Prediction of ICU admission for COVID-19 patients: a Machine Learning approach based on Complete Blood Count data

Lorenzo Famiglini*, Giorgio Bini*, Anna Carobene†, Andrea Campagner*, and Federico Cabitza*
* DISCo, Università degli Studi di Milano-Bicocca, Milano, Italy
† Laboratory Medicine, IRCCS San Raffaele Scientific Institute, Milano, Italy

2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS) | 978-1-6654-4121-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/CBMS52027.2021.00065

***Abstract*—In this article we discuss the development of prognostic Machine Learning (ML) models for COVID-19 progression: specifically, we address the task of predicting intensive care unit (ICU) admission in the next 5 days. We developed three ML models on the basis of 4995 Complete Blood Count (CBC) tests. We propose three ML models that differ in terms of interpretability: two fully interpretable models and a black-box one. We report an AUC of .81 and .83 for the interpretable models (the decision tree and logistic regression, respectively), and an AUC of .88 for the black-box model (an ensemble). This shows that CBC data and ML methods can be used for cost-effective prediction of ICU admission of COVID-19 patients: in particular, as the CBC can be acquired rapidly through routine blood exams, our models could also be applied in resource-limited settings and to get fast indications at triage and daily rounds.**

***Index Terms*—eXplainable AI, Machine Learning, COVID-19, Prognosis, Complete Blood Count**

I. Introduction

One year after its appearance, the SARS-CoV-2 coronavirus has infected more than 100 million people and has resulted in almost three million deaths worldwide. To mitigate this unprecedented pandemic spread, the use of AI techniques to develop tools that are supportive of clinicians in various tasks has attracted increasing interest. Despite promising results for the diagnostic task [1]–[3] (i.e., the detection of COVID-19), the development of prognostic models, either to predict ICU admission or other outcomes (including death) or to stratify patients by risk, has so far lagged behind: recent surveys report important limitations (in terms of bias or risk of overfitting) in the existing solutions [4], [5].

To address these limitations, in this work we report a retrospective study aimed at developing prognostic Machine Learning (ML) models to predict ICU admission, which can be seen as a proxy of disease severity or an outcome of worsening conditions. A large dataset of hematologic parameters has been collected from COVID-19 patients admitted to one of the largest teaching hospitals in Lombardy (Northern Italy), which was one of most severely affected regions during the first wave of the pandemic.

More specifically, we used and processed one of the most reliable datasets made available so far for COVID-19 analysis [1] (which is shared on the European open-access repository Zenodo[1]), motivated by the promising results regarding the strong association between blood tests data and COVID-19 prognosis [6], [7]. From this dataset, we extracted a small set of features regarding routine blood exams that are both inexpensive and quick to get, the so-called Complete Blood Count - CBC, for its wide application in a number of diagnostic and monitoring tasks. To the best of our knowledge, this is the first work using ML algorithms to perform COVID-19 prognosis only on the basis of CBC parameters. To this aim, we present three models, which have been conceived as complementary decision support tools. One model, which is based on the ensembling of 3 models, has been selected for its high accuracy, despite its low clinical interpretability because of the black-box nature. The other two models, i.e. a decision tree and a logistic regression, have been selected because of their explainability, despite their lower accuracy with respect to the ensemble model mentioned above. Indeed, these models can provide clinicians with more interpretable indications that can help them in their decision-making during the management and treatment of COVID-19 patients.

[1]https://zenodo.org/record/4081318#.X_1UDxYo-Uk

II. Methods

The study protocol (BIGDATA-COVID19) was approved by the Institutional Ethical Review Board in agreement with the World Medical Association Declaration of Helsinki.

In what follows, we report the data characteristics regarding the model development according to the MINIMAR guidelines [8], which were recently proposed to increase the understandability and reproducibility of Machine Learning studies in medical settings.

The dataset used for this retrospective study encompasses the results of routine blood tests of 1218 patients, regularly admitted to the hospital Emergency Department for COVID-19 of the San Raffaele Hospital (OSR), Milan (Italy). The data collection was performed between February 19, and May 31, 2020, i.e. at the height of the first wave of the epidemic in Italy. In that period, healthcare facilities in Northern Italy were under unprecedented pressure, especially the intensive care units [9], which on the 3rd of April peaked at 133% of their nominal capacity with 1381 inpatients. In the collected records, the average age of the patients was 63.5 ± 0.85 (mean and 95% confidence interval), and the distribution of biological sex was 70.8% males (vs. 29.2% females). For each patient, with at least 24 hours of hospitalization, multiple

DOI 10.1109/CBMS52027.2021.00065

TABLE I
COMPLETE LIST OF PREDICTIVE COVARIATES USED FOR THE MODEL DEVELOPMENT

Feature	Unit of Measure	Missing rate (%)
Sex	Male/Female	0
Age	Years	0
White Blood Cells (WBC)	$10^9/L$	0.4
Red Blood Cells (RBC)	$10^12/L$	0.4
Hemoglobin (HGB)	g/dL	0.4
Hematocrit (HCT)	%	0.4
Mean Corpuscular Volume (MCV)	fL	0.4
Mean Corpuscular Hemoglobin (MCH)	pg/Cell	0.4
Mean Corpuscular Hemoglobin Concentration (MCHC)	g Hb/dL	0.4
Erythrocyte Distribution Width (RDW)	CV%	0.5
Platelets (PLT)	$10^9/L$	0.4
Mean Platelet Volume (MPV)	fL	3.5
Neutrophils Count (NE—NET)	% — $10^9/L$	8.4
Lymphocytes Count (LY—LYT)	% — $10^9/L$	8.4
Basophils Count (BA—BAT)	% — $10^9/L$	8.4
Eosinophils Count (EO—EOT)	% — $10^9/L$	8.4
Monocytes Count (MO—MOT)	% — $10^9/L$	8.4

observations (approximately one for each day of hospital stay) were considered. In total, the dataset encompasses 4995 observations: for each instance (that is, one day of hospital stay for each given patient), the target corresponds to *whether the given patient would be admitted to the ICU within the next 5 days* (starting from the date of the observation).

The data exploration revealed an imbalance with respect to the target variable, skewed in favor of the negative class: the number of observations for which the patient was admitted to the ICU (within the 5 days time interval) was 1359 (27% of the total observations). We addressed this imbalance by means of the SMOTE re-sampling procedure (see below) and by considering balanced metrics.

As covariate features, we selected a set of 22 variables: namely gender, age and the Complete Blood Count (CBC), including the leukocyte formula (analyzed through a Sysmex XN 9000 hematology analyzer). We decided to focus on this set of features for two main reasons: first, these variables guaranteed the highest completeness rate for the dataset at hand (see Table I); second, and most important, these hematologic parameters can be obtained through rapid, widely available and cost-effective routine blood exams. The full set of features, with the respective missing data rates, is reported in Table I. The complete dataset, in compliance with medical ML reporting guidelines [8], has been made publicly available on Zenodo https://zenodo.org/record/4686707.

In order to perform missing data imputation, we used a multi-variate iterative imputation approach [10], for its capability to better take into account the latent distribution of the missing values compared with standard constant-based imputation strategies. Due to the relatively low number of missing values ($< 10\%$), we do not expect significant differences with respect to other multi-variate imputation strategies [11]

In regard to data imbalance, we applied the SMOTE oversampling method [12]. This approach, compared with standard under- or oversampling approaches, allows to better capture the distribution of the minority class (under standard smoothness assumptions).

Model testing was performed through a patient code-based train/test split: we used a 80%-20% data split with the additional constraint that all observations pertaining to each given patient were all in the same data fold. This setting was selected in order to reduce performance over-estimation due to potential auto-correlations among different observations pertaining to the same patient.

As anticipated in Section I, we considered three classes of models: two interpretable models, i.e., a decision tree and a (regularized) logistic regression; and a black-box ensemble model. According to the tenets of eXplainable AI [13], the more interpretable models were chosen to guarantee a sufficient accuracy level according to the general expectations of clinicians for prognostic tasks [14] in combination with a high level of interpretability, so that their output could be understood and examined by the clinicians involved. In regard to the black-box model, this was developed as a solution to maximize the model's discriminative performance. This model was obtained by determining the best combination among 5 different ML model classes: Gradient Boosting, logistic regression, Support Vector Machine, Random Forest and a Decision Tree.

In order to perform hyper-parameter selection, we used a Sequential Model-Based Optimization (SMBO) approach, implemented through the Optuna framework, which allows to perform a computationally efficient and model-agnostic search through the parameter space [15].

Hyper-parameter selection, model training and validation were performed on the training set through a 10-time repeated 7-fold Cross-Validation. Indeed, as shown in [16], this procedure has lower over-estimation bias compared to either boostrapping and standard Cross-Validation, while being less computationally intensive than Nested Cross-Validation.

The target metric for hyper-parameter selection was the F_2 score:

$$F_2 = 5\frac{PPV \cdot Sensitivity}{4 \cdot PPV + Sensitivity} \quad (1)$$

We chose this metric in order to improve sensitivity (which the above formula considers twofold more important than positive predictive value) and hence reduce the amount of false negatives. False negatives for the task at hand are worse than the false positives, as the former ones contribute to underestimate the number of ICU beds necessary in the near future.

For each model, we report five different metrics, evaluated on the test set, namely: area under the ROC curve (AUC), sensitivity, specificity, F_2 score and the Brier score (as a measure of calibration). In particular, we reported the F_2 score as a way to better account for the presence of label imbalance in the used dataset.

We also report the performance of the models on the instances that were associated with a probability score greater

than 75% [14]: this allows to assess the performance of the models on the instances these models were more "confident" about, or for which the prediction uncertainty was lower [17]. In the former case, we denote the related measures as "highly confident" - HC.

III. Results

The selected hyper-parameter values are as follows:

- **Decision Tree**: split criterion = entropy, maximum depth = 4, minimum samples per split = 26, minimum samples per leaf = 3;
- **Logistic Regression**: regularization norm = l1, regularization coefficient C = 0.0058, solver = liblinear;
- **Ensemble**: optimal configuration = XGBoost, Random Forest, Logistic Regression.

The performance of the developed models, in terms of sensitivity, specificity, F_2 score, AUC and Brier score is reported in Table II. Table II also reports performance scores evaluated on the instances with an associated probability score greater than 75% (denoted with a HC- prefix, where HC stands for Highly Confident).

A graphical representation of the models' performance in the ROC space is reported in Figure 1, while the calibration of the models is reported in Figure 2.

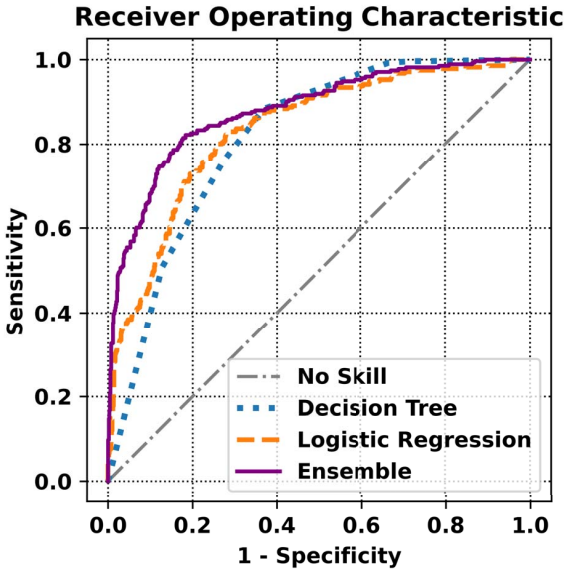

Fig. 1. ROC Curve evaluated on the Test Set.

IV. Discussion

The results reported in Section III show that our proposal to support the interpretation of COVID-19 cases is at the same level of - if not better than - the main contributions to the current literature. In what follows, we outline the main characteristics of these solutions and compare them with ours, to highlight the respective strengths and limitations.

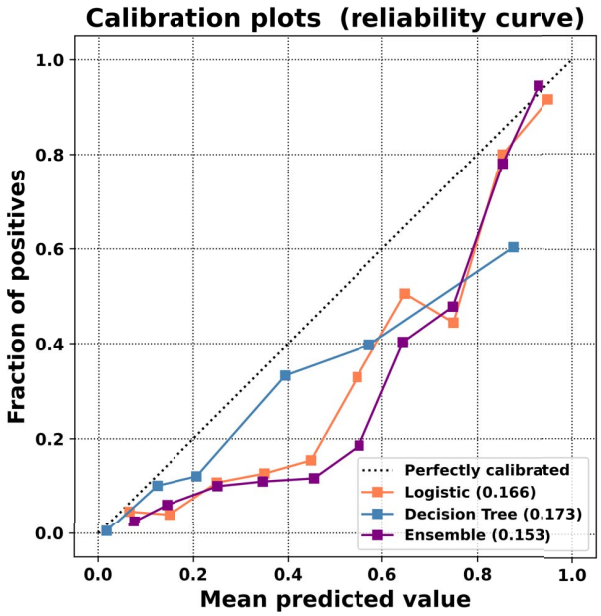

Fig. 2. Reliability curve for each model

Rodriguez-Nava et al. [18] developed clinical scores to predict ICU admission, and reported an AUC of .76. Although the proposed score is interpretable (like our logistic regression and decision tree models), it was developed on a relatively small sample encompassing 300 patients, and it was validated on the same data used to develop the score, with no method to control overfitting.

Wu et al. [19] developed a Logistic Regression model for risk prediction that they also externally validated: on the external validation sets the authors report an average AUC of .87, average sensitivity of .86 and average specificity of .71. The model, however, was developed and validated only with data collected between February and March 2020; also, compared with our proposed method, the model employs a large set of features encompassing hemato-chemical parameters, symptomatology and radiological findings. This could hamper its applicability in real-world medical practice, especially in resource-limited settings; our model, by contrast, only employs CBC data, i.e., a rapid, widely available and economic blood test. Also, while we address the task of ICU admission prediction, the authors of [19] consider a composite binary prediction task: a patient was considered severe in case of either ICU admission, organ failure, shock or death; this, in turn, can reduce the usefulness of the ML method in the management of severe cases.

Klann et al. [20] developed generalized-linear and gradient-boosting models for severity prediction based on *computable phenotypes* (that is, vector-based representations of a patient's clinical history and EHR data): for the task of ICU admission prediction, they report an average sensitivity of .77 and an average specificity of .79. While the reported results are comparable with our findings, we notice that our logistic re-

TABLE II
RESULTS OBTAINED ON THE TEST SET.

Model	Sensitivity	Specificity	AUC	F_2	Brier score	HC-AUC	HC-Sensitivity	HC-Specificity	HC-F_2	Coverage
Decision Tree	0.76	0.73	0.81	0.69	0.17	0.86	0.60	0.93	0.63	0.72
Logistic Regression	0.83	0.70	0.83	0.74	0.17	0.92	0.76	0.94	0.78	0.43
Ensemble	0.85	0.74	0.88	0.77	0.15	0.93	0.75	0.94	0.78	0.58

gression model achieves higher performance while being fully interpretable: this in turn could improve model understanding and foster trust in the clinical users and, ultimately, bring wider adoption in clinical practice as a medical decision support. Furthermore, the model discussed in [20] uses all information collected in patients' health records as predictive features, while the method that we propose only requires CBC data.

More in general, four recent reviews [5], [21]–[23] surveyed the state-of-the-art with respect to prognostic ML models for COVID-19: most of the surveyed works were found to be subject to a high risk of bias. This is due to limitations related to model development and data collection [24], lack of reporting standards, lack of procedures to control or mitigate over-fitting, and lack of data sharing [25] which, in turn, affects replicability.

As a final comment, we note some general important differences between the proposed approach and the surveyed works. First, all discussed models consider the task of severity (either death and/or ICU admission) prediction with a potentially unlimited prediction horizon: a case is considered to be severe if any severe adverse outcome occurs during the hospital stay, irrespective of its length. While this approach could help reducing data imbalance, it could also incur the risk of disregarding important confounding factors, such as the therapy; or it could require data which would not be available shortly after admission. In order to mitigate the impact of these confounding factors, in our method we considered prediction on a fixed 5-day horizon, which is nonetheless clinically meaningful and potentially useful. Second, our proposed approach is based only on CBC data and this is a major advantage for the following reasons:

- CBC can be acquired through routine exams;
- CBC can be acquired rapidly and with small costs compared to other more specialized biomarkers' data;
- Compared with other exams related to clinical chemistry [26], inflammatory markers [27], or coagulation parameters [28], CBC is less affected by both pre-analytical (that is, how specimens are collected, handled, and identified), analytical (which regards differences in the testing methods in different laboratories or with different equipment [29]) and biological variability (that is related to the fluctuations of biomarkers along patient's life [30]).

For these reasons, and in light of previous studies that highlighted key associations between CBC indicators and COVID-19 prognosis [6], [7], our CBC-based approach could be particularly useful for developing countries or for countries facing any resource shortage (e.g. in terms of specialized personnel), in that it provides a cost-effective method to predict ICU admission and, therefore, support the clinicians in ICU allocation planning.

In what follows, we discuss the results reported in Section III: as anticipated above, we observe that all three developed models achieved good results; in particular, all models achieve an AUC score greater than 80%. While these results are promising, we also acknowledge the following limitation: the generalizability of the developed models was not evaluated (e.g. through external validation), either on data collected from different settings, or collected from the same hospital but in a different period. Nonetheless, as previously mentioned, the adopted model development procedures were selected with the aim of increasing model robustness and reduce overfitting.

Interestingly, we note that the interpretable models (in particular, the logistic regression model) achieved good performance (see Table II): this shows that these approaches could be fruitfully used as decision-support tools that provide much more information, compared with the black-box model, and thus aid the clinicians in the decision-making task without undermining the predictive performance. Nonetheless, we note that, from a purely quantitative perspective, the ensemble model achieved the best performance.

Focusing on the interpretable models, the Decision Tree and the logistic regression coefficients are reported in Figures 3, 4, and 5. In both cases, the Neutrophils count feature is considered the most important prognostic variable, with higher Neutrophils count increasing the odds of ICU admission for both models. This information is consistent with the existing literature [31] where the prognostic role of the neutrophils-to-lymphocyte ratio is often discussed: interestingly, both models also associate a negative predictive power with the lymphocytes count (i.e. increased lymphocytes reduce the odds of ICU admission). Similar points can also be made for other leukocyte formula components, though the role for prognostic purposes of these other biomarkers has been less studied [32], [33]. As a further potential limitation of our study, we note that both interpretable models associate a decreasing odd of ICU admission with increasing age: this could be a consequence of the relatively small proportion of young patients in our sample (just around 25% of the involved patients was younger than 50 years), or an indication of more aggressive therapeutic interventions in older patients. In future work, we aim to collect data from more patients with possibly different age distributions, and see whether information about comorbidities (presence/absence or even the type) would have predictive power (as highly plausible). Also, we aim to validate our models also on COVID-19 negative patients: This would allow

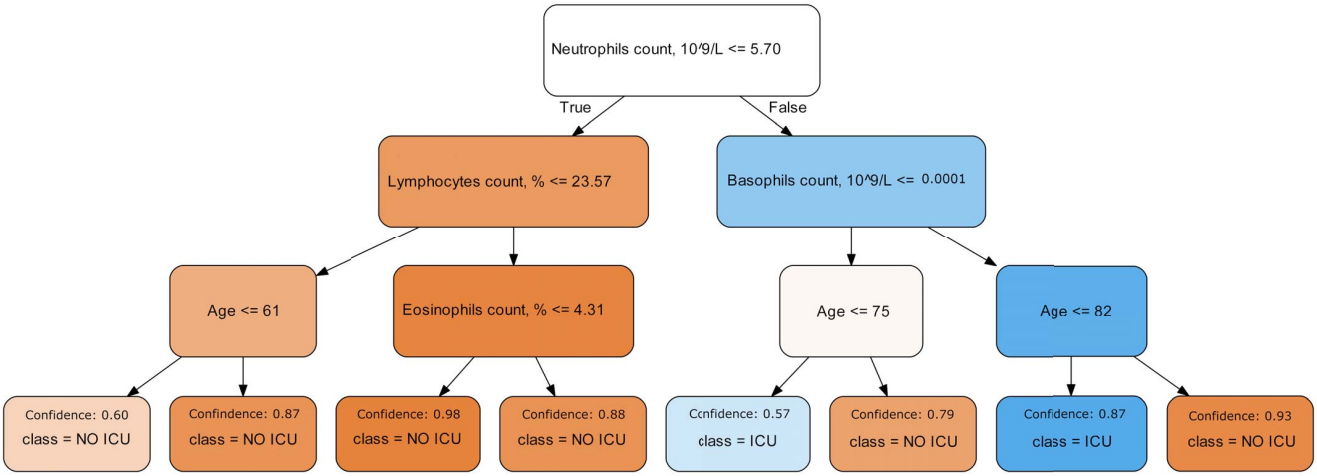

Fig. 3. A graphical representation of the Decision Tree

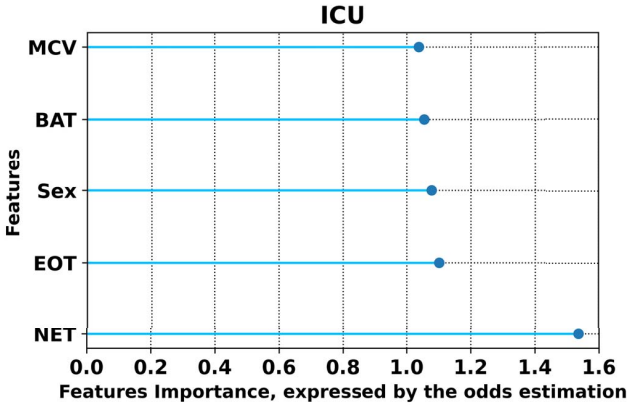

Fig. 4. Feature importance based on Logistic Regression coefficients, for the positive class (that is, admission to ICU)

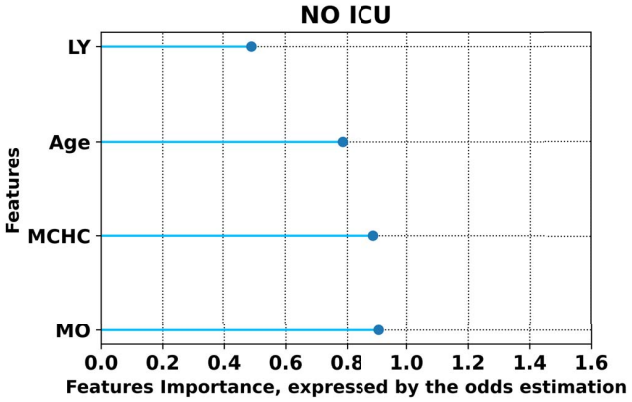

Fig. 5. Feature importance based on Logistic Regression coefficients, for the negative class.

the models to be applied also on cases that could be affected by false negatives in the RT-PCR test results.

In regard to calibration, we see from Table II that all models reported a good Brier score (we recall that the lower the Brier score, the better the model calibration): this can also be seen from Figure 2, where we can observe that all models over-estimate the probability scores, and especially so in the middle part of the plot (that is, on the more uncertain instances).

In order to better understand the performance of the models (as a function of the probability scores), we can observe the *Highly Confident* scores in Table II: the AUC and Specificity of the models increase when considering instances for which confidence is higher than 75%, while the Sensitivity decreases. This highlights the fact that the models tend to assign lower probability scores to positive- rather than negative-predicted instances: this is consistent with the decision to optimize the models for the F_2 score (which weights sensitivity more than PPV).

V. Conclusions

In summary, we reported a retrospective study to address the challenging task of predicting whether a COVID-19 patient will have to be transferred to the ICU within the next 5 days during their hospital stay. The proposed approach, based on both interpretable and black-box models, reported good results. Also, our methods are parsimonious, as they ground on two demographic features and the CBC test results, only: this is the main strength of our approach in light of acceptable accuracy. For this reason, our models can be useful in resource-limited settings, such as healthcare facilities which have to manage a surge of ill patients and that cannot afford the execution of more COVID-specific exams (e.g., inflammatory markers, interleukins and coagulation parameters [34]) on a daily basis.

For future work, we aim to externally validate our models with data coming from other hospitals and other time periods: This would allow to test the model in light of possible virus mutations and different patient management and therapeutic

policies. Since these latter ones depend on the number of cases to deal with and on the continuous advancement of what we know about COVID-19 and its effective treatment (changing its prognosis), phenomena related to *concept drift* cannot be ruled out in any existing predictive model, including ours.

REFERENCES

[1] F. Cabitza, A. Campagner, D. Ferrari, C. Di Resta, D. Ceriotti, E. Sabetta, A. Colombini, E. De Vecchi, G. Banfi, M. Locatelli *et al.*, "Development, evaluation, and validation of machine learning models for covid-19 detection based on routine blood tests," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 59, no. 2, 2021.

[2] S.-G. Chen, J.-Y. Chen, Y.-P. Yang, C.-S. Chien, M.-L. Wang, and L.-T. Lin, "Use of radiographic features in covid-19 diagnosis: Challenges and perspectives," *Journal of the Chinese Medical Association*, vol. 83, no. 7, p. 644, 2020.

[3] H. S. Yang, Y. Hou, L. V. Vasovic, P. A. Steel, A. Chadburn, S. E. Racine-Brzostek, P. Velu, M. M. Cushing, M. Loda, R. Kaushal *et al.*, "Routine laboratory blood tests predict sars-cov-2 infection using machine learning," *Clinical chemistry*, vol. 66, no. 11, pp. 1396–1404, 2020.

[4] P. Bachtiger, N. S. Peters, and S. L. Walsh, "Machine learning for covid-19—asking the right questions," *The Lancet Digital Health*, vol. 2, no. 8, pp. e391–e392, 2020.

[5] L. Wynants, B. Van Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, M. M. Bonten, D. L. Dahly, J. A. Damen, T. P. Debray *et al.*, "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *bmj*, vol. 369, 2020.

[6] E. J. Favaloro and G. Lippi, "Recommendations for minimal laboratory testing panels in patients with covid-19: Potential for prognostic monitoring." in *Seminars in Thrombosis and Hemostasis*, vol. 46, 2020, pp. 379–382.

[7] J. Linssen, A. Ermens, M. Berrevoets, M. Seghezzi, G. Previtali, H. Russcher, A. Verbon, J. Gillis, J. Riedl, E. de Jongh *et al.*, "A novel haemocytometric covid-19 prognostic score developed and validated in an observational multicentre european hospital-based study," *Elife*, vol. 9, p. e63195, 2020.

[8] T. Hernandez-Boussard, S. Bozkurt, J. P. Ioannidis, and N. H. Shah, "Minimar (minimum information for medical ai reporting): Developing reporting standards for artificial intelligence in health care," *Journal of the American Medical Informatics Association*, vol. 27, no. 12, pp. 2011–2015, 2020.

[9] G. Grasselli, A. Zangrillo, A. Zanella, M. Antonelli, L. Cabrini, A. Castelli, D. Cereda, A. Coluccello, G. Foti, R. Fumagalli *et al.*, "Baseline characteristics and outcomes of 1591 patients infected with sars-cov-2 admitted to icus of the lombardy region, italy," *Jama*, vol. 323, no. 16, pp. 1574–1581, 2020.

[10] S. v. Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of statistical software*, pp. 1–68, 2010.

[11] A. K. Waljee, A. Mukherjee, A. G. Singal, Y. Zhang, J. Warren, U. Balis, J. Marrero, J. Zhu, and P. D. Higgins, "Comparison of imputation methods for missing laboratory data in medicine," *BMJ open*, vol. 3, no. 8, 2013.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[13] R. Goebel, A. Chander, K. Holzinger, F. Lecue, Z. Akata, S. Stumpf, P. Kieseberg, and A. Holzinger, "Explainable ai: the new 42?" in *International cross-domain conference for machine learning and knowledge extraction*. Springer, 2018, pp. 295–303.

[14] F. Cabitza, A. Campagner, F. Del Zotti, F. Sternini, and A. Ravizza, "All you need is higher accuracy? on the quest for minimum acceptable accuracy for medical artificial intelligence," in *e-Health 2020, Proceddings of the 12th International Conference on e-health*, ser. E-health. IADIS, 2020, pp. 159–166.

[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631.

[16] J.-H. Kim, "Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap," *Computational statistics & data analysis*, vol. 53, no. 11, pp. 3735–3745, 2009.

[17] A. Campagner, F. Cabitza, and D. Ciucci, "The three-way-in and three-way-out framework to treat and exploit ambiguity in data," *International Journal of Approximate Reasoning*, vol. 119, pp. 292–312, 2020.

[18] G. Rodriguez-Nava, M. A. Yanez-Bello, D. P. Trelles-Garcia, C. W. Chung, H. J. Friedman, and D. W. Hines, "Performance of the quick covid-19 severity index and the brescia-covid respiratory severity scale in hospitalized patients with covid-19 in a community hospital setting," *International Journal of Infectious Diseases*, 2020.

[19] G. Wu, P. Yang, Y. Xie, H. C. Woodruff, X. Rao, J. Guiot, A.-N. Frix, R. Louis, M. Moutschen, J. Li *et al.*, "Development of a clinical decision support system for severity risk prediction and triage of covid-19 patients at hospital admission: an international multicentre study," *European Respiratory Journal*, vol. 56, no. 2, 2020.

[20] J. G. Klann, G. M. Weber, H. Estiri, B. Moal, P. Avillach, C. Hong, V. Castro, T. Maulhardt, A. L. Tan, A. Geva *et al.*, "Validation of a derived international patient severity algorithm to support covid-19 analytics from electronic health record data," *medRxiv*, 2020.

[21] N. Alballa and I. Al-Turaiki, "Machine learning approaches in covid-19 diagnosis, mortality, and severity risk prediction: A review," *Informatics in Medicine Unlocked*, p. 100564, 2021.

[22] J. Nayak, B. Naik, P. Dinesh, K. Vakula, B. K. Rao, W. Ding, and D. Pelusi, "Intelligent system for covid-19 prognosis: a state-of-the-art survey," *Applied Intelligence*, pp. 1–31, 2020.

[23] M. Sperrin, S. Grant, and N. Peek, "Prediction models for diagnosis and prognosis in covid-19," *BMJ*, no. 369, p. m1464, 2020.

[24] R. A. Hallman, A. Chikkula, and T. Prioleau, "Predicting criticality in covid-19 patients," in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2020, pp. 1–7.

[25] Y. Gao, G.-Y. Cai, W. Fang, H.-Y. Li, S.-Y. Wang, L. Chen, Y. Yu, D. Liu, S. Xu, P.-F. Cui *et al.*, "Machine learning based early warning system enables accurate mortality risk prediction for covid-19," *Nature communications*, vol. 11, no. 1, pp. 1–10, 2020.

[26] P. A. Kavsak, K. De Wit, and A. Worster, "Clinical chemistry tests for patients with covid-19–important caveats for interpretation," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 58, no. 7, pp. 1142–1143, 2020.

[27] A. Carobene, A. K. Aarsand, E. Guerra, W. A. Bartlett, A. Coşkun, J. Díaz-Garzón, P. Fernandez-Calle, N. Jonker, M. Locatelli, S. Sandberg *et al.*, "European biological variation study (eubivas): within-and between-subject biological variation data for 15 frequently measured proteins," *Clinical chemistry*, vol. 65, no. 8, pp. 1031–1041, 2019.

[28] E. J. Favaloro and J. Thachil, "Reporting of d-dimer data in covid-19: some confusion and potential for misinformation," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 58, no. 8, pp. 1191–1199, 2020.

[29] M. Seghezzi, S. Buoro, G. Previtali, V. Moioli, B. Manenti, R. Simon-Lopez, C. Ottomano, and G. Lippi, "A preliminary proposal for quality control assessment and harmonization of leukocytes morphology-structural parameters (cell population data parameters)," *Journal of medical biochemistry*, vol. 37, no. 4, pp. 486–498, 2018.

[30] S. Buoro, A. Carobene, M. Seghezzi, B. Manenti, A. Pacioni, F. Ceriotti, C. Ottomano, and G. Lippi, "Short-and medium-term biological variation estimates of leukocytes extended to differential count and morphology-structural parameters (cell population data) in blood samples obtained from healthy people," *Clinica Chimica Acta*, vol. 473, pp. 147–156, 2017.

[31] G. D. Pimentel, M. C. D. Vega, and A. Laviano, "High neutrophil to lymphocyte ratio as a prognostic marker in covid-19 patients," *Clinical Nutrition ESPEN*, vol. 40, pp. 101–102, 2020.

[32] M. Fraissé, E. Logre, H. Mentec, R. Cally, G. Plantefève, and D. Contou, "Eosinophilia in critically ill covid-19 patients: a french monocenter retrospective study," *Critical Care*, vol. 24, no. 1, pp. 1–4, 2020.

[33] G. Huang, A. J. Kovalic, and C. J. Graber, "Prognostic value of leukocytosis and lymphopenia for coronavirus disease severity," *Emerging infectious diseases*, vol. 26, no. 8, p. 1839, 2020.

[34] R. M. Elshazli, E. A. Toraih, A. Elgaml, M. El-Mowafy, M. El-Mesery, M. N. Amin, M. H. Hussein, M. T. Killackey, M. S. Fawzy, and E. Kandil, "Diagnostic and prognostic value of hematological and immunological markers in covid-19 infection: A meta-analysis of 6320 patients," *PloS one*, vol. 15, no. 8, p. e0238160, 2020.