, Beadsworth MBJ. Covid-19: testing times. *BMJ*. 2020;369. <https://doi.org/10.1136/bmj.m1403> <https://www.bmj.com/content/369/bmj.m1403>.
- Blagus R, Lusa L. Smote for high-dimensional class-imbalanced data. *BMC Bioinformatics*. 2013;14(1):106.
- Borges L. Medidas de acurácia diagnóstica na pesquisa cardiovascular. *Int J Cardiovascul 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. p. 144–52.
- Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
- Burog AILD, Yacapin CPRC, Maglente RRO, Macalalad-Josue AA, Uy EJB. Should IgM/IgG rapid test kit be used in the diagnosis of COVID-19? *Asia Pacific Center Evid Based Healthc*. 2020;2020(04):1–12.
- Cheng J, Greiner R. Learning Bayesian belief network classifiers: algorithms and system. *Adv Artif Intell*. 2001;2056(1):141–51.
- Commowick O, Istace A, Kain M, Laurent B, Leray F, Simon M, et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Sci Rep*. 2018;8(1):1–17.
- 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, Santos WPD, Silva-Filho AG. Analysis of supervised and semi-supervised growcut applied to segmentation of masses in

- mammography images. *Comp Methods Biomech Biomed Eng.* 2017;5(4):297–315.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* 1995;20(3):273–97.
- 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.
- de Lima SM, da Silva-Filho AG, dos Santos WP. 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): IEEE. 2014. p. 988–91.
- 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 Prog Biomed.* 2016;134:11–29.
- 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.
- Döhla M, Boesecke C, Schulte B, Diegmann C, Sib E, Richter E, et al. Rapid point-of-care testing for SARS-CoV-2 in a community screening setting shows low sensitivity. *Public Health.* 2020;182:170–2.
- Eberhart, Shi Y. Particle swarm optimization: developments, applications and resources. In: *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, vol. 1; 2001. p. 81–6.
- Egner W, Beck DCC, Davis G, Dayan C, El-shanawany T, Griffiths E, et al. Statement from RCPATHS immunology' specialty advisory committee on COVID-19/SARS CoV2 antibody evaluation. *Pathology.* 2020;04:2020.
- Fan BE, Chong VCL, Chan SSW, Lim GH, Lim KGE, Tan GB, et al. Hematologic parameters in patients with COVID19 infection. *Am J Hematol.* 2020;04:2020.
- Feng L, Ong Y, Jiang S, Gupta A. Autoencoding evolutionary search with learning across heterogeneous problems. *IEEE Trans Evol Comput.* 2017;21(5):760–72.
- Gao Y, Li T, Han M, Li X, Wu D, Xu Y, et al. Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *J Med Virol.* 2020;92:791–6.
- Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* 2006;63(1):3–42.
- Guncar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M, et al. An application of machine learning to haematological diagnosis. *Sci Rep.* 2018;8(1):1–12.
- Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis.* 2020a;03:2020.
- Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Military Med Res.* 2020b;7(1):1–10.
- Haykin S. *Neural networks: principles and practice.* Bookman. 2001;11:900.
- Hoffman T, Nissen K, Krambrich J, Rönnerberg B, Akaberi D, Esmailzadeh M, et al. Evaluation of a covid-19 igm and igg rapid test; an efficient tool for assessment of past exposure to sars-cov-2. *Infect Ecol Epidemiol.* 2020;10(1):1754538.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395(10223):497–506.
- Jóhannesson GH, Bligaard T, Ruban AV, Skriver HL, Jacobsen KW, Nørskov JK. Combined electronic structure and evolutionary search approach to materials design. *Phys Rev Lett.* Jun 2002;88:255506. <https://doi.org/10.1103/PhysRevLett.88.255506>.
- Kaggle. Diagnosis of COVID-19 and its clinical spectrum. Kaggle, 2020. <http://www.kaggle.com/einsteindata4u/covid19>. Accessed 07 Apr 2020.
- Kennedy J, Eberhart R. Particle swarm optimization. In *Proceedings of ICNN'95 International Conference on Neural Networks.* IEEE. 1995;4:1942–48.
- Kim Y, Street WN, Menczer F. Feature selection in unsupervised learning via evolutionary search. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.* 2000. p. 365–9.
- Lerner B, Levinstein M, Rosenberg B, Guterman H, Dinstein L, Romem Y. Feature selection and chromosome classification using a multi-layer perceptron neural network. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94).* IEEE. 1994;6:3540–5.
- Li Z, Yi Y, Luo X, Xiong N, Liu Y, Li S, et al. Development and clinical application of a rapid igm-igg combined antibody test for sars-cov-2 infection diagnosis. *J Med Virol.* 2020;(2020):1518–24.
- Liang K-H, Yao X, Newton C. Evolutionary search of approximated ndimensional landscapes. *Int J Knowledge Based Intell Eng Syst.* 2000;4(3):172–83.
- 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 *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society.* 2015;(2015):801–4.
- Lin D, Liu L, Zhang M, Hu Y, Yang Q, Guo J, et al. Evaluations of serological test in the diagnosis of 2019 novel coronavirus (SARS-CoV2) infections during the COVID-19 outbreak. *medRxiv.* 2020.
- Lippi G, Plebani M. Laboratory abnormalities in patients with Covid-2019 infection. *Clin Chem Lab Med (CCLM).* 2020;1(7):1131–4.
- Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clin Chim Acta.* 2020;506:145–8.
- Liu J, Li S, Liu J, Liang B, Wang X, Wang H, et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine.* 2020a;(2020):102763.
- Liu Y, Liu Y, Diao B, Ren F, Wang Y, Ding J, et al. Diagnostic indexes of a rapid IgG/IgM combined antibody test for SARS-CoV-2. *medRxiv.* 2020b.
- 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.
- Okba NM, Muller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, et al. SARS-CoV-2 specific antibody responses in covid-19 patients. *medRxiv.* 2020.
- Patel R, Babady E, Theel ES, Storch GA, Pinsky BA, George KS, et al. Report from the american society for microbiology covid-19 international summit, 23 march 2020: value of diagnostic testing for SARS-CoV-2/COVID-19, 2020.
- Pereira JMS, Santana MA, Lima RCF, Lima SML, Santos WP. Method for classification of breast lesions in thermographic images using elm classifiers. In: dos Santos WP, de Santana MA, da Silva WWA, editors. *Understanding a Cancer Diagnosis.* 1st ed. New York: Nova Science; 2020a. p. 117–32.
- Pereira JMS, Santana MA, Lima RCF, Santos WP. Lesion detection in breast thermography using machine learning algorithms without previous segmentation. In: dos Santos WP, de Santana MA, da Silva WWA, editors. *Understanding a Cancer Diagnosis.* 1st ed. New York: Nova Science; 2020b. p. 81–94.
- 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: dos Santos WP, de Santana MA, da Silva WWA, editors. *Understanding a*

- Cancer Diagnosis. 1st ed. New York: Nova Science; 2020c. p. 95–118.
- 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.
- Poli R, Kennedy J, Blackwell T. Particle swarm optimization. *Swarm Intell.* 2007;1(1):33–57.
- Rodrigues AL, de Santana MA, Azevedo WW, Bezerra RS, Barbosa VA, de Lima RC, et al. 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.
- Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev.* 1958;65(6):386.
- Santana MA d, Pereira JMS, Silva FLD, Lima NMD, Sousa FND, Arruda GMSD, et al. Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Res Biomed Eng.* 2018;34:45–53 ISSN 2446–4740.
- 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: dos Santos IWP, de Santana MA, da Silva WWA, editors. *Understanding a Cancer Diagnosis.* 1st ed. New York: Nova Science; 2020. p. 65–80.
- Shi Y, Eberhart RC. Empirical study of particle swarm optimization. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99* (Cat. No. 99TH8406). IEEE. 1999;3:1945–50.
- 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: Gandhi TK, Bhattacharyya S, De S, Konar D, Dey S, editors. *Advanced machine vision paradigms for medical image analysis.* London: Elsevier; 2020.
- Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang Y-Q, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal Transduct Target Therapy.* 2020;5(1):1–3.
- Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost.* 2020a;18:844–7.
- Tang Y-W, Schmitz JE, Persing DH, Stratton CW. The laboratory diagnosis of covid-19 infection: current issues and challenges. *J Clin Microbiol.* 2020b;3:2020.
- Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, Lai YL, 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). <https://doi.org/10.1371/journal.pntd.0000196>.
- Turner AJ, Hiscox JA, Hooper NM. Ace2: from vasopeptidase to sars virus receptor. *Trends Pharmacol Sci.* 2004;25(6):291–4.
- van den Bergh F, Engelbrecht A. A study of particle swarm optimization particle trajectories. *Inf Sci.* 2006. ISSN 0020-0255;176(8):937–71. <https://doi.org/10.1016/j.ins.2005.02.003>.
- Wan S, Xiang Y, Fang W, Zheng Y, Li B, Hu Y, et al. Clinical features and treatment of COVID-19 patients in Northeast Chongqing. *J Med Virol.* 2020;92:797–806.
- 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.
- WHO. Coronavirus disease (COVID-19) pandemic. World Health Organization, 2020. <http://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed 02 Apr 2020.
- Zhang C, Shi L, Wang F-S. Liver injury in Covid-19: management and challenges. *Lancet Gastroenterol Hepatol.* 2020a;5(5):428–30.
- Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol.* 2020b;30:1346–1351.e2.
- 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.
- Zhu Q, Samanta A, Li B, Rudd RE, Frolov T. Predicting phase behavior of grain boundaries with evolutionary search and machine learning. *Nat Commun.* 2018;9. <https://doi.org/10.1038/s41467-018-02937-2>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.