COVIDAL: a machine learning classifier for digital COVID-19 diagnosis in German hospitals

Christina C. Bartenschlager, Stefanie S. Ebel, Sebastian Kling, Janne Vehreschild, Lutz T. Zabel, Christoph D. Spinner, Andreas Schuler, Axel R. Heller, Stefan Borgmann, Reinhard Hoffmann, Siegbert Rieg, Helmut Messmann, Martin Hower, Jens O. Brunner, Frank Hanses, Christoph Römmele

Angaben zur Veröffentlichung / Publication details:

Bartenschlager, Christina C., Stefanie S. Ebel, Sebastian Kling, Janne Vehreschild, Lutz T. Zabel, Christoph D. Spinner, Andreas Schuler, et al. 2023. "COVIDAL: a machine learning classifier for digital COVID-19 diagnosis in German hospitals." *ACM Transactions on Management Information Systems* 14 (2): 14. https://doi.org/10.1145/3567431.

COVIDAL: A Machine Learning Classifier for Digital COVID-19 Diagnosis in German Hospitals

CHRISTINA C. BARTENSCHLAGER, STEFANIE S. EBEL, and SEBASTIAN KLING, Chair of Health Care Operations/Health Information Management, Faculty of Business and Economics, Faculty of Medicine, University of Augsburg, Universitätsstraße 16, Augsburg, Germany
JANNE VEHRESCHILD, Department II of Internal Medicine, Hematology/Oncology, Goethe University, Frankfurt, Frankfurt am Main, Germany; Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany; German Centre for Infection Research (DZIF), partner site Bonn-Cologne, Cologne, Germany
LUTZ T. ZABEL, Laboratory Medicine, Alb Fils Kliniken GmbH, Eichertstraße 3, Göppingen, Germany
CHRISTOPH D. SPINNER, Technical University of Munich, School of Medicine, University Hospital rechts der Isar, Department of Internal Medicine II, Ismaninger Str. 22, Munich, Germany
ANDREAS SCHULER, Gastroenterology, Alb Fils Kliniken GmbH, Eichertstraße 3, Göppingen, Germany
AXEL R. HELLER, Anaesthesiology and Operative Intensive Care Medicine, Medical Faculty, University of Augsburg, Stenglinstrasse 2, Augsburg, Germany

---

Frank Hanses and Christoph Römmele shared last authorship.
Authors' addresses: C. C. Bartenschlager, S. S. Ebel, S. Kling, and J. O. Brunner, Chair of Health Care Operations/Health Information Management, Faculty of Business and Economics, Faculty of Medicine, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany; emails: christina.bartenschlager@uni-a.de, stefanie.ebel@uni-a.de, sebastian.kling@uni-a.de, jens.brunner@uni-a.de; J. Vehreschild, Department II of Internal Medicine, Hematology/Oncology, Goethe University, Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany; German Centre for Infection Research (DZIF), partner site Bonn-Cologne, Cologne, Germany; email: Janne.Vehreschild@kgu.de; L. T. Zabel, Laboratory Medicine, Alb Fils Kliniken GmbH, Eichertstraße 3, Göppingen, Germany; email: Lutz.Zabel@af-k.de; C. D. Spinner, Technical University of Munich, School of Medicine, University Hospital rechts der Isar, Department of Internal Medicine II, Ismaninger Str. 22, Munich, Germany; email: Christoph.Spinner@mri.tum.de; A. Schuler, Gastroenterology, Alb Fils Kliniken GmbH, Eichertstraße 3, Göppingen, Germany; email: Andreas.Schuler@af-k.de; A. R. Heller, Anaesthesiology and Operative Intensive Care Medicine, Medical Faculty, University of Augsburg, Stenglinstrasse 2, 86156 Augsburg, Germany; email: axel.heller@uk-augsburg.de; S. Borgmann, Hygiene and Infectiology, Klinikum Ingolstadt, Krumenauerstraße 25, 85049 Ingolstadt, Germany; email: stefan.borgmann@klinikum-ingolstadt.de; R. Hoffmann, Laboratory Medicine and Microbiology, Medical Faculty, University of Augsburg, Stenglinstrasse 2, 86156 Augsburg, Germany; email: Reinhard.Hoffmann@uk-augsburg.de; S. Rieg, Clinic for Internal Medicine II - Infectiology, University Hospital Freiburg, Hugstetter Straße 55, 79106 Freiburg im Breisgau, Germany; email: siegbert.rieg@uniklinik-freiburg.de; H. Messmann and C. Römmele (on behalf of the LEOSS study group), Clinic for Internal Medicine III - Gastroenterology and Infectious Diseases, University Hospital Augsburg, Stenglinstraße 2, 86156 Augsburg, Germany; emails: Helmut.Messmann@uk-augsburg.de, Christoph.Roemmele@uk-augsburg.de; M. Hower, Department of Pneumology, Infectious Diseases and Intensive Care, Klinikum Dortmund gGmbH, Hospital of University Witten/Herdecke, Münsterstraße 240, 44145 Dortmund, Germany; email: Martin.Hower@klinikumdo.de; F. Hanses, Emergency Department, University Hospital Regensburg, Germany, Department for Infection Control and Infectious Diseases, University Hospital Regensburg, Franz-Josef-Strauß-Allee 11, 93053 Regensburg, Germany; email: frank.hanses@ukr.de.

https://doi.org/10.1145/3567431

STEFAN BORGMANN, Hygiene and Infectiology, Klinikum Ingolstadt, Germany
REINHARD HOFFMANN, Laboratory Medicine and Microbiology, Medical Faculty, University of Augsburg, Stenglinstrasse 2, Augsburg, Germany
SIEGBERT RIEG, Clinic for Internal Medicine II - Infectiology, University Hospital Freiburg, Germany
HELMUT MESSMANN, Clinic for Internal Medicine III - Gastroenterology and Infectious Diseases, University Hospital Augsburg, Stenglinstraße 2, Augsburg, Germany
MARTIN HOWER, Department of Pneumology, Infectious Diseases and Intensive Care, Klinikum Dortmund gGmbH, Hospital of University Witten/Herdecke, Dortmund, Germany
JENS O. BRUNNER, Chair of Health Care Operations/Health Information Management, Faculty of Business and Economics, Faculty of Medicine, University of Augsburg, Universitätsstraße 16, Augsburg, Germany
FRANK HANSES, Emergency Department, University Hospital Regensburg, Germany; Department for Infection Control and Infectious Diseases, University Hospital Regensburg, Germany
CHRISTOPH RÖMMELE (ON BEHALF OF THE LEOSS STUDY GROUP), Clinic for Internal Medicine III - Gastroenterology and Infectious Diseases, University Hospital Augsburg, Stenglinstraße 2, Augsburg, Germany

For the fight against the COVID-19 pandemic, it is particularly important to map the course of infection, in terms of patients who have currently tested SARS-CoV-2 positive, as accurately as possible. In hospitals, this is even more important because resources have become scarce. Although polymerase chain reaction (PCR) and point of care (POC) antigen testing capacities have been massively expanded, they are often very time-consuming and cost-intensive and, in some cases, lack appropriate performance. To meet these challenges, we propose the COVIDAL classifier for AI-based diagnosis of symptomatic COVID-19 subjects in hospitals based on laboratory parameters. We evaluate the algorithm's performance by unique multicenter data with approximately 4,000 patients and an extraordinary high ratio of SARS-CoV-2-positive patients. We analyze the influence of data preparation, flexibility in optimization targets, as well as the selection of the test set on the COVIDAL outcome. The algorithm is compared with standard AI, PCR, POC antigen testing and manual classifications of seven physicians by a decision theoretic scoring model including performance metrics, turnaround times and cost. Thereby, we define health care settings in which a certain classifier for COVID-19 diagnosis is to be applied. We find sensitivities, specificities, and accuracies of the COVIDAL algorithm of up to 90 percent. Our scoring model suggests using PCR testing for a focus on performance metrics. For turnaround times, POC antigen testing should be used. If balancing performance, turnaround times, and cost is of interest, as, for example, in the emergency department, COVIDAL is superior based on the scoring model.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms** • **Applied computing** → *Life and medical sciences*; *Decision analysis*;

Additional Key Words and Phrases: Machine learning, COVID-19 diagnosis, multicenter data

ACM Reference format:
Christina C. Bartenschlager, Stefanie S. Ebel, Sebastian Kling, Janne Vehreschild, Lutz T. Zabel, Christoph D. Spinner, Andreas Schuler, Axel R. Heller, Stefan Borgmann, Reinhard Hoffmann, Siegbert Rieg, Helmut Messmann, Martin Hower, Jens O. Brunner, Frank Hanses, and Christoph Römmele (on behalf of the LEOSS study group). 2023. COVIDAL: A Machine Learning Classifier for Digital COVID-19 Diagnosis in German Hospitals. *ACM Trans. Manag. Inform. Syst.* 14, 2, Article 14 (March 2023), 16 pages.
https://doi.org/10.1145/3567431

1 INTRODUCTION

For the fight against the COVID-19 pandemic, it is particularly important to map the course of infection, in terms of patients who have currently tested SARS-CoV-2 positive, as accurately as

Table 1. Number of Publications in Different Categories based on Literature Reviews on COVID-19 and AI

Review	Hospital admission	Contact tracing	Drugs and vaccination	SARS-CoV-2 prognosis	SARS-CoV-2 diagnosis	Diagnsotic imaging
Wynants et al. [2020]	1	–	–	9	5	12
Lalmuanawma et al. [2020]	–	36	3	4	2	2
Alballa and Al-Turaiki [2021]	–	–	–	34	18	–

possible. In hospitals, this is even more important because resources such as nursing and medical staff have become scarce [Melman et al. 2021]. Although polymerase chain reaction (PCR) and point of care (POC) antigen testing capacities for COVID-19 diagnosis have been massively expanded, they are often very time-consuming and cost-intensive, and in some cases, lack appropriate performance. PCR tests are generally known as the gold standard in COVID-19 diagnosis in which a nasal or throat swab is drawn and analyzed in the laboratory. POC antigen tests also require a nasal or throat swab, but the test result can be autonomously interpreted using an antigen test kit. To meet the drawbacks of PCR as well as POC testing, many machine learning and Artificial Intelligence approaches have been proposed and reviewed by Wynants et al. [2020], Lalmuanawma et al. [2020], and Alballa and Al-Turaiki [2021]. Many articles, for example, Bertsimas et al. [2021a], Bertsimas et al. [2021b], Avetisian et al. [2021], Wang et al. [2021], or Zokaeinikoo et al. [2021], focus on the prognosis of COVID-19 patients in terms of the severity of the disease, contact tracing, or diagnostic imaging, and approximatelyimately 20 different paper study digital SARS-CoV-2 diagnosis as we do (see Table 1).

A detailed analysis of the 20 papers shows that most authors apply standard machine learning techniques, such as Support Vector Machines, Logistic Regression, or Neural Networks (see, e.g., Schwab et al. [2020] or Tschoellitsch et al. [2020]), to single-center data with a median data set size of 1,482 patients involving 280 SARS-CoV-2 positive cases (median). Usually, research on the influence of concrete data processing plays a subordinate role. A minority of researchers, such as Brinati et al. [2020], for example, suggest a readily available tool for the application of their algorithms or use alternative data for validation. Retrospective evaluation in terms of a head-to-head comparison with physicians experienced in COVID-19 care is not in the focus of researchers, although providing an important benchmark for performance measurement. Applied features are, for example, vital signs (e.g., Feng et al. [2020]), demographic characteristics (e.g., Shoer et al. [2020]), symptoms (e.g., Feng et al. [2020]) or laboratory parameters (e.g., Goodman-Meza et al. [2020]). Similarly, Alballa and Al-Turaiki [2021] find 18 frequently reported laboratory features. In our study, we focus on laboratory data for the diagnosis of COVID-19 and review the state-of-the-art in classification models and laboratory data for COVID-19 diagnosis. After that, we will briefly identify significant conclusions and limitations discovered in our literature review. To begin, red blood cell images of 24 volunteers are used in a highly comparative time-series analysis, to classify patients in COVID-19 positive and negative cases in O'Connor et al. [2022]. A deep Neural Network is used on routine blood tests in Rikan et al. [2022]. The authors find their algorithm performs better than comparable studies in Brinati et al. [2020], Cabitza et al. [2020], and Alakus and Turkoglu [2020]. A limitation of the aforementioned articles is that large multicenter data and offline classification approaches, such as expert opinions or POC testing, for comparison are not used. Thell et al. [2021] use blood test parameters to build a support tool to distinguish between COVID-19 positive and negative patients. They do not however evaluate their method with other potential approaches, concerning cost and turnaround times. Kukar et al. [2021] use machine learning and blood tests to classify patients in a single center. Zhang et al. [2022] use dif-

ferent machine learning models and blood tests to classify COVID-19 positive patients into mild and severe cases. Chadaga et al. [2022] use blood parameters and machine learning in a single center study. De Freitas Barbosa et al. [2022] use 24 different best-in-class blood tests to classify the COVID-19 status of patients. They determine their method saves time and cost. They do not compare the quality of their support system with other offline means of classification. Thimoteo et al. [2022] compare explainable machine learning models with black-box models for the diagnosis of COVID-19, using blood counts. A limitation of their studies is that they do not consider other means of diagnosis and they do not discuss cost or turnaround times. AlJame et al. [2020] propose an ensemble learning algorithm to determine the COVID-19 status of patients. The authors compare their results to other machine learning models but not to physician classification or POC testing. The authors also provide the first comprehensive review into machine learning diagnosing models using routine laboratory/blood data for the classification of potential COVID-19 cases. We therefore advise their review to the interested reader for an in-depth look into different additional modelling approaches.

To the best of our knowledge, no model has been trained using German data or German multicenter data. Quantitative performance measurement usually involves frequently applied AI metrics such as sensitivity, specificity, or positive predictive values without explicitly considering turnaround times or classification cost. In our research, we introduce and evaluate the COVIDAL classifier for diagnosing symptomatic COVID-19 subjects in German hospitals by unique multicenter data, with a strong focus on Germany. COVIDAL is a new, non-standard machine learning algorithm based on laboratory parameters and optimizes, in its base version, the sensitivity and accuracy outcome. COVIDAL might be readily applied in German hospitals by our COVIDAL-APP, depending on the results of wide multicenter training and validation. Our data contains approximately 4,000 patients and an extraordinary high ratio of subjects tested SARS-CoV-2 positive. We analyze the influence of data preparation, flexibility in optimization targets, other than sensitivity and accuracy, in addition to the selection of the test set on the COVIDAL outcome. The algorithm is compared with standard AI, PCR, POC antigen testing and COVID-19 classifications of seven physicians of the University Hospital of Augsburg. Besides classical metrics, i.e., sensitivity and specificity, we explicitly evaluate the performance of the different classifiers applied in our work, adding turnaround times and classification cost to a decision theoretic scoring model. With the help of the model, we aim at defining settings in which a certain classifier is to be applied. We find high sensitivities, specificities, and accuracies of COVIDAL of up to 90% for our multicenter data. In many cases, COVIDAL outperforms standard AI or physicians' classification. Regarding data preparation, COVIDAL, other than standard AI techniques, works best with real-valued data and thus minimum data processing is necessary, which is a crucial finding for future applications in hospitals. The optimization strategy has a minor influence on the overall level of the outcome while a decision maker is enabled to flexibly adjust for varying predefined metrics. Our scoring model suggests using PCR testing for a focus on performance metrics as is the case in elective surgery, for instance. For a focus on turnaround times as in hospital visitor admission, POC antigen testing should be used. If there is an interest on balancing performance metrics, turnaround times and cost as, for example, in the emergency department, COVIDAL is superior based on our scoring model and the COVIDAL-APP might support physicians in their daily work.

Our work is structured as follows: Section 2 presents our data integration process, the different classifiers for COVID-19 diagnosis including a detailed description of the COVIDAL algorithm, performance metrics, and the scoring model. Section 3 depicts our results, which are discussed in Section 4. In Section 5, we conclude and give an outlook to future research.

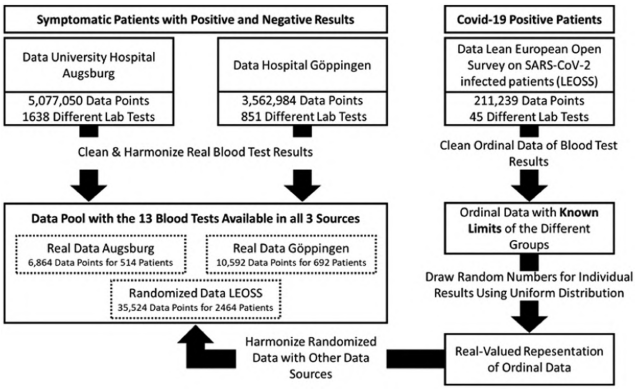

Fig. 1. Data harmonization process for real-valued data.

2 METHODS

2.1 Data Preparation

Data from different sources are included in our analyses: The University Hospital of Augsburg, the Alb Fils Kliniken, and the Lean European Open Survey on SARS-CoV-2 infected patients (LEOSS, https://leoss.net/). Data from the University Hospital of Augsburg, Germany, contains data of symptomatic patients in the early days of the pandemic, when screening was not yet done in German hospitals (March to June 2020). By omitting screening, we guarantee that only symptomatic patients and their blood results are considered for classification and asymptomatic patients do not distort the results. Symptomatic patients are defined in terms of respiratory infection/cold. The data provided by the Alb Fils Kliniken, Göppingen, Germany, where we consider data for the same timeframe, is similar.

While there are SARS-CoV-2-positive patients in both datasets, most patients are expected to test negative, having the prevalence rates in Germany in mind. Therefore, we consider a data export of LEOSS. In our LEOSS baseline data, laboratory parameters of SARS-CoV-2-positive patients across Europe, with a strong focus on the first pandemic wave in Germany are included (March 2020 to November 2020). Based on the LEOSS baseline definition, laboratory parameters of 48 hours after the first positive test result are considered. Unlike the data of the University Hospital of Augsburg and the Alb Fils Kliniken, the LEOSS data does not provide real-valued, i.e., cardinally scaled blood counts, but ordinally scaled data where blood counts are provided in predefined ranges. Varying data structure enables us to research the influence of data preparations on our results in detail. In the LEOSS protocol patients can be included via PCR confirmed diagnosis or rapid tests as an acceptable alternative. Approval for LEOSS was obtained by the applicable local ethics committees of participating centers and registered at the German Clinical Trials Register (DRKS, No. S00021145). The study at hand has been reported to the local ethics committee, too (20-465, BKF 2020-26). The LEOSS registry was supported by the German Centre for Infection Research (DZIF) and the Willy Robert Pitzer Foundation. To ensure anonymity in all steps of the analysis process, an individual LEOSS Scientific Use File (SUF) was created, which is based on the LEOSS Public Use File (PUF) principles described in Jakob et al. [2020].

Figure 1 shows the harmonization process for real-valued data, i.e., cardinally scaled blood counts. The data of both Augsburg and Alb Fils Kliniken first requires identical naming of blood tests and cleaning of implausible data. The LEOSS data is then translated from ordinal to real-valued data. While the specific values of patients' blood tests were not provided, the limits of the different ordinal categories are known. Using the upper and lower limits for the intervals of

different blood tests, random values are drawn from a uniform distribution. The number of laboratory tests as well as the number of data points varies between the three sources. Harmonization therefore requires not only renaming, but also determining the most important test parameters, which are available in all three data pools.

After discussions with clinical experts, restrictions for which patients and blood counts to include were implemented. For SARS-CoV-2 test results, we are mainly interested in the binary information whether a patient is SARS-CoV-2 positive or negative. If the first test of a symptomatic patient is negative and a later test, within the same hospital stay, for the same patient is positive, the patient is designated as SARS-CoV-2 positive for data concerning Augsburg or Alb Fils Kliniken. The logic behind this is that a test might be a false negative, but the blood counts might already indicate a SARS-CoV-2 positive patient. A patient's individual blood counts are included if they are taken 24 hours before or after the first PCR test. In case there are several laboratory determinations of the same kind for a patient within the period, the first value is included. For most of the included blood tests, we ensure that the specific blood test is done for more than 250 patients within 24 hours of the respective patient's first PCR test. A patient is not included in the combined data pool if any of the C-reactive protein, serum lactate dehydrogenase, hemoglobin, platelets, erythroblasts, or leukocytes tests is missing. These markers give important information about inflammation and cell death. Finally, three test codes, i.e., d-dimer, serum creatinine, and serum alanine aminotransferase, which inform about potential organ failure, are included into the set even if there are not 250 patient observations.

The real-valued samples x_{lk} for patient $l = 1, \ldots, L$ and blood parameter $k = 1, \ldots, K$, prepared as described before, are further processed using feature scaling or feature normalization. For feature normalization, samples are transformed using mean μ_k and standard deviation σ_k for feature k with $\frac{x_{lk}-\mu_k}{\sigma_k}$. Minima and maxima transform the samples in case of feature scaling: $\frac{x_{lk}-\min_l x_{lk}}{\max_l x_{lk}-\min_l x_{lk}}$. Besides the three real-valued datasets, we also include a range dataset, defined by the ranges of LEOSS.

In a final step, we divide the data, for range (data preparation 1), real-valued (data preparation 2), feature normalization (data preparation 3), and feature scaling (data preparation 4) preparations, into training and test data. The first test-split covers a randomly selected Augsburg subset with approximately 25% of Augsburg patients (test 1). Training data contains all data but the subset, i.e., training 1. For the second training-test-split, results for a ratio of 65% training data (training 2) to 35% test data (test 2) for each of the individual data sources are provided. In case of missing values of a patient's blood counts these values are temporarily dropped if they belong to the training set of COVIDAL whereas in the test set of COVIDAL and within standard AI they are filled with the mean value of the respective blood count.

2.2 Classifiers for COVID-19 Diagnosis

This work engages with COVID-19 diagnosis, i.e., the binary classification of symptomatic patients regarding an infection with SARS-CoV-2. For the implementation of the classification, we compare machine-learning, human and biochemical COVID-19 classifiers. Biochemical classifiers considered are Polymerase-Chain-Reaction (PCR) and Point-of-Care (POC) antigen tests. While PCR tests are widely known as gold-standard of SARS-CoV-2 diagnosis, POC antigen tests with readily available results have developed notable meaning from the second and third pandemic wave in Europe. Seven Augsburg physicians of different qualification levels and specialities, but experienced in the treatment of COVID-19 patients, function as our human classifiers. The physicians have been provided with the patients in the 25% Augsburg test data set (test set 1), however the dataset only contained blood parameters of the patients without their classification regarding a present COVID-19 disease. Different standard supervised machine learning algorithms for

classification problems are considered regarding machine learning algorithms. We apply a Logistic Regression model (LR), a Support Vector Machine (SVM) and an Artificial Neural Network (NN). In addition, we propose a new machine learning classifier, the COVIDAL algorithm, which is specifically developed for SARS-CoV-2 diagnosis of symptomatic subjects in German hospitals.

With the new algorithm, we aim at an approach (1) based on minimum data pre-preparation only, (2) trustable, (3) readily applicable, and (4) highly sensitive in the base version. The distinct construction of COVIDAL by the principle of Multiple Classifier Systems (MCS), which is reviewed in detail by Britto et al. [2014], targets goals (1) and (2). Goal (3) is reached by our COVIDAL-APP (see Appendix A.1). Here, we used Flask framework to develop a web interface where users may enter a patient's blood parameters and receive a classification of this patient. Goal (4) targets the fact that the actual detection of SARS-CoV-2 positive cases (i.e., sensitivity) might be of particular importance due to infection-prevention in hospitals and, in turn, we focus a sensitivity-centred optimization routine. The advanced version of COVIDAL involves substituting goal (4) with goal (4*) flexible in performance metrics, for example, specificity. The pool of 10 different classifiers employed for COVIDAL, its input parameters, their combination, selection and a thereupon calculation of the SARS-CoV-2 diagnosis is defined in detail in the Appendix A.2 by the COVIDAL Algorithm and a COVIDAL flowchart.

2.3 Performance Metrics and Evaluation

We measure the performance of biochemical, machine learning and human classifiers by sensitivity, specificity, accuracy, classification cost, and turnaround time. The turnaround time for the different categories is determined by the mean execution time of PCR, POC antigen, and other laboratory blood tests in the University Hospital of Augsburg as well as interviews with the physicians that classified patients by hand. Classification cost highlight the mean execution cost including material and personnel cost resulting from either cost of physicians or health care workers per day in the University Hospital of Augsburg. Sensitivity and specificity of PCR testing are based on manufacturer's information. For POC antigen tests, we use clinical outcomes of a study at the University Hospital of Augsburg [Kahn et al. 2021] confirmed by broad findings in thereupon literature (e.g., Scohy et al. [2020] or Torres et al. [2021]). Metrics for the machine learning and physician classifiers present the performance of the algorithms for the test dataset, while we select a superior performance based on balanced accuracy as our main performance metric. Balanced accuracy is calculated by the average of sensitivity and specificity, i.e., $\frac{\text{Sensitivity+Specificity}}{2}$. In addition, we report Area under the Curve (AUC) and F1-statistics to provide a holistic view on COVIDAL.

Note, a decision maker is basically interested in maximizing sensitivity and specificity while minimizing classification cost and turnaround time. We provide insights on this tradeoff by incorporating a decision theoretic evaluation scheme, namely a scoring model under certainty, into our analyses [Parmigiani and Inoue 2009]. The different classifiers a_i with $i \in \{1, \ldots, I\}$, i.e., PCR test, POC antigen test, physicians, standard machine learning and COVIDAL algorithms, and the performance metrics z_j with $j \in \{1, \ldots, J\}$, i.e., sensitivity ($j = 1$), specificity ($j = 2$), turnaround time ($j = 3$), and classification cost ($j = 4$), form the decision problem. Thereby, we score the alternatives, i.e., the classifiers a_i, by a weighted utility function $\sum_{j=1}^{J} w_j \cdot u_{ij}$ with varying weights w_j and $\sum_{j=1}^{J} w_j = 1$. Based on predefined weights and the performance metrics p_{ij}, an optimal classifier a_* is defined as follows:

$$a_* = \underset{i}{\operatorname{argmax}} \sum_{j=1}^{J} w_j \cdot u_{ij} \; with \; u_{ij} = \begin{cases} \frac{p_{ij} - \min_i p_{ij}}{\max_i p_{ij} - \min_i p_{ij}} & \forall j \in \{1, 2\} \\ \frac{p_{ij} - \max_i p_{ij}}{\min_i p_{ij} - \max_i p_{ij}} & \forall j \in \{3, 4\} \end{cases}$$

Table 2. Blood Parameters and Abbreviations

Abbr.	Blood parameter	Abbr.	Blood parameter
DDIM	d-Dimer	PTT	partial thromboplastin time
HGB	hemoglobin	cCRP	C-reactive protein
PLT	platelets	WBC	leukocytes
RBC	erythroblasts	cDBIL	serum direct bilirubin
cGGT	serum gamma-glutamyl transferase	cGPT	serum alanine aminotransferase
cHST	serum urea	cKREA	serum creatinine
cLDH	serum lactate dehydrogenase		

Table 3. Overview on Training and Test Sets for Data Preparation 1 (Range Data), 2 (Real-valued Data without Feature Scaling), 3 (Real-valued Data Including Feature Normalization) and 4 (Real-valued Data Including Feature Scaling)

Data preparation	Type of data	Test set	Description of test set	Positive	Negative	Training set	Positive	Negative
1	Ordinally scaled blood counts in ranges	Test 1 Test 2	25% of Augsburg data 35% of combined data	21 888	97 363	Training 1 Training 2	2,595 1,728	856 590
2, 3, 4	Cardinally scaled real valued blood counts	Test 1 Test 2	25% of Augsburg data 35% of combined data	21 890	103 396	Training 1 Training 2	2,606 1,737	940 674

The calculation of the utility u_{ij} for classifier i and metric j scales the decision matrix p_{ij} into the interval of $[0; 1]$ where values close to one highlight a superior outcome. The transformation is necessary, because p_{ij} is a combined matrix with high values for superior sensitivity and specificity, and low values for superior execution time and classification cost.

3 RESULTS

After running data preparation, we arrive at a combined data set of 3,670 patients for which up to 13 blood parameters are given (see Table 2). We include 95 SARS-CoV-2-positive patients out of 692 in case of Alb Fils Kliniken, 68 SARS-CoV-2-positive patients out of 514 total patients in the data of Augsburg and 2,464 SARS-CoV-2 positive LEOSS cases. Table 3 depicts the test and training sample sizes for the four different data preparations. Due to data harmonization, some samples are excluded in case of the range data (data preparation 1), for example slightly fewer patients than 3,670 are included here. Physicians have been presented with 124 patients (test set 1) where a subset of 21 patients tested SARS-CoV-2 positive. The data set contains a total of 3,546 patients with 940 negative correlates. In case of the training-test-split 2, 890 positive and 396 negative samples are tested based on a training set with 1737 SARS-CoV-2 positive and 674 negative patients.

The sensitivity of physicians presented with the 25% Augsburg test set (test set 1, data preparation 2) varies between 44.4% and 88.5%, while they achieved specificities from 50.4% to 89.8%. On the same data, COVIDAL classifies 90.48% of SARS-CoV-2-positive patients as SARS-CoV-2 positive (i.e., sensitivity). 87.38% of negative patients are classified as SARS-CoV-2 negative (i.e., specificity), while AUC is 79.06% and the F1 statistic is 71.70%. The standard AI algorithms, namely the Neural Network, the Logistic Regression, and the Support Vector Machine, maintain sensitivities from 90.48% to 100.00% and specificities from 0.00% to 30.10%. While the accuracy of the standard AI algorithms is rather low (16.94%, 33.06%, and 40.32%), physicians vary from 57.0% to 871% accuracy. COVIDAL provides 87.90% accuracy for the real-valued data processing (data preparation 2). In case of feature normalization (data preparation 3) and feature scaling (data preparation 4), the

Table 4. Performance Metrics for COVIDAL, Standard AI and Physicians' Classification, Different Data Preparations and Test Set 1

Data preparation	Algorithm	Sensitivity [%]	Specificity [%]	Accuracy [%]
1	NN	100.00	0.00	17.80
	LR	57.14	80.41	76.27
	SVM	38.10	86.60	77.97
	COVIDAL	100.00	49.48	58.47
2	NN	100.00	0.00	16.94
	LR	95.24	20.39	33.06
	SVM	90.48	30.10	40.32
	COVIDAL	90.48	87.38	87.90
	Physician 1	51.9	78.4	74.7
	Physician 2	88.5	50.4	57.0
	Physician 3	85.2	76.1	77.4
	Physician 4	51.9	84.4	79.9
	Physician 5	44.4	88.8	82.4
	Physician 6	70.4	72.5	72.2
	Physician 7	70.4	89.8	87.1
3	NN	71.43	80.58	79.03
	LR	95.24	30.10	41.13
	SVM	38.10	86.41	78.23
	COVIDAL	85.71	84.47	84.68
4	NN	52.38	84.47	79.03
	LR	100.00	12.62	27.42
	SVM	19.05	93.20	80.65
	COVIDAL	95.24	49.51	57.26

Please see Table 3 for descriptions on data preparation and test set.

Table 5. Performance Metrics of COVIDAL for Different Data Preparations and Test Sets

Data preparation	Algorithm	Test set	Sensitivity [%]	Specificity [%]	Accuracy [%]	AUC [%]	F1 [%]
1	COVIDAL	Test 1	100.00	49.48	58.47	49.48	46.15
	COVIDAL	Test 2	83.78	80.44	82.81	67.40	87.38
2	COVIDAL	Test 1	90.48	87.38	87.90	79.06	71.70
	COVIDAL	Test 2	89.66	90.40	89.89	81.06	92.47
3	COVIDAL	Test 1	85.71	84.47	84.68	72.40	65.45
	COVIDAL	Test 2	89.78	89.90	89.81	80.71	92.42
4	COVIDAL	Test 1	95.24	49.51	57.26	47.16	43.01
	COVIDAL	Test 2	89.78	89.90	89.81	80.71	92.42

Please see Table 3 for descriptions on data preparation and test set.

performance of the Neural Network is significantly improved which is no surprise to the known reader. COVIDAL maintains high performance metrics for data preparation 3. Logistic Regression and Support Vector Machine show a high variability for the metrics. Compared to data preparation 2, data preparation 1 leads to declining metrics for COVIDAL with 58.47% accuracy and increasing performance of the Logistic Regression model with 76.27% accuracy. The results are summarized in Table 4.

If the performance of COVIDAL is compared for different test sets, i.e., the physician test set with 25% of Augsburg data (test set 1) and the 35% test split (test set 2), we find a balanced behavior of COVIDAL on the latter. Sensitivities, specificities, and accuracies are similar to the performance of COVIDAL for the real-valued physician preparation (test set 1, data preparation 2) with approximately 90% in all metrics (see Table 5).

Table 6. Performance Metrics of COVIDAL for Different Optimization Strategies, Test Set 1 and Data Preparation 2

Optimization strategy	Sensitivity [%]	Specificity [%]	Accuracy [%]	AUC [%]	F1 [%]
Sensitivity and Accuracy	90.48	87.38	87.90	79.06	71.70
Specificity and Accuracy	71.43	86.41	83.87	61.72	60.00
Sensitivity and Balanced Accuracy	90.48	82.52	83.87	74.66	65.52
Specificity and Balanced Accuracy	95.24	71.84	75.81	68.42	57.14
F1 Positive and Accuracy	95.24	83.50	85.48	79.52	68.97
F1 Negative and Accuracy	85.71	69.90	72.58	59.92	51.43
F1 Positive and Balanced Accuracy	95.24	70.97	75.00	67.50	56.34
F1 Negative and Balanced Accuracy	95.24	78.64	81.45	74.90	63.49

Please see Table 3 for descriptions on data preparation and test set.

Table 7. Decision Matrix Based on the Results for Test Set 1 and Data Preparation 2

Classifier	Sensitivity [%]	Specificity [%]	Turnaround time [min]	Classification cost [€]
POC antigen	59.00	99.00	20	15
PCR	97.00	100.00	300	15
Physicians	66.10	77.20	60	4.76
COVIDAL	90.48	87.38	60	2.68
Standard AI	95.24	16.83	60	2.68

Please see Table 3 for descriptions on data preparation and test set.

A crucial characteristic of COVIDAL is the flexibility regarding an optimization strategy when selecting classifiers for the combination of two features, i.e., laboratory parameters. While the base version focuses a maximization of sensitivity and accuracy, we also research the impact of varying such metrics (i.e., advanced version). In particular, we focus eight strategies based on sensitivity, specificity, F1-Score, accuracy and balanced accuracy. Table 6 captures the results for the physician test set (test set 1) and real-valued data preparation 2. For the base case (second line in Table 6), the performance with 90.48% sensitivity, 87.38% specificity and 87.90% accuracy has been discussed before. Sensitivities vary from 71.43% to 95.24%, specificities vary from 69.90% to 86.41% and accuracies vary from 72.58% to 85.48% for optimization strategies two to eight. Thus, the optimization strategy has a minor influence on the overall level of the outcome while a decision maker is enabled to flexibly adjust for varying predefined metrics.

A summary of the performance metrics for all classifiers, data preparation 2 and the physician test set (test set 1) is presented in the decision matrix in Table 7. For physicians and standard AI algorithms, we average individual results. The turnaround time of POC antigen tests is 20 minutes, 300 minutes for the PCR test, i.e., 5 hours, and, for other classifiers, the turnaround time is approximately 60 minutes including a five-minute actual classification time for physician's decision or application analysis by an experienced health care worker. Classification cost are given by 15 € for POC antigen and PCR testing, 4.76 € for physician's classification and 2.68 € personnel cost for running an application. Per day, cost per physician in the University Hospital of Augsburg are 457.21 € and cost per health care worker are 257.06 €. We assume an eight-hour shift per employee and day.

By the scoring model for the decision matrix given in Table 7, PCR testing is superior for a sole sensitivity, i.e., $\boldsymbol{w} = (1, 0, 0, 0)$, or specificity perspective, i.e., $\boldsymbol{w} = (0, 1, 0, 0)$. POC antigen testing is suggested if one is interested in turnaround time, i.e., $\boldsymbol{w} = (0, 0, 1, 0)$. A decision maker having a

Table 8. Outcomes of the Scoring Model for Varying Weights

$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 0.5 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 0 \\ 0.5 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0.5 \\ 0.5 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0.5 \\ 0 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0.5 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.33 \\ 0.33 \\ 0.33 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0.33 \\ 0.33 \\ 0.33 \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}$
POC antigen	0.00	0.99	**1.00**	0.00	0.49	0.50	0.00	0.99	0.49	0.50	0.66	0.66	0.50
PCR	**1.00**	**1.00**	0.00	0.00	**1.00**	0.50	0.50	0.50	0.50	0.00	0.66	0.33	0.50
Physicians	0.19	0.73	0.86	0.83	0.46	0.52	0.51	0.79	0.78	0.84	0.58	0.80	0.65
COVIDAL	0.83	0.85	0.86	**1.00**	0.84	0.84	0.91	**0.85**	**0.92**	**0.93**	**0.84**	**0.89**	**0.88**
Standard AI	0.95	0.00	0.86	**1.00**	0.48	**0.91**	**0.98**	0.43	0.50	0.93	0.60	0.61	0.70

Values printed in bold highlight respective maxima.

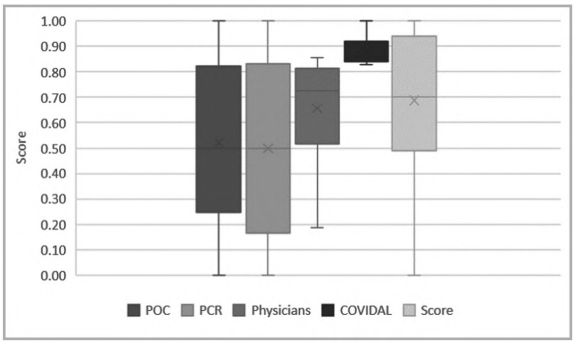

Fig. 2. Boxplots of the outcomes of the scoring model for the different classifiers and varying weights $\boldsymbol{w}$.

cost induced view, i.e., $\boldsymbol{w} = (0, 0, 0, 1)$, will prefer COVIDAL or other AI testing. When considering a balanced view on performance metrics (i.e., sensitivity and specificity), turnaround time and classification cost, i.e., $\boldsymbol{w} = (0.25, 0.25, 0.25, 0.25)$, COVIDAL is superior for most weight combinations with utility scores varying from 0.83 to 1 (see Table 8). POC antigen, PCR and standard AI scores vary from 0 to 1. The scores of physicians are in the range of 0.19 to 0.86 while being not optimal in any of the settings researched on. The rather low variability of COVIDAL scores including a high score-level compared with the respective scores of physicians, POC antigen, PCR, and standard AI (see Figure 2) highlights the balance of the COVIDAL outcome considering performance metrics, turnaround time, and classification cost.

4 DISCUSSION

Taking the results into consideration, COVIDAL maintains high performance rates, i.e., sensitivity, specificity, and accuracy, especially for minimal data preparation with real-valued cardinally scaled laboratory parameters. Standard AI techniques require elaborate data preparation including normalization, for example, to perform similar. From an application perspective, this is a crucial finding because minimum data preparation is necessary to run COVIDAL, which might contribute to trustworthiness and actual application. On the other hand, the algorithm depends on this exact data and the loss of information due to scaling does have an impact on specificity, especially. Test data and optimization strategies have a minor influence on the overall level of COVIDAL outcomes. A decision maker can thus flexibly create an output under consideration of different metrics. COVIDAL outperforms physicians' classification, which is a unique comparison in the literature stream on SARS-CoV-2 diagnosis and AI.

A major motivation of AI-based SARS-CoV-2 diagnosis is to relieve hospital resources, in terms of monetary aspects, personnel, testing, and other resources. We quantify this tradeoff by a decision-theoretic scoring model which captures performance metrics, turnaround time and classification cost. COVIDAL thereby guarantees a balance of the three measures while PCR tests are superior in case of an isolated view on performance metrics. In addition, the COVIDAL-APP may be operated by experienced health care workers rather than by physicians themselves and thus supports physicians in their daily work. From the user's point of view, COVIDAL is therefore well suited for situations in which a balance of the metrics is to be achieved and at the same time laboratory parameters are usually determined routinely for symptomatic patients. The emergency department with limited personnel capacity and time pressure could be a well-defined use case. The focus is different for elective surgeries, for example, where the dates are usually fixed a certain time in advance and thus testing can be outsourced to test centres. Here, decision makers in hospitals might opt for PCR testing with a superior performance regarding sensitivity and specificity. POC antigen testing functions as an option for situations in which the turnaround time plays an important role as, for example, in hospital visitor admission with limited space capacities in the entrance area in hospitals [Bartenschlager et al. 2022]. Visitors usually stay in hospital a limited time and other protective measures such as N95 face masks are mandatory. This may shift the prioritization towards the metric on turnaround time.

The limitations of our study include the following: While we try to balance our dataset with the help of LEOSS, the ratios of SARS-CoV-2-positive and -negative cases are unequal including caution in interpreting the accuracy. Nonetheless, we decided to not include simulation techniques to balance the dataset in advance, because the under-represented class forms a considerable amount of data, we are already manipulating data for mapping LEOSS range to real-valued data and over- and under-sampling techniques are controversial (see, e.g., Visa and Ralescu [2005]). In addition, our LEOSS data mainly contains data of University hospitals and non-University hospitals but there might also be patients from institutes and medical practices included at a low scale. In the LEOSS dataset, we do not, differentiate between symptomatic and asymptomatic patients, other than for Augsburg and Göppingen data, because LEOSS contains SARS-CoV-2-positive cases only and we focus data at a rather early stage of the pandemic. We have been trying to include negative correlates from other German hospitals, too. Although we are provided with an ethical statement to do so, the significant workload during COVID-19 pandemics in hospitals hinders trials so far. Another concern might arise from the timeframe defined for including laboratory parameters into the training and test data. LEOSS refers to the positive test and Augsburg and Alb Fils Kliniken refer to the first PCR test. Given that we focus on risk-averse proceedings regarding a classification and the latter, other than LEOSS, provide SARS-CoV-2-positive and -negative patients justifies our approach. The performance metrics for POC antigen tests are based on a study in the University Hospital of Augsburg, but the outcomes do not capture the patients involved in the study at hand. A similar concern might arise from evaluating the gold-standard PCR analyses by manufacturer's information while determining POC antigen testing measurements by clinical outcomes. Another concern might arise from resistance of clinicians in potentially using COVIDAL. The use of machine learning models is particularly questioned in the medical context due to concerns regarding black-box behaviour and understandability of the results. Ethical, legal, and social issues also play a role here. On the other hand, the COVID-19 pandemic has impressively shown what happens when resources reach their limits. This is where Clinical Decision Support Systems, e.g., the COVIDAL-APP, might help. The results of COVIDAL strongly depend on the input data and the pandemic situation in which the data has been collected. Varying lockdown measures and prevalence rates, for example, might influence the performance. Therefore, re-running the

model-building step and rebuilding the model periodically are important to verify that the model still provides sufficient accuracy for diagnostic purposes.

5 CONCLUSION

In this work, we research the application of the COVIDAL classifier for digital COVID-19 diagnosis in German hospitals. The conclusions are based on extensive sensitivity analyses and an innovative scoring model considering performance metrics, turnaround times, and classification cost. PCR, POC antigen tests, physicians‘ and standard AI applications function as benchmarks for comparative analyses. Among others, crucial aspects of our research are the unique multicenter data with a strong focus on Germany and an actual head-to-head comparison with physicians.

We find high sensitivities, specificities, and accuracies of COVIDAL of up to 90% for our multicentre data. In many cases, COVIDAL outperforms standard AI or physicians‘ classification. The findings highlight the applicability of COVIDAL in hospitals. Regarding data preparation, COVIDAL, other than standard AI techniques, works best with real-valued data and thus minimum data processing is necessary, which is a crucial finding having actual applications in hospitals in mind. The optimization strategy and the test set have a minor influence on our overall outcome and indicate a certain robustness of the results. Our scoring model suggests using PCR testing for a pure focus on performance metrics, e.g., for elective surgery. For a focus on turnaround times, e.g., in hospital visitor admission, POC antigen testing should be used. If there is an interest on balancing performance metrics, turnaround times and cost as, e.g., in the emergency department, COVIDAL is superior in most instances and might support physicians in their daily work.

For future research and an actual application of COVIDAL, an interdisciplinary discussion regarding AI- and COVIDAL-based decision support for SARS-CoV-2 diagnosis is to be held. This includes an ethical and legal framework for the practice of digital COVID-19 diagnosis. In addition, efforts should be made to include data from more hospitals for broad use of the COVIDAL-APP.

ACKNOWLEDGMENTS

We express our deep gratitude to the physicians of the University Hospital of Augsburg, among others Georg Braun, Alanna Ebigbo, Johannes Manzeneder, Lars Sudhoff, and Tobias Weber, for their valuable manual assessment of the test data. It was their extraordinary engagement that made the important head-to-head comparison possible. We express our deep gratitude to all study teams supporting the LEOSS study. The LEOSS study group contributed at least 5 per mille to the analyses of this study: University Hospital Regensburg (Frank Hanses), Technical University of Munich (Christoph Spinner), Hospital Ingolstadt (Stefan Borgmann), University Hospital Freiburg (Siegbert Rieg), Klinikum Dortmund gGmbH, Hospital of University Witten / Herdecke (Martin Hower), University Hospital Frankfurt (Maria Vehreschild), University Hospital Jena (Maria Madeleine Ruethrich), Practice at Ebertplatz Cologne (Christoph Wyen), Hospital Bremen-Center (Bernd Hertenstein), Hospital Passau (Julia Lanznaster), University Hospital Augsburg (Christoph Roemmele), Johannes Wesling Hospital Minden Ruhr University Bochum (Kai Wille), Hospital Ernst von Bergmann (Lukas Tometten), University Hospital Essen (Sebastian Dolff), University Hospital Munich/ LMU (Michael von Bergwelt-Baildon), University Hospital Heidelberg (Uta Merle), Robert-Bosch-Hospital Stuttgart (Katja Rothfuss), University Hospital Wuerzburg (Nora Isberner), University Hospital Cologne (Norma Jung), University Hospital Tuebingen (Siri Göpel), Hospital Maria Hilf GmbH Moenchengladbach (Juergen vom Dahl), Municipal Hospital Karlsruhe (Christian Degenhardt), University Hospital Erlangen (Richard Strauss), University Hospital Ulm (Beate Gruener), Hospital Leverkusen (Lukas Eberwein), Catholic Hospital Bochum (St. Josef Hospital) Ruhr University Bochum (Kerstin Hellwig), Bundeswehr Hospital Koblenz (Dominic Rauschning), Evangelisches Hospital Saarbruecken (Mark Neufang), Marien Hospital Herne Ruhr

University Bochum (Timm Westhoff), Tropical Clinic Paul-Lechler Hospital Tuebingen (Claudia Raichle), Hacettepe University (Murat Akova), University Hospital Duesseldorf (Bjoern-Erik Jensen), Elbland Hospital Riesa (Joerg Schubert), Center for Infectiology Prenzlauer Berg Berlin (Stephan Grunwald), University Hospital Schleswig-Holstein Kiel (Anette Friedrichs), University Hospital of Giessen and Marburg (Janina Trauth), University Hospital Dresden (Katja de With), Clinic Munich (Wolfgang Guggemos), Hospital Braunschweig (Jan Kielstein), Agaplesion Diakonie Hospital Rotenburg (David Heigener), Hospital Fulda (Philipp Markart), University Hospital Saarland (Robert Bals), Petrus Hospital Wuppertal (Sven Stieglitz), Elisabeth Hospital Essen (Ingo Voigt), Richmond Research Institute (Jorg Taubel), Malteser Hospital St. Franziskus Flensburg (Milena Milovanovic).

The LEOSS study infrastructure group: Jörg Janne Vehreschild (Goethe University Frankfurt), Susana M. Nunes de Miranda (University Hospital of Cologne), Carolin E. M. Jakob (University Hospital of Cologne), Melanie Stecher (University Hospital of Cologne), Lisa Pilgram (Goethe University Frankfurt), Nick Schulze (University Hospital of Cologne), Sandra Fuhrmann (University Hospital of Cologne), Max Schons (University Hospital of Cologne), Annika Claßen (University Hospital of Cologne), Bernd Franke (University Hospital of Cologne) und Fabian Praßer (Charité, Universitätsmedizin Berlin).

No specific financial support was received for this study. The LEOSS registry was supported by the German Centre for Infection Research (DZIF) and the Willy Robert Pitzer Foundation.

REFERENCES

T. B. Alakus and I. Turkoglu. 2020. Comparison of deep learning approaches to predict COVID-19 infection. *Chaos, Solitons & Fractals* 140 (2020), 110120.

M. AlJame, I. Ahmad, A. Imtiaz, and A. Mohammed. 2020. Ensemble learning model for diagnosing COVID-19 from routine blood tests. *Informatics in Medicine Unlocked* 21 (2020), 100449.

N. Alballa and I. Al-Turaiki. 2021. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. *Informatics in Medicine Unlocked* 2021; Online First.

M. Avetisian, I. Burenko, K. Egorov, V. Kokh, A. Nesterov, A. Nikolaev, A. Ponomarchuk, E. Sokolova, A. Tuzhilin, and D. Umerenkov. 2021. CoRSAI: A system for robust interpretation of CT scans of COVID-19 patients using deep learning. *ACM Transactions on Management Information Systems* 12, 4 (2021), 28.

C. C. Bartenschlager, R. Frey, M. Freitag, J.-M. Classen, H. Messmann, J. O. Brunner, S. Temizel, and C. Römmele. 2022. Managing hospital visitor admission during Covid-19: A discrete-event simulation by the data of a German University hospitalIn. *Proceedings of the Hawaii International Conference on System Sciences*.

D. Bertsimas, A. Borenstein, L. Mingardi, O. Nohadani, A. Orfanoudaki, B. Stellato, H. Wiberg, P. Sarin, D. J. Varelmann, V. Estrada, C. Macaya, and I. J. N. Gil. 2021a. Personalized prescription of ACEI/ARBs for hypertensive COVID-19 patients. *Health Care Management Science* 24 (2021a), 339–355.

D. Bertsimas, L. Boussioux, R. Cory-Wright, et al. 2021b. From predictions to prescriptions: A data-driven response to COVID-19. *Health Care Management Science* 24 (2021b), 253–272.

D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza. 2020. Detection of COVID-19 infection from routine blood exams with machine learning: A feasibility study. *Journal of Medical Systems* 44 (2020), 135.

A. S. Britto, R. Sabourin, and L. E. S. Oliveira. 2014. Dynamic selection of classifiers—A comprehensive review. *Pattern Recognition* 47 (2014), 3665–3680.

F. Cabitza, A. Campagner, D. Ferrari, C. Di Resta, D. Ceriotti, E. Sabetta, A. Colombini, E. De Vecchi, G. Banfi, M. Locatelli, and A. Carobene. 2020. Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. *Clinical Chemistry and Laboratory Medicine* 59 (2020), 421–431.

K. Chadaga, C. Chakraborty, S. Prabhu, S. Umakanth, V. Bhat, and N. Sampathila. 2022. Clinical and laboratory approach to diagnose COVID-19 using machine learning. *Interdisciplinary Sciences: Computational Life Sciences* 14 (2022), 452–470.

V. A. de Freitas Barbosa, J. C. Gomes, M. A. de Santana, J. E. de A. Albuquerque, R. Gomes de Souza, R. E. de Souza, and W. Pinheiro dos Santos. 2022. Heg.IA: An intelligent system to support diagnosis of Covid-19 based on blood tests. *Research on Biomedical Engineering* 38 (2022), 99–116.

C. Feng, L. Wang, X. Chen, Y. Zhai, F. Zhu, H. Chen, Y. Wang, X. Su, S. Huang, L. Tian, W. Zhu, W. Sun, L. Zhang, Q. Han, J. Zhang, F. Pan, L. Chen, Z. Zhu, H. Xiao, Y. Liu, G. Liu, W. Chen, and T. Li. 2020. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected covid-19 pneumonia in fever clinics. medRxiv 2020.

D. Goodman-Meza, A. Rudas, J. N. Chiang. 2020. A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity. *PloS One* 15 (2020), e0239474.

C. E. M. Jakob, F. Kohlmayer, T. Meurers, J. J. Vehreschild, and F. Prasser. 2020. Design and evaluation of a data anonymization pipeline to promote Open Science on COVID-19. *Scientific Data* 7 (2020), 435.

M. Kahn, L. Schuierer, C. Bartenschlager, S. Zellmer, R. Frey, M. Freitag, C. Dhillon, M. Heier, A. Ebigbo, C. Denzel, S. Temizel, H. Messmann, M. Wehler, R. Hoffmann, E. Kling, and C. Römmele. 2021. Performance of antigen testing for diagnosis of COVID-19 – a direct comparison of a lateral flow device to nucleic acid amplification-based tests. *BMC Infectious Diseases* (2021), 798.

M. Kukar, G. Gunčar, T. Vovko, S. Podnar, P. Černelč, M. Brvar, M. Zalaznik, M. Notar, S. Moškon, and M. Notar. 2021. COVID-19 diagnosis by routine blood tests using machine learning. *Scientific Reports* 11 (2021), 1–9.

L. Kuncheva. 2014. *Combining Pattern Classifiers. Methods and Algorithms*. Wiley 2014.

S. Lalmuanawma, J. Hussain, and L. Chhakchhuak. 2020. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons and Fractals* 139 (2020), 110059.

N. Lawrence, M. Seeger, and R. Herbrich. 2002. Fast sparse gaussian process methods: The informative vector machine. In *Proceedings of the 16th Annual Conference on Neural Information Processing Systems*.

A. López Pineda, Y. Ye, S. Visweswaran, G. F. Cooper, M. M. Wagner, and F. Tsui. 2015. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *Journal of Biomedical Informatics* 58 (2015), 60–69.

G. Melman, A. Parlikad, and E. Cameron. 2021. Balancing scarce hospital resources during the COVID-19 pandemic using discrete-event simulation. *Health Care Management Science* (2021), 356–374.

T. O'Connor, S. Santaniello, and B. Javidi. 2022. COVID-19 detection from red blood cells using highly comparative time-series analysis (HCTSA) in digital holographic microscopy. *Optics Express* 30 (2022), 1723–1736.

G. Parmigiani and L. Inoue. 2009. *Decision Theory: Principles and Approaches*. John Wiley & Sons (2009).

S. Babaei Rikan, A. Sorayaie Azar, A. Ghafari, J. Bagherzadeh Mohasefi, and H. Pirnejad. 2022. COVID-19 diagnosis from routine blood tests using artificial intelligence techniques. *Biomedical Signal Processing and Control* 72 (2022), 103263.

P. Schwab, A. D. Schütte, B. Dietz, and S. Bauer. 2020. Predcovid-19: A Systematic Study of Clinical Predictive Models for Coronavirus Disease 2019. arXiv 2020; 2005.08302.

A. Scohy, A. Anantharajah, M. Bodéus, B. Kabamba-Mukadi, A. Verroken, and H. Rodriguez-Villalobos. 2020. Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis. *Journal of Clinical Virology* 129 (2020), 104455.

S. Shoer, T. Karady, A. Keshet, S. Shilo, H. Rossman, A. Gavrieli, T. Meir, A. Lavon, D. Kolobkov, I. Kalka, A. Godneva, O. Cohen, A. Kariv, O. Hoch, M. Zer-Aviv, N. Castel, C. Sudre, A. E. Zohar, A. Irony, T. Spector, B. Geiger, D. Hizi, V. Shalev, R. Balicer, and E. A. Segal. 2020. A prediction model to prioritize individuals for SARS-CoV-2 test built from national symptom surveys. *Clinical and Translational Resource and Technology Insights* 2, 2 (2020), 196–208.

R. Thell, J. Zimmermann, M. Szell, S. Tomez, P. Eisenburger, M. Haugk, A. Kreil, A. Spiel, A. Blaschke, A. Klicpera, O. Janata, W. Krugluger, C. Sebesta, H. Herkner, and B. Laky. 2021. Standard blood laboratory values as a clinical support tool to distinguish between SARS-CoV-2 positive and negative patients. *Scientific Reports* 11 (2021), 1–9.

L. M. Thimoteo, M. M. Vellasco, J. Amaral, et al. 2022. Explainable artificial intelligence for COVID-19 diagnosis through blood test variables. *Journal of Control, Automation and Electrical Systems* 33 (2022), 625–644.

I. Torres, S. Poujois, E. Albert, J. Colomina, and D. Navarro. 2021. Evaluation of a rapid antigen test (Panbio^{TM} COVID-19 Ag rapid test device) for SARS-CoV-2 detection in asymptomatic close contacts of COVID-19 patients. *Clinical Microbiology and Infection* 27, 4 (2021), 636.e1–636.e4.

T. Tschoellitsch, M. Dünser, C. Böck, K. Schwarzbauer, and J. Meier. 2020. Machine learning prediction of sars-cov-2 polymerase chain reaction results with routine blood tests. *Laboratory Medicine* 52, 2 (2020), 146–9.

S. Visa and A. Ralescu. 2005. Issues in mining imbalanced data sets - a review paperIn. *Proceedings of the 16th Midwest Artificial Intelligence and Cognitive Science Conference*.

S. Wang, X. Zhang, and Y.-D. Zhang. 2021. DSSAE: Deep stacked sparse autoencoder analytical model for COVID-19 diagnosis by fractional fourier entropy. *ACM Transactions on Management Information Systems* 13, 1 (2021), 2.

L. Wynants, B. Van Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, M. M. J. Bonten, D. L. Dahly, J. A. A. Damen, T. P. A. Debray, V. M. T. de Jong, M. De Vos, P. Dhiman, M. C. Haller, M. O. Harhay, L. Henckaerts, P. Heus, M. Kammer, N. Kreuzberger, A. Lohmann, K. Luijken, J. Ma, G. P. Martin, D. J. McLernon, C. L. Andaur Navarro, J. B. Reitsma, J. C. Sergeant, C. Shi, N. Skoetz, L. J. M. Smits, K. I. E. Snell, M. Sperrin, R. Spijker, E. W. Steyerberg, T. Takada, I. Tzoulaki, S. M. J. van Kuijk, B. van Bussel, I. C. C. van der Horst, F. S. van Royen, J. Y. Verbakel, C. Wallisch, J. Wilkinson, R. Wolff, L. Hooft, K. G. M. Moons, and M. van Smeden. 2020. Prediction models for diagnosis and prognosis of Covid-19 infection: systematic review and critical appraisal. *BMJ* 369 (2020), 1–16.

R. K. Zhang, Q. Xiao, S. L. Zhu, H. Y. Lin, and M. Tang. 2022. Using different machine learning models to classify patients into mild and severe cases of COVID-19 based on multivariate blood testing. *Journal of Medical Virology* 94 (2022), 357–365.

M. Zokaeinikoo, P. Kazemian, P. Mitra, and S. Kumara. 2021. AIDCOV: An interpretable artificial intelligence model for detection of COVID-19 from chest radiography images. *ACM Transactions on Management Information Systems* 12, 4 (2021), 29.