J. Health Inform. 2021 Abril-Junho; 13(2): 49-56

Artigo Original

JHI JOURNAL OF HEALTH INFORMATICS

Diagnosis of patients with blood cell count for COVID-19: An explainable artificial intelligence approach

Diagnóstico de pacientes com hemograma para COVID-19: Uma abordagem com inteligência artificial explicável

Diagnóstico de pacientes con hemograma para COVID-19: un enfoque explicable de inteligencia artificial

Kaike Wesley Reis[1], Karla Patricia Oliveira-Esquerre[2]

Keywords: Artificial Intelligence; Diagnosis; Blood Cell Count

ABSTRACT

Objective: Present an explainable artificial intelligence (AI) approach for COVID-19 diagnosis with blood cell count. **Methods:** Five AI algorithms were evaluated: Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting and eXtreme Gradient Boosting. A Bayesian optimization with 5-Fold cross-validation was used to hyper-parameters tuning. The model selection evaluated three results: cross validation performance, test set prediction performance and a backtest: performance on identifying patients negative for COVID-19, but positive for others respiratory pathologies. Shapley Additive explanations (SHAP) was used to explain the chosen model. **Results:** A Random Forest model was obtained with 77.7% F1-Score (IC95%:57.1;92.3), 85.9% AUC (IC95%:73.7;95.9), 74.4% Sensitivity (IC95%:50.0;92.1) and 97.5% Specificity (IC95%:93.6;100.0). The main features were leukocytes, platelets and eosinophils. **Conclusion:** The research highlights the importance of model interpretability, demonstrating blood cell count as a possibility for COVID-19 diagnosis. The methodological structure developed, using TRIPOD's guidelines, can be extrapolated to other pathologies.

Descritores: Inteligência Artificial; Diagnóstico; Contagem de Células Sanguíneas

RESUMO

Objetivo: Propor uma abordagem com inteligência artificial explicável para diagnóstico de COVID-19 com hemograma. **Métodos:** Cinco algoritmos de IA foram testados: Regressão Logística, Florestas Aleatórias, Máquina de Vetores de Suporte, *Gradient Boosting* e *eXtreme Gradient Boosting*. Os hiper-parâmetros foram definidos através da otimização bayesiana com validação cruzada 5-*Fold*. A seleção de modelo utilizou três resultados de desempenho para definir o melhor modelo: validação cruzada, conjunto de teste e rendimento na identificação de pacientes negativos para COVID-19, porém positivos para outras patologias respiratórias (*backtest*). Ao final, *Shapley Additive explanations* (SHAP) foi utilizado para explicar o modelo escolhido. **Resultados:** Obteve-se um modelo *Random Forest* com F1-Score de 77.7% (IC95%:57.1;92.3), AUC de 85.9% (IC95%:73.7;95.9), Sensibilidade de 74.4% (IC95%:50.0;92.1) e Especificidade de 97.5% (IC95%:93.6;100.0). As principais variáveis foram leucócitos, plaquetas e eosinófilos. **Conclusão:** A pesquisa destaca a importância da interpretabilidade do modelo, demonstrando o hemograma como uma possibilidade para diagnosticar COVID-19. A estrutura metodológica desenvolvida no estudo, utilizando as diretrizes do TRIPOD, pode ser extrapolada para detecção de outras patologias.

Descriptores: Inteligencia Artificial; Diagnóstico; Recuento de Células Sanguíneas

RESUMEN

Objetivo: Proponer un enfoque explicable de inteligencia artificial (IA) para el diagnóstico de COVID-19 con el uso de hemograma. **Métodos:** Cinco modelos de IA fueron evaluados: *Logistic Regression*, *Random Forest*, *Support Vector Machine*, *Gradient Boosting* e *eXtreme Gradient Boosting*. Los hiper-parámetros fueron definidos a través de optimización bayesiana con validación cruzada 5-*Folds*. La selección del modelo se utilizó tres resultados: rendimiento del validación cruzada, rendimento en conjunto de pruebas y el análisis de desempeño en identificación de pacientes negativos para COVID-19, pero positivos para otras patologías respiratorias (*backtest*). *Shapley Additive explanations* (SHAP) fue utilizado para explicar el modelo elegido. **Resultados:** Se obtuvo un modelo *Random Forest* con F1-Score de 77.7% (IC95%:57.1;92.3), AUC de 85.9% (IC95%:73.7;95.9), Sensibilidad de 74.4% (IC95%:50.0;92.1) y Especificidad de 97.5% (IC95%:93.6;100.0). Las principales variables fueron leucocitos, plaquetas y eosinófilos. **Conclusión:** La investigación presenta la importancia de la interpretabilidad del modelo, demostrando el uso de hemograma como posibilidad para diagnosticar COVID-19. La estructura elaborada, siguiendo las directrices de TRIPOD, puede ser extrapolar para otras patologías.

[1] *Associate Professor, Industrial Engineering Graduate Program (PEI), Federal University of Bahia – UFBA, Salvador, Bahia (BA), Brazil.*
[2] *Undergraduate student, Process Control and Automation Engineering, Federal University of Bahia – UFBA, Salvador, Bahia (BA), Brazil.*

Autor Correspondente: **Kaike Wesley Reis**
e-mail: **kaikewesley@hotmail.com**

Artigo recebido: **06/08/2020**
Aprovado: **13/04/2021**

INTRODUCTION

The pandemic scenario caused by the SARS-CoV-2 virus, known as COVID-19, is requiring unprecedented responses of exceptional intensity and agility from governments. In total, over 16 million cases with at least 660 thousands deaths were reported around the world, where the United States of America and Brazil lead these values[1]. However, to decrease and control the virus spread is necessary several measures, including population testing. This action in particular leads to quick identification, early treatment, and immediate isolation to prevent more infections[2]. Given the enormous global urgency for testing in a short period, it may be common for some countries to suffer from a lack of stock. This identify a necessity for alternative methods to identify COVID-19 presence.

In this current scenario, several papers presented artificial intelligence as a solution to predict this pathology using medical sources, mainly X-ray and Computed tomography (CT) images, applying well-established Deep Learning (DL) models[3-6]. However, the intense urge for results, lack of database information, a short time for model training and production creates doubts about their impacts, results and mainly: trustiness.

Naudé, 2020[3] points out that most of the growing number of publications reporting on using AI for diagnostic and predictive purposes with X-ray and CT images so far tend to use small, possibly biased, and mostly Chinese-based samples. Besides that, the lack of quality on datasets, most of them noisy or with outliers, can be a major problem. Maguolo, 2020[4] presents a critical evaluation of automatic detection for COVID-19, showing that DL models were learning spurious COVID-19 patterns from the dataset and thus enhancing the necessity for better protocols to construct fair train and test sets.

In a more general study about AI for COVID-19 diagnosis, Wynants, 2020[5] shows alarming results: all models described were rated at high or unclear risk of bias, mostly because of a non-representative selection of control patients. Besides that, Wynants, 2020[5] pointed out a high risk of model overfitting, vague reporting, absence of any description of the study population or intended use of the models, model predictions calibration was rarely assessed and lack of well-defined model report.

To address those issues, Wynants, 2020[5] and Schwalbe, 2020[6] recommended TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) as a guideline standard to achieve reliable, reproducible and interpretable results[6]. Besides X-ray and CT images, blood cell count can be used to address COVID-19 prediction.

Through a medical literature review, a strong relationship in changing rates of various blood components for patients with COVID-19[7-13] was found. In total, three countries are covered in these studies: China[7-10,12], Singapore[11,13] and Italy[10] where most of them highlighted an intense decrease in rates of lymphocytes[7-9,11-13] and leukocytes[7,9,11-13] for positive cases.

Studies focused on using AI to diagnose COVID-19 with blood features were also found with promising results. Brinati, 2020[14] presents a Random Forest with a sensitivity of 90% and specificity of 65%. Bao, 2020[15] shows a Support Vector Machine (SVM) with a sensitivity of 88% and a precision of 92%. Kukar, 2020[16] used an eXtreme Gradient Boosting (XGBoost) that achieved a sensitivity of 81.9% and specificity of 97.9% with an importance score. Schwab, 2020[17] presents a XGBoost with sensitivity of 75%(95CI: 67%,81%) and a specificity of 49%(95% CI: 46%,51%). Finally, Soares, 2020[18] shows an ensemble of ten trained SVM-based SMOTEBoost models with a specificity of 86%(95% CI: 85%,87%) and a sensitivity of 70%(95% CI: 67%,73%). Table 1 shows a diversity of sample sizes and countries evaluated in these studies, including two works developed with the same original dataset analyzed in this paper[17-18].

Even though most studies carried out some kind of model's interpretation, presented in Table 1, none addressed possible model's pitfalls or how a feature change could affect the prediction. Besides that, only two studies show a result using a clear 95% Confidence Interval[17-18] and a patient flow diagram[16,18], both TRIPOD guideline requirements[19].

Despite the problems[3-4,19], applications with AI are already used in the health field for diagnosis[20-21] and decision making[21], but to increase confidence in the results it is important to keep and maintain a pre-established guideline[6].

To this aim, this paper presents a well-defined AI approach, following the TRIPOD guideline to diagnose COVID-19 using blood cell count. The development and results presented here focus on model's interpretation and feature understanding, not yet seen in the literature, to improve trustiness by health professionals in model reports. In addition, all the code and results are available to the public[22] to increase reproducibility.

METHODS

The studied dataset was collected, anonymized and

Table 1 - General information from studies with AI for COVID-19 diagnosis

Paper Reference	Study sample size	Country dataset	Interpretability
Brinati, 2020[14]	279	Italy	Feature Importance Decision Tree Analysis
Bao, 2020[15]	412	China	Feature Ranking Decision Tree Analysis
Kukar, 2020[16]	5493	Slovenia	Feature Importance
Schwab, 2020[17]	5644	Brazil	Feature Importance
Soares, 2020[18]	599	Brazil	---

Reference – Developed by the authors

public available by the Hospital Israelita Albert Einstein in São Paulo, Brazil in the early months of 2020. The data collection is hosted publicly at Kaggle's platform[23] with 5,644 samples and 108 medical exams outcomes[18]*.

The dataset presents a high number of missing values with 88.1% of your records absent. Furthermore, an unbalancing problem between positive and negative COVID-19 cases was observed. To adress both issues, a four-step approach was carried out: Pre-processing, Model development, Results evaluatuon and Quantitative analysis.

Pre-processing. In this step, the study dataset was prepared.The authors performed a missing value analysis, but any kind of imputation could introduce either unreal or noisy data. Most of the columns presented over 90% of missing values. Furthermore, there was a lack of more detailed information about the patients such as gender (female and male) or the presence of some type of comorbidity.

Therefore, a complete subset of 598 samples was extracted from the original dataset including: COVID-19 exam result, Patient admission (the actual patient place in the hospital), Patient age quantile, Hematocrit, Red Blood Cells (RBC), Basophils, Eosinophils, Mean Corpuscular Volume (MCV), Leukocytes, Mean Corpuscular Hemoglobin Concentration (MCHC), Platelets, Hemoglobin, Monocytes, Mean Platelet Volume (MPV), Lymphocytes, Mean Corpuscular Hemoglobin (MCH), Red Blood Cell distribution Width (RDW).

The "patient age quantile" column had it values divided by the maximum quantile to remove scale problems compared to other numeric variables, standardized previously by the hospital. With exception of "patient age quantile", all remaining columns are related to red blood count. Table 2 presented the infectious respiratory agents/ disease utilized to create a new column called "Any respiratory disease detection", used in backtest process (explained in Results evaluation).

The created column shown at Table 2 is a binary variable where a "True" value occurs if it is detected in the patient at least one infectious respiratory agent in your sample and zero otherwise. Figure 1 summarizes the information from the selected subset considering backtest target variable.

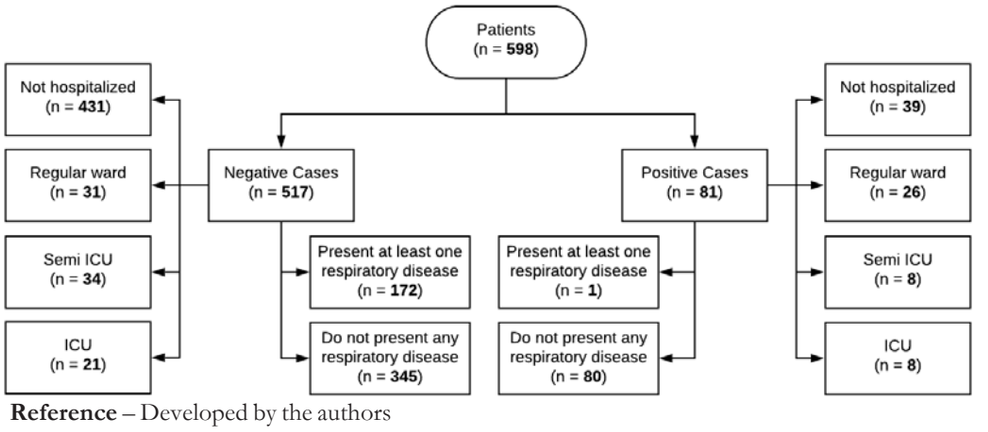
Figure 1 shows that the balancing problem for positive and negative cases persists. Besides that, most of the patients were not hospitalized and the majority that present any infectious respiratory agent/disease were negative cases.

A correlation analysis was then performed by using

Table 2 – Respiratory pathologies agents used to create backtest process target

Respiratory pathologies agents	Backtest target
Respiratory syncytial virus	Any respiratory disease detection
Influenza A virus	
Influenza A virus (H1N1)	
Influenza B virus	
Parainfluenza 1	
Parainfluenza 3	
Parainfluenza 4	
Adenovirus	
Rhinovirus/Enterovirus	
Coronavirus HKU1	
Coronavirus NL63	
Coronavirus 229e	
Coronavirus C43	
Bordetella pertussis bacteria	
Metapneumovirus	
Chlamydophila pneumoniae bacteria	

Reference – Developed by the authors



Reference – Developed by the authors

Figure 1 - Patient flow diagram for preprocessed dataset (ICU: Intensive Care Unit)

* According to article 1, sole paragraph, items 2 and 3 of CONEP (National Health Council) resolution n° 510, April 2016, it is not necessary a CEP (Ethics and Research Committee) research approval for developed works that used public data sources.

the Spearman coefficient for the blood features to remove highly correlated features, considering 0.8 as a threshold. Through this analysis, RBC, hematocrit and MCH were removed given the high correlation with hemoglobin (the first two) and MCV respectively. In the end, the feature space was composed of the remaining blood variables and patient age quantile. The feature patient age quantile was kept given the ease which it can be extracted during an exam.

Model development. In this step, the selected subset was split into training and test using a proportion of 85% and 15%, respectively. This split was made with a stratification based in the COVID-19 exam result column, i.e., was kept the same proportion of positive (13.6%) and negative (86.4%) of the COVID-19 exam result target where kept in these datasets.

After that, the Synthetic Minority Oversampling Technique (SMOTE) was applied in the training set. SMOTE is an oversampling technique to remove the balancing problem between the classes in training set by creating new synthetic positive cases[24] (was selected 5 nearest neighbours for the method). The authors selected five well-known AI algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GBM), and eXtreme Gradient Boosting (XGBoost). The authors opted for those methods, instead of more advanced as deep learning, given their generalization power in learning from small datasets, such as the selected subset.

The hyperparameters tuning for each algorithm was defined by a Bayesian Optimization with 5-Fold Cross-Validation (BOCV-5) in defined parameter grids spaces[22] using F1-Score as a metric to be optimized by the algorithm.

Results evaluation

In this step, the authors picked the best model. The model selection analyzed three major results: F1-Score performance in BOCV-5, test set performance considering F1-Score and AUC (Area Under the receiver operating characteristic Curve) metrics and backtest process.

The backtest process consists in each model accuracy performance in a subset extracted from the test set of 26 patients negative for COVID-19, but positive for other respiratory pathologies, defined by "Any respiratory disease detection" column. The authors developed this procedure to evaluate if the created models presented any possible confusion between COVID-19 and other respiratory diseases.

At the end, the selected AI model presented F1-Score, AUC, Sensitivity and Specificity metrics using a stratified Bootstrap 95% Confidence Intervals (BCI95%) and calculated through 999 bootstrapped samples extracted from the test set[17,25] , with the same test set size and proportion for positive and negative cases.

Qualitative analysis

In this step, the best model was analyzed with SHAP (Shapley Additive explanations)[26] seeking to interpret how it made the predictions, the importance behind this analysis and possible pitfalls used to get those insights.

RESULTS AND DISCUSSION

Table 3 shows the results for F1-Score during BOCV-5.

Except for the LR models, the results show an excellent performance during cross-validation for the trainning set, with metrics above 90%. However, the estimated metrics for the test set highlights RF as the best model (see Table 4). In addition, the disparity between this model and the others in F1-Score demonstrates a greater generalization power and prediction of patients positive for COVID-19, since the training set is an unbalanced set where the majority class are negative cases and F1-Score penalizes more false positives.

Table 5 shows backtest accuracy results for each trained model.

GBM and XGBoost models shown the best results in backtest process, followed by RF, SVM and LR models (Table 5). Despite being a positive result for GBM and XGBoost, these models may have given preference in the prediction of negative classes, given the low performance in F1-Score compared to the RF model shown in Table 4. In addition, the difference between them and the RF model is just one misclassified sample.

The authors selected the RF model as the best model, given its greater generalization power for negative and mainly positive patients for COVID-19 demonstrated by

Table 3 – BOCV-5 overall results for AI algorithms in training set

AI Algorithm	F1-Score Mean (%)	F1-Score Standard Deviation (%)
LR	83.4	3.0
SVM	94.8	1.3
RF	94.6	2.0
GBM	96.0	2.0
XGBoost	93.7	2.4

Reference – Developed by the authors

Table 4 – Test set performance results. In bold are highlighted the best scores

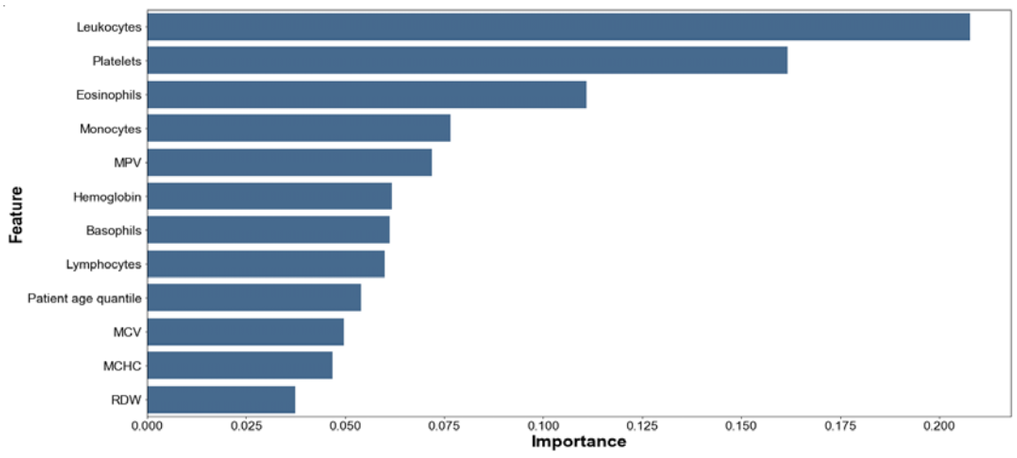
AI Algorithm	F1-Score (%)	AUC (%)
LR	48.6	78.9
SVM	51.9	85.6
RF	**78.3**	**94.4**
GBM	57.1	90.0
XGBoost	58.3	88.9

Reference – Developed by the authors

Table 5 – Backtest evaluation results. In bold are highlighted the best score

AI Algorithm	Accuracy (%)	Misclassified samples
LR	85.0	4
SVM	85.0	4
RF	92.0	2
GBM	**96.0**	**1**
XGBoost	**96.0**	**1**

Reference – Developed by the authors



Reference – Developed by the authors

Figure 2 - Random Forest feature importance results

the good results in Table 3, 4 and 5.

The best model metrics calculated using BCI95% in test set was: 77.7% F1-Score (IC95%:57.1;92.3), 85.9% AUC (IC95%:73.7;95.9), 74.4% Sensitivity (IC95%:50.0;92.1) and 97.5% Specificity (IC95%:93.6;100.0).

These results show an improvement in specificity and an equivalent AUC compared to the study carried out by Soares, 2020[18], using a less complex model and results much easier to be interpreted as shown later. In addition, compared to the study developed by Schwab, 2020[17], the space of features is much less complex given the pre-processing step. Both studies cited above use the same data source publicly available on the Kaggle's platform[23]. The RF model hyperparameters and other technical informations can be found at the GitHub repository[22].

Figure 2 ranks the feature importance, retrieved from selected RF model, created with gini impurity criteria. The three most important variables identified for this study were leukocytes, platelets and eosinophils with summed importance of ~45% (the importance was normalized to a range from zero to one). Similar importance results were found in developed AI models using datasets from Slovenia[16] (eosinophils), Italy[10] (eosinophils and platelets) and China[15] (leukocyte) showing a possible pattern from COVID-19.

Figure 3 presents a relation between the feature value (represented by the color in y-axis) and the SHAP calculated values for each training sample output (represented by x-axis). As the SHAP value increases, the model gains strength to report that the patient is positive for COVID-19. In addition, by decreasing the SHAP value, the model gains more strength to say the opposite: that the patient is negative for COVID-19.

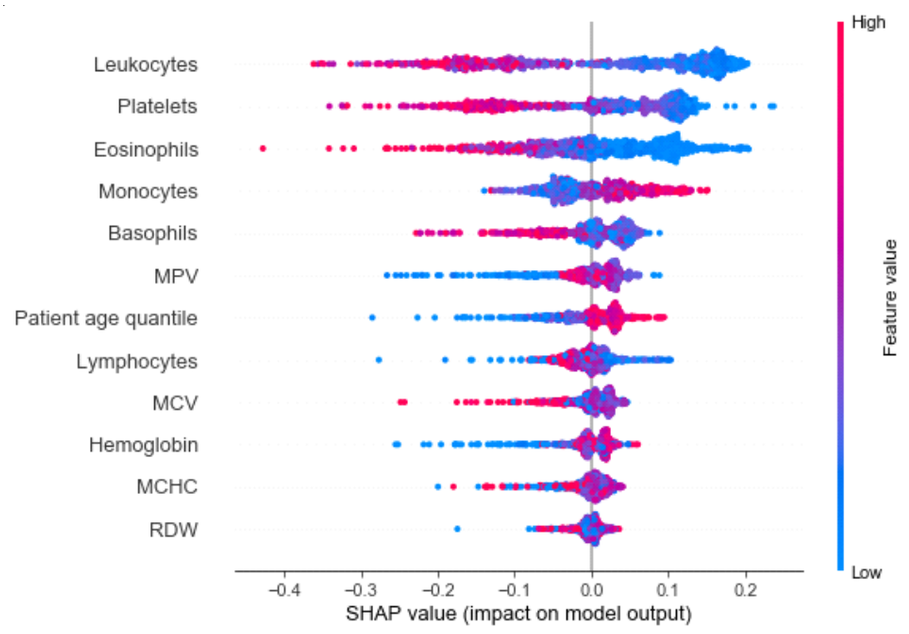
Through this summary analysis, the authors show that most of the features presented patterns: decreasing leukocytes, platelets and eosinophils implies a tendency to report positive, the same goes to increasing monocyte and patient age quantile. For negative reports, more patterns were found while observing the increasing leukocytes, platelets, eosinophils, basophils and MCV, and decreasing MPV, patient age quantile and hemoglobin more confidence for the RF to predict a negative case.

Still in Figure 3, as you increase leukocytes, platelets and eosinophils count shown as most important features in Figure 2, higher is the model confidence in reporting a negative case compared to positive cases given how lower are the negative SHAP values. This clear pattern for leukocytes, platelets and eosinophils may explain, for example, why they were the most important for the model in the feature importance presented in Figure 2.

Those patterns associated with positive cases were reported in medical papers for lower values of leukocytes[9,11-13], eosinophils[11-12] and platelets[11-12]. Surprisingly, the lymphocytes did not present a meaningful importance for positive COVID-19 cases as happens with Formica, 2020[10]. The lymphocyte decreasing values in patients are highly correlated with COVID-19 in several studies[7-9,11-13].

However, latest studies presented that patients with severe and fatal COVID-19 infection had significantly increased leukocyte count and decreased lymphocyte and platelet count compared to non-severe patient and survivors[27-28], different from results shown in Figure 3 for leukocyte. This adverse result can be explained by the period of sample collection (at the beginning of the pandemic scenario in Brazil) and the low sample size of patients with severe COVID-19 as shown in Figure 1, making it difficult to generalize the model.

The results described here show SHAP analysis importance AI interpretability, where it is possible to



Reference – Developed by the authors

Figure 3 - RF SHAP analysis for positive cases in the training set

understand how the model achieved that prediction through changes in the features, bringing the possibility of comparison with hypotheses and results at the medical literature. It may facilitate the understanding of the model and the discussion of the results obtained in comparison to the current medical literature.

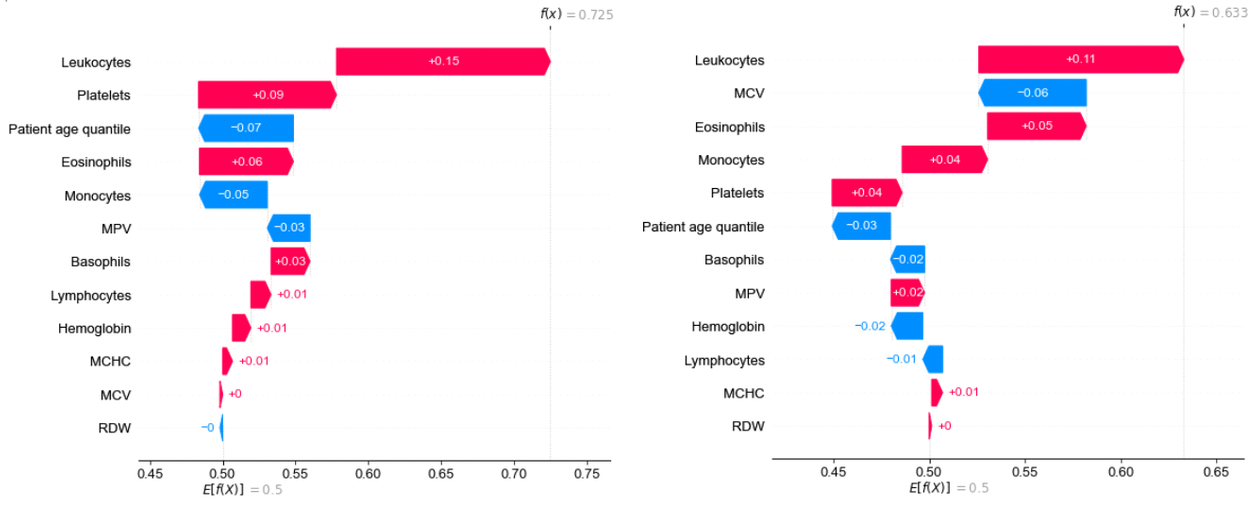
Given the data anonymization, situations such as leukopenia or lymphopenia in patients cannot be assessed. Figure 4 demonstrated some false positive cases evaluated.

The observed misclassification from the selected AI algorithm can be used to adress possible pitfalls. In red are presented supports to induce the RF model predict a positive report for COVID-19 and in blue are opposite supports: predict that the patient does not have COVID-19, the correct outcome.

Given the higher importance for leukocyte in the model if the patient is negative for COVID-19, but still appears with a lower value for this feature given any other disease such as anemia, the model may incorrectly classify as a positive case for COVID-19. The same goes to platelets and eosinophils that appear with higher SHAP values for a misclassification.

This can present a possible pitfall in the model and so precautions are recommended when creating predictive models for COVID-19 based on blood features. One possible solution would be to add other features such as patient symptoms and symptom period. This would further improve the model to get a prediction more focused on COVID-19.

In addition, this result shows that for this dataset, the leukocyte count increase is highly associated with patients negative for COVID-19, different from the current



Reference – Developed by the authors

Figure 4 - SHAP values for samples wrongly classified as positive in backtest

literature[27-28].

CONCLUSION

The developed RF model indicates goods results compared to developed studies using the same dataset. The model presented in this study demonstrates a high predictive power for detecting patients negative for the disease given the BCI95% specificity. Compared to the literature, this work brought the easiest model analysis to be interpreted and explained, where variations in features can be mapped for the final prediction and possible pitfalls can be found through an explanaible AI approach with SHAP.

The developed approach demonstrated here, based on the TRIPOD, could be applied to any other pathology detection for decision-making softwares where SHAP analysis can be used to medical interpretation, improving trustiness in model's prediction.

Even with great results, the use of this model for COVID-19 would need to be carried out with caution: given the data anonymization, many important information was hidden, such as the patient's gender and the presence of any comorbidity, highly correlated to COVID-19 mortality. In addition, the model generalization power decreases given the fact that the collected samples were carried out in one city (São Paulo, Brazil) at the pandemic beginning and due to the prevalence of a financial distinction between the target audience of the Albert Einstein Hospital and the overall population.

In conclusion, this work presents a strong approach for AI applications in the health field, leaving out black box solutions and adopting explainable results to support medical decisions.

ACKNOWLEDGMENTS

The authors are grateful to Jairo da Silva Freitas Júnior for the fruitful discussions about the development of AI models applied in healthcare.

REFERENCES

1. Johns Hopkins University CSSE. Coronavirus COVID-19 Global Cases: ArcGIS. [cited 2020 Jul 29]. COVID-19 Dashboard by CCSE at JHU. Available from: https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6.
2. Sanchez E. COVID-19 science: Why testing is so important. American Heart Association [Internet]. [cited 2020 Jul 29]. Available from: https://www.heart.org/en/news/2020/04/02/covid-19-science-why-testing-is-so-important.
3. Naudé W. Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. AI and Society. 2020; 35: 761-65.
4. Maguolo G, Nanni L. A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images. arXiv:2004.12823v3 [Preprint]. 2020 [cited 2020 Jul 28]. Available from: https://arxiv.org/abs/2004.12823.
5. Wynants L, Calster V, Bonten B, et al. Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal. The BMJ. 2020; 369: 1-14.
6. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. The Lancet. 2020; 395(10236): 1579-86.
7. Wang D, Hu B, Hu C. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA. 2020; 323(11): 1061-9.
8. Wei X, Jing D, Jia B, et al. Characteristics of in peripheral blood of 70 hospitalized patients and 8 diarrhea patients with COVID-19. Int J Med Sci. 2020; 17(9): 1142-6.
9. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020; 395(10223): 497-506.
10. Formica V, Minieri M, Bernardini S, et al. Complete blood count might help to identify subjects with high probability of testing positive to SARS-CoV-2. Clin Med (London). 2020; 20(4): 114-9.
11. Sun Y, Koh V, Marimuthu K, et al. Epidemiological and Clinical Predictors of COVID-19. Clin Infect Dis. 2020; 71(15): 786-92.
12. Lu G, Wang J. Dynamic changes in routine blood parameters of a severe COVID-19 case. Clin Chim Acta. 2020; 508: 98-102.
13. Fan B. Hematologic parameters in patients with COVID-19 infection. Am J Hematol. 2020; 95(6): 131-4.
14. Brinati D, Campaigner A, Ferrari D, et al. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: a Feasibility Study. J Med Syst. 2020; 44(8): 135-46.
15. Bao F, He Y, Liu J, et al. Triaging moderate COVID-19 and other viral pneumonias from routine blood tests. arXiv:2005.06546v1 [Preprint]. 2020 [cited 2020 Jul 28]. Available from: https://arxiv.org/abs/2005.06546.
16. Kukar M, Gunèar G, Vovko T, et al. COVID-19 diagnosis by routine blood tests using machine learning. arXiv:2006.03476v1 [Preprint]. 2020 [cited 2020 Jul 28]. Available from: https://arxiv.org/abs/2006.03476.
17. Schwab P, Schütte A, Dietz B, et al. predCOVID-19: A Systematic Study of Clinical Predictive Models for Coronavirus Disease 2019. arXiv:2005.08302v1 [Preprint]. 2020 [cited 2020 Jul 28]. Available from: https://arxiv.org/abs/2005.08302.
18. Soares F, Villavicencio A, Fogliatto F, et al. A novel specific artificial intelligence-based method to identify COVID-19 cases using simple blood exams. medRxiv:v3 [Preprint]. 2020 [cited 2020 Aug 01] Available from: https://doi.org/10.1101/2020.04.10.20061036.
19. Yusuf M, Atal I, Li J, et al. Reporting quality of studies using machine learning models for medical diagnosis: a systematic review. BMJ Open. 2020; 10(3): 1-7.
20. Junior E. COVID-19: Desafios para modelagem e análise de dados. J Health Inform. 2020; 12(2): 12-3.
21. Salimon C, Macedo M. Aplicações de Business Intelligence na Saúde: Revisão de Literatura Business. J Health Inform. 2017; 9(1): 31-5.
22. Reis K, Oliveira-Esquerre K. Reproducible Paper - Diagnosis of patients with blood cell count for COVID-19: An explainable artificial intelligence approach [github]. [cited 2020 Jul 29]. Available from: https://github.com/KaikeWesleyReis/paper-code-for-explainable-AI-in-COVID-prediction.
23. Albert Einstein Hospital [Internet]. Diagnosis of COVID-19 and its clinical spectrum: Kaggle. [cited 2020 Jul 28]. Available from: https://www.kaggle.com/einsteindata4u/covid19.
24. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002; 16:321-57.
25. Rajkomar A, Oren A, Chen K, et al. Scalable and accurate deep learning with electronic health records. Nat P J Dig Med. 2018; 1(1): 1-10.
26. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020; 2: 56–67.
27. Grassly N, Pons-Salort M, Parker E, et al. Comparison of

molecular testing strategies for COVID-19 control: a mathematical modelling study. The Lancet. 2020; 20(12): 1381-9

28. Henry B, Santos de Oliveira M, Benoit S, et al. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med.2020; 58(7):1021-8