Informatics in Medicine Unlocked 21 (2020) 100449

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Ensemble learning model for diagnosing COVID-19 from routine blood tests

Maryam AlJame [a], Imtiaz Ahmad [a,*], Ayyub Imtiaz [b], Ameer Mohammed [a]

[a] Computer Engineering Department, Kuwait University, Kuwait
[b] College of Medicine, Kuwait University, Kuwait

ARTICLE INFO

Keywords:
COVID-19
Routine blood tests
Diagnostic model
Ensemble
Machine learning

ABSTRACT

Background and objectives: The pandemic of novel coronavirus disease 2019 (COVID-19) has severely impacted human society with a massive death toll worldwide. There is an urgent need for early and reliable screening of COVID-19 patients to provide better and timely patient care and to combat the spread of the disease. In this context, recent studies have reported some key advantages of using routine blood tests for initial screening of COVID-19 patients. In this article, first we present a review of the emerging techniques for COVID-19 diagnosis using routine laboratory and/or clinical data. Then, we propose ERLX which is an ensemble learning model for COVID-19 diagnosis from routine blood tests.

Method: The proposed model uses three well-known diverse classifiers, extra trees, random forest and logistic regression, which have different architectures and learning characteristics at the first level, and then combines their predictions by using a second level extreme gradient boosting (XGBoost) classifier to achieve a better performance. For data preparation, the proposed methodology employs a KNNImputer algorithm to handle null values in the dataset, isolation forest (iForest) to remove outlier data, and a synthetic minority oversampling technique (SMOTE) to balance data distribution. For model interpretability, features importance are reported by using the SHapley Additive exPlanations (SHAP) technique.

Results: The proposed model was trained and evaluated by using a publicly available data set from Albert Einstein Hospital in Brazil, which consisted of 5644 data samples with 559 confirmed COVID-19 cases. The ensemble model achieved outstanding performance with an overall accuracy of 99.88% [95% CI: 99.6–100], AUC of 99.38% [95% CI: 97.5–100], a sensitivity of 98.72% [95% CI: 94.6–100] and a specificity of 99.99% [95% CI: 99.99–100].

Discussion: The proposed model revealed better performance when compared against existing state-of-the-art studies (Banerjee et al., 2020; de Freitas Barbosa et al., 2020; de Moraes Batista et al., 2020; Soares et al., 2020) [3,22,56,71] for the same set of features employed by them. As compared to the best performing Bayes Net model (de Freitas Barbosa et al., 2020) [22] average accuracy of 95.159%, ERLX achieved an average accuracy of 99.94%. In comparison with AUC of 85% reported by the SVM model (de Moraes Batista et al., 2020) [56], ERLX obtained AUC of 99.77% in addition to improvements in sensitivity, and specificity. As compared with ER-COV model (Soares et al., 2020) [71] average sensitivity of 70.25% and specificity of 85.98%, ERLX model achieved sensitivity of 99.47% and specificity of 99.99%. The ERLX model obtained a considerably higher score as compared with ANN model (Banerjee et al., 2020) [3] in all performance metrics. Therefore, the model presented is robust and can be deployed for reliable early and rapid screening of COVID-19 patients.

1. Introduction

The pandemic of novel coronavirus disease 2019 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is spreading rapidly all over the world and resulting in a massive death toll. As of July 11, 2020, more than 12 million confirmed cases have been reported in 216 countries and territories with more than 500,000 deaths due to this pandemic [75]. The COVID-19 pandemic has impacted virtually every aspect of human society in all geographic locations. Therefore, tremendous global efforts have been made for its quick and precise early detection and timely treatment to avoid the spread of the virus. The current gold standard test in COVID-19 diagnosis is the Reverse

* Corresponding author.
E-mail addresses: maryam.aljame@eng.ku.edu.kw (M. AlJame), imtiaz.ahmad@ku.edu.kw (I. Ahmad), ayyub@hsc.edu.kw (A. Imtiaz), ameer.mohammed@ku.edu.kw (A. Mohammed).

https://doi.org/10.1016/j.imu.2020.100449
Received 11 August 2020; Received in revised form 28 September 2020; Accepted 7 October 2020
Available online 20 October 2020

Transcription Polymerase Chain Reaction (RT-PCR) with DNA sequencing and identification [16,17], but it is time consuming, costly, needs specialized equipment and has a roughly 20% false-negative rate [47]. Moreover, there is a shortage of the availability of RT-PCR test kits globally. Tests based on IgM/IgG antibodies have been used, but their drawbacks are a very low sensitivity (18.8%) and specificity (77.8%) in diagnosing COVID-19 during its early phase [13]. Therefore, other testing methods are imperative which are less expensive and more accessible.

There has been a profound interest in exploring the potential of machine learning (ML) tools to combat the COVID-19 pandemic by contributing to disease diagnosis and prognosis, forecasting, prevention, treatment and management, disease surveillance and antiviral drug discovery [12,45]. In this article, we only address the use of ML techniques to help medical specialists in the accurate and reliable early detection of COVID-19. In this context, ML based medical imaging such as computed tomography (CT) scans and chest X-rays images [74], have demonstrated promising results to complement the conventional diagnostic techniques of COVID-19 such as molecular biology (RT-PCR) and immune (IgM/IgG) assays. Due to the immense traction in use of machine learning for biomedical image analysis for COVID-19, there have been several recent reviews with exclusive focus on X-rays or CT scans [18,28,58,68]. However, CT scans cannot be utilized for screening tasks because of the radiation doses, the relative low number of devices available, and the related high costs. Recently, some research groups have advocated point-of-care ultrasound imaging for being non-invasive and radiation-free for COVID-19 detection specially for children and pregnant women [7,18]. Other research groups have explored the opportunities of speech and sound analysis for COVID-19 detection [34,65] by using ML systems.

Recently, a number of clinical studies have revealed [5,19,24,41,61,70] that blood parameters of COVID-19 patients exhibit considerable change and the identification of these parameters can play a key role in the initial screening for COVID-19 [2,21,40,41,69]. Initial screening process provides an early probabilistic indication of the presence of the disease, while diagnosing confirms the presence/absence of disease. As reported in the study [27], it is difficult for even an experienced physician to extract all the information contained in routine blood tests. However, machine learning algorithms can learn and differentiate among various patterns observed in the routine blood test parameters. Therefore, some initial efforts have started in developing machine learning algorithms for the identification of COVID-19 from routine blood samples [11,22,43,56,66,71,79] as explained in details in related work section. However, the research work is still in the infancy stage.

The blood tests based ML framework for COVID-19 early detection will provide a fast, easy to use, more accessible and less expensive alternative to costly and time consuming tests, such as imaging based studies and RT-PCR. Such a system will have a major impact in developing and low-income countries which suffer from a shortage of testing kits, laboratory supplies and specialized centers for PCR related exams. Other key advantages of such a quick and inexpensive system include smooth patient flow and speeding up results for potentially infected patients, and thus curbing the pandemic [11,66,71].

The aims of this article are two folds. First, we present a review of emerging COVID-19 automatic diagnosing models using routine laboratory and/or clinical data. No such review exists to date for these techniques. We analyze these techniques from the perspective of dataset, feature selection, machine learning classifiers employed and performance evaluation. It is envisioned that this review will provide readers with an overview of the state-of-the-art techniques for COVID-19 detection from routine laboratory tests and will inspire researchers in developing better models to combat COVID-19. Then, we propose an ensemble learning model named ELRX for initial screening of COVID-19 from routine blood tests. As compared with existing techniques, the ELRX model utilizes two levels of classifiers for enhanced performance. The first level diverse classifiers include random forest, extra trees and logistic regression whose outputs are fed to the second level extreme gradient boosting classifier. In addition, the ERLX model uses KNNImputer algorithm to handle null values in the dataset, iForest to remove outlier data, and synthetic minority oversampling technique (SMOTE) to balance data distribution. Furthermore, features importance are reported by using the SHapley Additive exPlanations (SHAP) technique for the model interpretability requirement in medical settings. The ERLX model is robust and achieved considerable improvement in performance metrics for diagnosing COVID-19 when compared against existing state-of-the-art studies for a publicly available dataset from Albert Einstein Hospital in Brazil.

The rest of the paper is structured as follows: Section 2 reviews the related work in the field. The proposed design methodology is detailed in Section 3 and experimental results are discussed in Section 4. Finally, concluding remarks are made in Section 5.

2. Related work

This section presents review of almost all the machine learning techniques reported at the time of writing this article for early detection of COVID-19 based on routine laboratory tests and/or clinical data.

Wu et al. [79] reported the first study by employing a random forest (RF) classification algorithm [10] to detect COVID-19 from blood tests. The authors collected a dataset of 253 blood samples with 105 confirmed COVID-19 cases from different hospitals in Lanzhou, China. The authors identified the 11 key features out of the 49 features in the given blood samples. The model trained with 11 key features resulted in a sensitivity of 95.12%, a specificity of 96.97% and an overall accuracy of 96.95%. However, the performance of the algorithm was a bit lower on external blood samples. Furthermore, the dataset has a quite high ratio of positives (41.5%) and performance will suffer for low ratio of COVID-19 positives in the dataset.

Wu et al. [77] designed a least absolute shrinkage and selection operator (LASSO) logistic regression (LR) model [53] based on blood results for COVID-19 detection. The dataset consisted of 110 patient blood samples from Tongji Hospital, China, where 80% of the data was used for training the model and the remaining 20% for validation. The blood features were reduced from 47 to 15 by removing non-significant features by applying the maximum relevance minimum redundancy (mRMR) algorithm [60]. The application of LASSO further reduced the features to 7 for training the model. The model achieved 98% [93%, 100%] sensitivity and 91% [84%, 99%] specificity in COVID-19 prediction. However, the model considered few features and the data size is very small from a single center to be applicable in real settings.

Yan et al. [80] used an extreme gradient boosting (XGBoost) machine learning model [15] to predict the survival rate of critically ill patients with COVID-19 infection based on epidemiological and clinical data. The authors tested single-tree and multi-tree variants of XGBoost algorithm with data of 375 patients from Tongji Hospital of Wuhan. The authors identified the three key clinical features (lactic dehydrogenase, lymphocyte and C-reactive protein) to quickly assess the risk of death. Although the study is useful, the dataset is of a very small size and from a single source.

Feng et al. [20] developed an innovative predictive model for an early identification of COVID-19 on admission. For the model, four different classifiers were chosen including LR with LASSO, LR with Ridge regularization [53], decision trees (DT) [63] and Adaboost algorithms. The key strength of the model lies in the selection of candidate features which included 2 variables of demographic information, 4 variables of vital signs, 20 variables of blood routine values, 17 variables of clinical signs and symptoms, 2 infection-related biomarkers and 1 other variable related to admission. The dataset from 132 patients (26 positives) with the required features were collected from the First Medical Center, General Hospital of People's Liberation Army, Beijing, China. With LASSO, only 18 features among the above 46 features were selected to train the model. Based on the results, LR with LASSO

achieved the best performance, with an area under curve (AUC) of 93.8%, a sensitivity of 100% and a specificity of 77.8%. The study has several strengths such as integration of the most routinely available data features on admission for accurately identifying the COVID-19. However, because of the very small data size from a single center, further external validation is required for the success of the model in real settings.

Soares et al. [71] designed a ML based framework called ER-CoV based on blood exams to perform initial screening of suspect COVID-19 patients. The proposed model utilized a combination of three techniques: support vector machine (SVM) [8], SMOTEBoost [14] and ensemble [62]. To improve the classification performance of SVM due to the small number of positive samples in the dataset, oversampling was performed using SMOTEBoost. In addition, ensemble methods were employed that combined predictions from 10 SVM-based SMOTEBoost models and the final prediction was based on the average probability from all 10 models. The ER-CoV was evaluated by using a publicly available dataset from Albert Einstein Hospital in Brazil, which consisted of 599 blood samples with 81 confirmed COVID-19 cases. From the 108 features of dataset samples, only 16 blood features were selected to train and test the model. The ER-CoV model achieved a sensitivity of 70.25%, a specificity of 85.98% and an AUC of 86.78%, respectively.

Banerjee et al. [3] tested four machine learning models based on blood tests to perform initial screening of suspect COVID-19 patients. The models utilized were RF [38], artificial neural network (ANN) [29], LR [31] and Lasso-elastic-net regularized generalized linear network (GLMNET). The models were evaluated by using the dataset from Albert Einstein Hospital in Brazil, which consisted of 598 blood samples with 81 confirmed COVID-19 cases. From the 108 features of dataset samples, only 14 blood features were selected to train and test the model. The models achieved an accuracy of 81–87%, a sensitivity of 43–65%, a specificity of 81–91% and an AUC of 80–86%, respectively. The authors also reported results with only 4 key features. However, more testing and validation is needed to evaluate the models in clinical settings.

Brinati et al. [11] investigated different classes of machine learning classifiers for COVID-19 detection from routine blood samples. The authors considered these models: DT [63]; extremely randomized trees (ET) [25]; K-nearest neighbors (KNN) [1]; LR [31]; Naive Bayes (NB) [46]; RF [38] and SVM [64]. In addition, the authors modified the RF algorithm to 3-way RF classifier in order to improve accuracy. A dataset of 279 routine blood samples was obtained from patients admitted to San Raffaele Hospital, Milan, Italy. The dataset contained 177 confirmed COVID-19 samples and only 15 features of blood samples were considered. The models were trained and evaluated resulting in an accuracy of 82%–86% and a sensitivity of 92%–95%. The RF model was the best performing classifier. Some of the limitations of their technique are the relatively small data size, a single source of data with a high ratio of positives (63.44%) and a limited set of blood sample features.

Batista et al. [56] reported a study to predict COVID-19 diagnosis using ML algorithms from emergency care blood samples. Five well-known machine learning models (neural networks (NN), RF, gradient boosting trees (GBT), LR and SVM) [23,29,31,38,64] were utilized for classification. The authors collected the dataset from Albert Einstein Hospital in Brazil, which consisted of 235 blood samples with 102 confirmed COVID-19 cases. From the blood samples, only 15 features were selected to train and test the model. The best predictive performance was obtained by the SVM algorithm with a sensitivity of 68%, a specificity of 85% and an AUC of 85%.

Bao et al. [4] investigated RF [38] and SVM [64] models for the early detection of COVID-19 based on routine blood tests. Dataset of 294 blood samples (including 208 positives) were collected from Wuhan Union Hospital, and Kunshan People's Hospital, in Kunshan, China. Three types of classification tasks were performed (moderate vs viral, severe vs. viral and severe vs. moderate). A maximum of fifteen blood features was selected to train the models. The SVM-based classifier performed the best with an accuracy of 84%, a sensitivity of 88%, a specificity of 80%, and a precision of 92%.

Kukar et al. [43] selected the extreme gradient boosting (XGBoost) machine learning model [15] instead of RF or deep neural networks (DNN) because of its higher performance, less computational resources requirement and its intrinsic ability to handle missing data. The dataset of 5333 blood samples with various bacterial and viral infections, and 160 confirmed COVID-19 samples were collected from University Medical Center Ljubljana, Slovenia. Out of the 117 dataset features, only the prominent 35 features were selected for the model. The proposed model was trained, tested and cross-validated with the given dataset. The results revealed a sensitivity of 81.9%, a specificity of 97.9% and an AUC of 97%. However, the dataset for building the model has a very low ratio of positives (2.91%), which makes it difficult to assess the quality of results.

Barbosa et al. [22] created a cheap COVID-19 detection system from routine blood samples by utilizing several ML classifiers such as multilayer perceptron [29], SVM [64], RT, RF [10,25,38], bayesian networks (BN) and NB [29,46]. The authors collected the dataset from Albert Einstein Hospital in Brazil, which consisted of 5644 data samples with 559 confirmed COVID-19 cases. SMOTE [1] was used for oversampling to improve the performance of their models due to small number of positive samples in the dataset. From the dataset with 108 features, Particle Swarm Optimization (PSO) [39] and Evolutionary Search (ES) [49] optimization algorithms were employed to reduce the features to 63 and 62, respectively. In order to further reduce cost and time of blood tests, features were reduced to 24 by hand to train and test their models. Results achieved high classification performance with 95.159% of an overall accuracy, a sensitivity of 96.8%, a precision of 93.8% and a specificity of 93.6%. Experiments revealed that BN had superior performance with respect to other models.

Yang et al. [81] constructed machine learning models incorporating patient demographic features (age, sex, gender, race) with 27 blood test features for COVID-19 detection. The classifiers considered were LR [31], DT [63], RF [38] and gradient boosted decision trees (GBDT) [23]. These models were trained and tested with a dataset of 3346 patients (1394 positives) collected from New York Presbyterian Hospital/Weil Cornell Medicine. However, for validation dataset of 1822 patients (549 positives) collected from New York Presbyterian Hospital/Lower Manhattan was used. The GBDT classifier reported the best results among the four classifiers with a sensitivity of 75.8%, a specificity of 80.2% and an AUC of 85.3%. However, the study was developed with a control group and cannot be generalized.

Sun et al. [57] employed five types of classification models: SVM [64], LR [31], DT [63], RF [38] as well as deep neural network (DNN) [29] to identify the best model for diagnosing the early COVID-19 infection with significant clinical value. The authors collected clinical data of 912 patients (361 positives) from 18 hospitals in Zhejiang Province. Each patient's clinical record contains 31 features including gender, age, coexisting diseases, epidemiological information, laboratory tests, clinical symptoms and imaging findings. The authors selected 10 features from the 31 features mentioned above. The LR classification model resulted in the best performance among five classifiers with an accuracy of 91%, a sensitivity of 87%, an AUC of 97.1%, and a specificity of 95%. The study revealed that the lack of epidemiological information greatly affected the accuracy, specificity and sensitivity of the model. However, more testing and validation is required for general clinical settings.

Joshi et al. [36] developed a LR [31] model to predict COVID-19 from 3 blood count components (absolute neutrophil count, absolute lymphocyte count, and hematocrit) and patient sex. The model was trained with dataset of 390 patient samples (with 33 confirmed positive) collected from Stanford Health Care. The trained model was validated with datasets from diverse locations including Seattle, Washington, Northern California, Chicago, and South Korea. The model achieved a sensitivity of 86–93% and a specificity of 35–55%. Although, the model was validated with diverse patient populations, the study only selected

few blood features.

Li et al. [48] reported a study aimed at finding the correlation among clinical variables (signs, symptoms and laboratory tests variables), clustering the patients into groups based on a similarity distance metric and to develop a COVID-19 diagnosis model based on important clinical variables. The authors compiled the key 42 clinical variables from 151 published studies comprising 413 patients. Then associations among variables were discovered by computing correlations among these 42 variables using different statistical tests based on the type of variable. A self-organizing-map (SOM) machine learning [42] algorithm was applied on patients to form clusters based on Euclidian distance measure. Only 27 clinical variables were found to be important in decision making. Finally, a XGBoost algorithm [15] based diagnostic model was trained using only 19 clinical variables with a dataset from multiple sources to differentiate between COVID-19 and influenza patients. The model achieved a sensitivity of 92.5% and a specificity of 97.9%. Despite the usefulness of the study in revealing associations among clinical variables, the sample size was too small for the model to be generalized.

Bayat et al. [6] developed a RF [10,38] based COVID-19 predictive model by using a combination of vital signs with common laboratory tests. The dataset consisting of 68 features of 5002 individuals with 1079 confirmed COVID-19 was collected from different Veterans Health Administration sites across USA. The authors applied pairwise correlation among features to select 54 most significant ones. The model was trained with 40–54 features resulting in an accuracy of 88.3%, a sensitivity of 83.4%, a specificity of 89.8% and an AUC of 92.8%. The authors also identified the minimum 9 key features that can produce an acceptable level of accuracy for the model. The proposed predictive model also has the ability to discriminate between patients with COVID-19 versus other respiratory virus infections such as Influenza, Respiratory Syncytial Virus, and seasonal human coronaviruses. One of the limitations of the model was that it was trained on older and mostly male patients.

Langer et al. [44] tested a number of ML models including variants of ANNs [29], DT [63], RF [38] and LR [31] for the early diagnosis of COVID-19 patients in emergency departments by using basic clinical, radiological and routine laboratory data. The dataset consisting of 74 features of 199 individuals with 127 confirmed COVID-19 was collected from one of the main hospitals in Milan, Italy. The authors applied a feature selection algorithm to select 42 most significant features of the dataset among 74 features to train the machine learning models. The best performing architecture was one of the ANNs model, which achieved an accuracy of 91.4%, a sensitivity of 94.1% and a specificity of 88.5%. The usefulness of the study lies in the good selection of the clinical data which is generally available in the emergency departments to make a quick decision to prevent the spread of the disease. However, the study suffers from some limitations such as a single-center study, a very small sample size and lack of some clinical and epidemiological data which may be valuable for improving the accuracy of the model.

Soltan et al. [72] investigated two early-detection predictive models to identify COVID-19 using routinely collected data. One of the models classify the patients at emergency departments as being COVID-19 positive or negative, while the second model determines whether the COVID-19 positive patients will be admitted to hospital or not. The authors employed LR [31], RF [38] and XGBoost [15] classifiers for their predictive models. A huge dataset of clinical variables consisting of laboratory blood tests, point-of-care blood gas readings, changes in laboratory blood results from pre-admission baseline, vital signs and Charlson Comorbidity Index (CCI) from Oxford University Hospitals, UK were analyzed. XGBoost classifier demonstrated the highest predictive performance for COVID-19 with an accuracy of 92.3%, a sensitivity of 77.4% and a specificity of 95.7%. A limitation of the study is the lack of diversity of the data source.

Schwab et al. [66] conducted an important study using machine learning models based on routinely collected clinical data to classify patients into four classes: (i) COVID-19 negative, (ii) COVID-19 positive which require swab test, (iii) require hospitalization, and (iv) require intensive care. Five different classification models were chosen including LR [31], NN [29], RF [38], SVM [64], and gradient boosting (XGB) [15,23]. The dataset from Albert Einstein Hospital in Brazil, which consisted of 5644 blood samples with 279 confirmed COVID-19 cases, were used. In the data clean-up stage, some features were removed resulting in 97 features to train and test the model. Based on the performance metrics (AUC, sensitivity and specificity) thresholds, the models were successful in classifying the patients into their proper class. The authors highlighted important clinical features appropriate for each classification class. However, the experimental evaluation was based on data collected from a single study site, and its results may therefore not generalize to settings with significantly different patient populations, admission criteria, and testing guidelines.

A summary of the related techniques with their key characteristics is given in Table 1. Most of the techniques utilize proprietary datasets and these datasets are generally small in size. The two most popular classifiers are the LR and RF, which is the reason for their selection in the proposed model. The feature set varies from one technique to another and so are the performance evaluations. As one can observe from Table 1, most of the techniques utilize single level classifiers with the exception of [71], which employed an ensemble method to combine predictions from 10 SVM models and the final prediction was based on the average probability from all 10 models. The existing studies did not exploit the opportunities of utilizing ensemble models to enhance COVID-19 prediction models based on routine laboratory and clinical data, which inspired us to explore this fertile area of research.

3. Design methodology

This section provides detailed explanation of the methodology used to develop the proposed classification model. The first subsection describes the dataset used and the selected features. The data preparation process is described next. The last subsection, demonstrates the implementation details of the proposed model.

3.1. Dataset description and features selection

The dataset used in our study was obtained from 5644 patients admitted to the Albert Einstein Israelita Hospital located in Saulo Paulo, Brazil [37]. Kaggle made the dataset available for public access. The dataset was collected from the March 28, 2020 to April 3, 2020, with more than 100 laboratory tests including blood tests, urine tests, SARS-CoV-2 test, rt-PCR test, influenza A viruses presence, to name a few. The clinical data were normalized to have a mean of zero and a unit standard deviation. Among 5644 patients, 559 patients were infected with SARS-Cov2. The Albert Einstein dataset has SARS-Cov2 attribute which gives COVID-19 diagnosis as string values: negative and positive. In the proposed model those values have been converted to integers where zero is assigned to negative cases and one for positive cases.

Feature selection is an important process in building a machine learning model. Selecting the appropriate features helps in reducing data redundancy and avoiding noisy data, thus, improving model performance. There are 108 features in the Albert Einstein dataset, in the proposed model 18 features have been selected based on their importance in detecting COVID-19 based on clinical studies [19,40,41,70] as well as reported by earlier machine learning prediction models [3,22,56,71]. Those features are: Hemoglobin, Platelets, Leukocytes, Lymphocytes, Basophils, Eosinophils, Monocytes, Neutrophils, Age, Urea, C reactive Protein, Creatinine, Potassium, Sodium, Alanine transaminase, Aspartate transaminase, International normalized ratio (INR), Albumin, D-Dimer, and Prothrombin time. However, the proposed model can handle any set of features.

Table 1
Comparison of related techniques.

Ref.	Dataset Source	Dataset Size (COVID-19)	Total Features (Selected)	Model Used	Accuracy	Sensitivity	Specificity
[79]	Hospitals, Lanzhou, China	253 (105)	49 (11)	RF	96.95%	95.12%	96.97%
[77]	Tongji Hospital of Wuhan, China	110 (–)	47 (7)	LASSO-LR	–	98%	91%
[80]	Tongji Hospital of Wuhan, China	375 (201)	300 (3)	XGBoost	–	83%	–
[20]	First Medical Center, Beijing, China	132 (26)	46 (18)	LASSO-LR, DT, Adaboost	–	100%	77.8%
[71]	Albert Einstein Hospital, Brazil	599 (81)	108 (16)	Ensemble of 10 SVM models	–	70.25%	85.98%
[3]	Albert Einstein Hospital, Brazil	598 (81)	108 (14)	RF, LR, GLMNET, ANN	81%–87%	43%–65%	81%–91%
[11]	San Raffaele Hospital, Milan, Italy	279 (177)	- (15)	DT, ET, KNN, LR, NB, RF, SVM	82%–86%	92%–95%	–
[56]	Albert Einstein Hospital, Brazil	253 (102)	108 (15)	NN, RF, GBT, LR, SVM	–	68%	85%
[4]	Hospitals in Wuhan, China	294 (208)	15 (–)	RF, SVM	84%	88%	80%
[43]	University Medical Center, Ljubljana, Slovenia	5333 (160)	117 (35)	XGBoost, RF, DNN	–	81.9%	97.9%
[22]	Albert Einstein Hospital, Brazil	5644 (559)	108 (24)	XMLP, SVM, RT, RF, BN, NB	95.159%	96.8%	93.6%
[81]	New York Presbyterian Hospital/WCM, LMH, USA	3346 (1394) 1822 (549)	685 (33)	RF, LR, DT, GBDT	–	75.8%	80.2%
[57]	Hospitals in Zhejiang, China	912 (361)	31 (10)	LR, DT, RF, SVM. DNN	91%	87%	95%
[36]	Stanford Health Care, CA, USA	390 (31)	- (4)	LR	–	86–93%	35–55%
[48]	–	398 (–)	42 (19)	SOM, XGBoost	–	92.5%	97.9%
[6]	Veterans Health Administration Sites, USA	5002 (1079)	68 (54)	RF	83.3%	83.4%	89.8%
[44]	Hospital in Milan, Italy	199 (127)	74 (42)	ANNs, LR, RF, DT	91.4%	94.1%	88.7%
[66]	Oxford University Hospitals, UK	40732 (437)	74 (–)	RF, LR, XGBoost	92.3%	77.4%	95.7%
[72]	Albert Einstein Hospital, Brazil	5644 (279)	106 (97)	LR, NN, RF, SVM, XGB	–	80%	98%

3.2. Data preparation

The data preparation process consists of three phases: handling missing values, outliers elimination, and balancing the data. To handle missing values the proposed ERLX model utilized KNNImputer from sklearn.impute Python library. The KNNImputer employs k-nearest neighbors to impute missing values using the mean value from nearest neighbors. In the proposed model the number of nearest neighbors is set after tuning it and found that seven neighbors was a suitable candidate.

Outliers detection is a method for detecting anomalies in a given dataset. Anomalies are different than normal records in terms of quantity and quality. Therefore, removing outliers helps in increasing the performance of a classification model. In this article, isolation forest (iForest) [50] has been used to eliminate outliers from the COVID-19 dataset. The iForest is applied from Python scikit-learn library [59] ensemble class. For a given dataset, iForest creates an ensemble of isolation trees (iTrees). To detect outliers, iForest computes the average path lengths for instances on the iTrees, outliers are those instances with short average length. In fact, iForest works efficiently with a small subsample size and a suitable number of trees. These two parameters in scikit-learn iForest are max_samples and n_estimators. After tuning the parameters, n_estimators the base estimators number has been set to 150. Each base estimator is trained with 621 samples by setting max_samples parameter to 621. Another important parameter is contamination that determines outliers proportion in the dataset. In ERLX model, contamination parameter has been set to 7%. Those parameters settings lead to 393 outliers in the COVID-19 dataset.

After outliers removal, the entire dataset was randomly divided into 80% for the training set and 20% for the test set. The next step is balancing the dataset. Imbalanced data has significant impact on classification model, specifically on the training data. In fact, imbalanced data makes the classification model tends to be biased toward the majority class. This increases the occurrence of both false positive and false negative which degrades the performance of the classification model. Therefore, the proposed classification model balances COVID-19 training data to gain performance improvement. The Albert Einstein dataset has 9.9% percent of SARS-CoV-2 positive patients and around 90.1% percent of SARS-CoV-2 negative cases. Hence, the dataset is obviously imbalanced towards negative cases. The proposed model utilizes SMOTE from imblearn Python library. SMOTE balances the data by randomly creating minority class instances, to over sample the minority class.

3.3. Ensemble learning classification model

Stacking [76] is an ensemble machine learning paradigm that combines several classification algorithms to generate a single model. The architecture of a stacking model consists of multiple levels. For instance,

a two levels stacking model has the base level that constructs from different machine learning models. The predictions of the base level are used as inputs of features to the second level. Stacking uses cross-validation to avoid overfitting. All the aforementioned methods make a stacking model robust with improved accuracy. In this context, the ERLX model utilizes vecstack [35] Python package to build a two level stacking model. As illustrated in Fig. 1, the first level consists of three classifiers including extra trees [67], random forest [38], and logistic regression [31]. To develop a robust stacking model, it is recommended to build the first level from algorithms that have different prediction methodology. In spite of the fact that both random forest and extra trees are ensemble learning model based on the decision tree algorithm. There are differences between their core methodology, random forest subsamples the data with replacement. Subsampling data increases data diversity which helps in training the model with highly discriminating training data. On the other hand, extra trees uses the whole data which reduces bias. Indeed, increasing data diversity and reducing bias enhance model performance and makes ERLX more effective. Further, both algorithms split nodes in a different way, random forest splits nodes by finding the optimum split while extra trees split nodes randomly. The random split in extra trees reduces data variance. Thereby, both extra trees and random forest in the first level of ERLX model keep bias and variance in an optimal balance. Another reason for selecting random forest and extra trees for the first level because they are algorithms based on an ensemble of decision trees which makes them perform better than the traditional decision tree algorithm. In addition to that in Ref. [33], results demonstrate that random forest (RF) is a good candidate classifier for diagnosing type 2 diabetes and hypertension. The last algorithm in the first level of ERLX is logistic regression (LR), which demonstrated enhanced performance for various health datasets [26,54,78]. Accordingly, logistic regression (LR) has been selected to build ERLX. The second level in the ERLX model is an extreme gradient boosting (XGBoost) classifier [15]. XGBoost is an ensemble algorithm based on decision trees and gradient boosting framework. XGBoost applies several system optimizations to enhance its performance including parallelization, tree pruning, and cross-validation, to name a few. In addition, XGBoost reduces overfitting by employing regularization technique. For those reasons, using XGBoost to build the second level in ERLX model produces superior results as will be shown in Section 4.

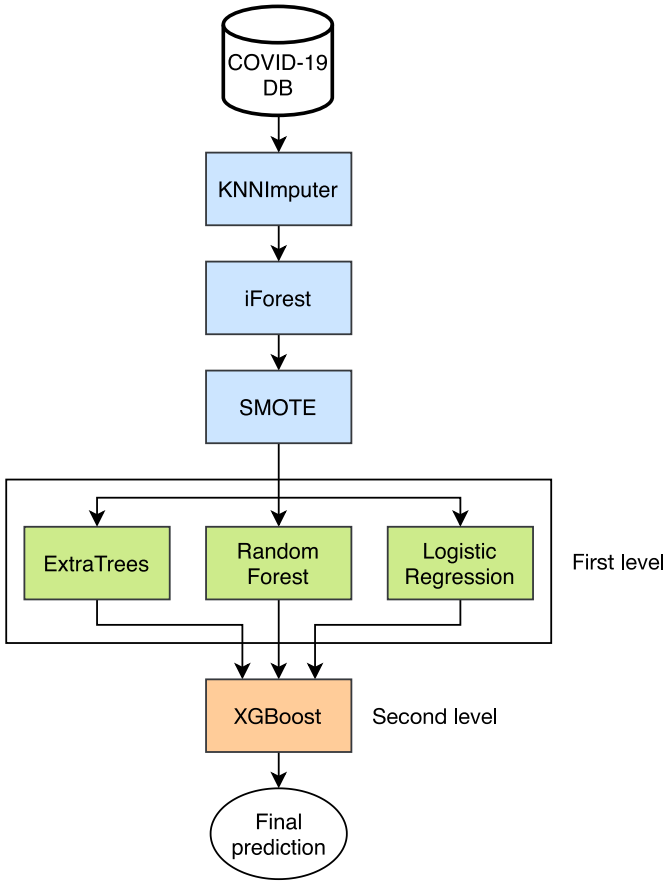

Fig. 1. The ERLX model.

For parameters setting, both extra trees and random forest classifiers have 300 estimators with the maximum depth set to 17. In logistic regression classifier, the max_iter parameter of logistic regression is set to 500 iterations. Moreover, the first level of the stacking model used ten fold cross-validation with data shuffling. The ERLX model sets number of estimators (trees) in XGBoost to 300 with a maximum depth of 17. To prevent overfitting, XGBoost uses a step size shrinkage named learning rate parameter, that ranges between 0 and 1. The learning rate in ERLX is set to one.

4. Performance evaluation and discussions

This section evaluates the proposed ERLX model and discusses results. The first subsection shows the impact of removing outliers and imputing COVID-19 dataset. The second subsection compares ERLX with other existing models.

4.1. Performance metrics

There are several performance metrics used to evaluate ML prediction models. To assess the accuracy of ERLX model the following performance metrics were employed: AUC, accuracy, sensitivity and specificity. Those performance metrics are computed based on confusion matrix values: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP and TN are the cases when the actual class is predicted correctly. On the other hand, FP and FN are the cases when the model predicted the actual class incorrectly. The receiver operating characteristic (ROC) curve (also known as AUC) depicts the relation between true positive (TP) and false positive (FP) rate, where the x-axis is FP and the y-axis is TP. The higher the AUC the better the model is in distinguishing between its two different classes. The study [9] demonstrated the need for AUC as a single figure to measure the classifier performance. Thus, AUC shows the trend of TP and FP.

Accuracy gives the proportion of correctly predicted cases (TP and TN) out of the whole dataset predictions. Accuracy exhibits a general view of model performance. The measurement of accuracy is given in equation (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100 \quad (1)$$

The proportion of true negative is called specificity. For instance, the proportion of the patients who are not infected with SARS-Cov2 and ERLX model predicted them as negative cases. Classification model with high specificity has higher value of TN and lower value of FP that boost model performance. Equation (2) gives the formula for computing specificity:

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

The proportion of true positive is called sensitivity. For example, the proportion of the actual COVID-19 patients that ERLX model correctly predicted them as COVID-19 patients. The higher the sensitivity is the better the model performs as higher sensitivity means higher TP and lower value of FN. Sensitivity is calculated as shown in equation (3):

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

4.2. Impact of eliminating outliers and balancing data

Table 2 lists accuracy, area under the curve (AUC), sensitivity, and specificity of ERLX model on the COVID-19 dataset. The first row shows the performance metric of the dataset with its outliers and imbalanced class distributions. While the last row represents the balanced dataset without outliers. Results of the remaining combination of data balancing and outliers are shown in the second and the third row. Results indicate that outliers impact AUC as the second row in Table 2 has the lowest AUC. Obviously, the balanced COVID-19 dataset without outliers outperforms imbalanced dataset with outliers in all performance metric, distinctly in AUC and sensitivity. Further, Fig. 2 illustrates the receiver operating characteristic (ROC) for ERLX model on the COVID-19 dataset after removing outliers and balancing the dataset. Moreover, the average confusion matrix obtained from 100 replications of ERLX model is illustrated in Fig. 3. On average the model was able to correctly predict 773.843 of COVID-19 negative cases and 65.4059 of COVID-19 positive cases. In addition, the false negatives average is 0.8614, while the false positives average is 0.1089, those low values corroborate the robustness of the ERLX model.

4.3. Comparison with other classification models

This subsection evaluates the proposed model ERLX by comparing it with other classification models from previous studies [3,22,56,71]. The comparison is in terms of model accuracy, the area under the curve (AUC) of the receiver operating characteristic (ROC), sensitivity, and specificity. Both ERLX and previous studies [3,22,56,71] used the same publicly available COVID-19 dataset provided by Kaggle [37]. All results from previous studies are taken from the reported results in their article. For comparison fairness, ERLX set the selected features same as each study that was compared with. However, the samples used may be different because of data balancing and other preprocessing steps. Experimental results for ERLX model are obtained from 100 repetitions. Each repetition has different data partitions where 80% of the dataset is allocated for the training data and 20% of the dataset is assigned for testing ERLX performance. The average of 100 repetitions and the 95% confidence interval are computed through bootstrapping the dataset randomly. The size of the bootstrap sample in each repetition is set to 80% of the whole dataset. From sklearn the resample () function is applied to select the bootstrap sample with replacement that means the records which are not in the bootstrap training sample are selected to the test sample data. Thereby, the 95% confidence intervals is computed in a robust way.

In this experiment, first the ERLX model is compared with the ER-CoV model [71]. The ER-CoV model is an ensemble-based classifier [62] built based on SMOTEBoost [14], SVM [8], and kNN algorithm [73] which handles C reactive protein missing values. There are similarities between ERLX and ER-CoV in the techniques used to build the two models. Both models balanced the imbalanced training data in

Table 2
Outliers removal and data balancing impacts on performance metric.

Dataset	Accuracy	AUC	Sensitivity	Specificity
Imbal w/ outliers	99.24% [95% CI: 98.7–99.7]	98.81% [95% CI: 97.1–100]	93.66% [95% CI: 88.7–98.7]	99.85% [95% CI: 99.5–100]
Bal w/ outliers	99.35% [95% CI: 98.7–99.8]	97.83% [95% CI: 94.1–99.9]	95.69% [95% CI: 90.2–100]	99.75% [95% CI: 99.3–100]
Imbal w/ o outliers	99.85% [95% CI: 99.5–100]	99.79% [95% CI: 98.8–100]	98.43% [95% CI: 95.0–100]	99.97% [95% CI: 99.8–100]
Bal w/o outliers	99.88% [95% CI: 99.6–100]	99.38% [95% CI: 97.5–100]	98.72% [95% CI: 94.6–100]	99.99% [95% CI: 99.99–100]

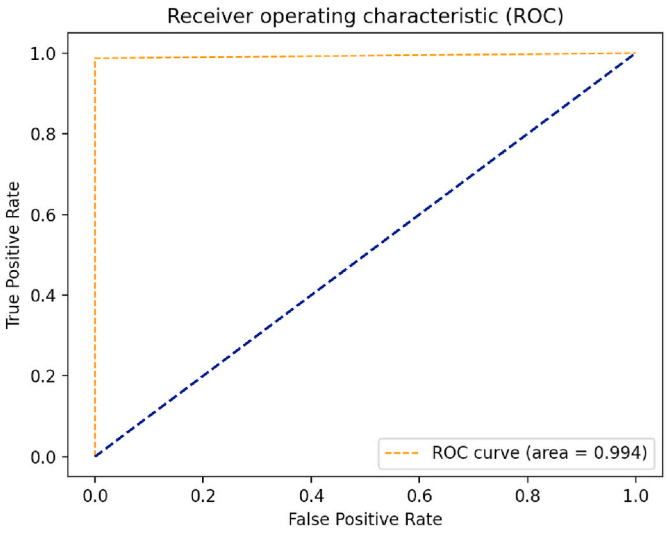

Fig. 2. The Receiver operating characteristic (ROC) curve for the test set.

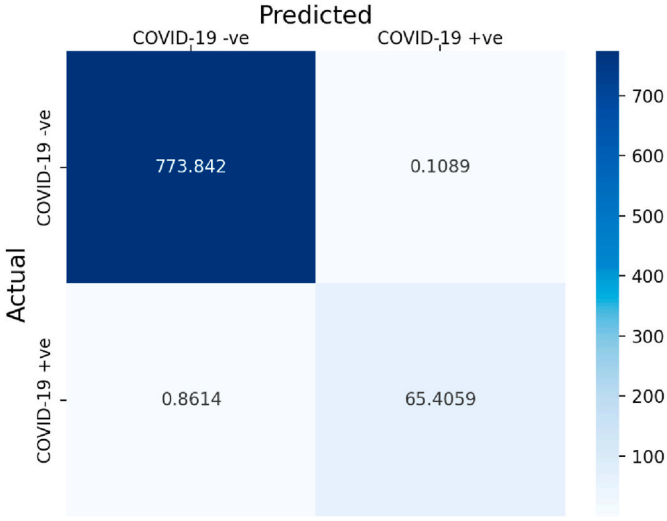

Fig. 3. Average confusion matrix obtained from 100 replications of ERLX.

Covid-19 dataset, however each model used a different algorithm: ERLX utilized SMOTE while ER-CoV used SMOTEBoost. The second similarity is handling missing values, ERLX utilized KNNImputer with seven neighbors while ER-CoV used kNN with five neighbors. Moreover, the ERLX model handled all missing values in its selected 18 features, while the ER-CoV handled just the missing values in C reactive protein feature. The third similarity is ensemble-based classifier, ER-CoV is based on ten SVM models, on the other hand, ERLX is a two levels ensemble-based classifier where the first level consists of three classifiers: extra trees, random forest, and logistic regression. Then, the second level is a XGBoost classifier. In the study [71], model accuracy is not reported. Thus, the comparison in this experiment is based on AUC, sensitivity,

Table 3
95% C.I. model performance of ERLX vs ER-CoV.

Model	AUC	Sensitivity	Specificity
ERLX with [71] features	99.73% [95%CI: 98.6–100]	99.47% [95%CI: 97.2–100]	99.99% [95%CI: 99.9–100]
ER-CoV [71]	86.78% [95% CI: 85.65–87.90]	70.25% [95% CI: 66.57–73.12]	85.98% [95% CI:84.94–86.84]

and specificity. Table 3 lists the 95% CI model performance for ERLX and ER-CoV. Obviously the proposed ERLX outperforms the ER-CoV in all metrics with an AUC, a sensitivity, and a specificity values of 99.73% [95% CI: 98.6–100], 99.47% [95% CI: 97.2–100], and 99.99% [95% CI: 99.9–100], respectively. Compared to the existing ER-CoV model which achieves an AUC, a sensitivity, and a specificity values of 86.78% [95% CI: 85.65–87.90], 70.25% [95% CI: 66.57–73.12], and 85.98% [95% CI: 84.94–86.84], respectively. Results confirm that ERLX is better than ER-CoV. Furthermore, the 99.73% [95% CI: 98.6–100] AUC of the proposed ERLX model is immensely higher than ER-CoV AUC 86.78% [95% CI: 85.65–87.90]. In addition, the low values of the false negatives and the false positives as shown in Fig. 4 imply that ERLX has significant ability to distinguish between positive and negative COVID-19 diagnosis. The results provide that the techniques used to build ERLX model lead to more accurate and reliable classification.

The second comparison in this experiment is shown in Table 4. In Ref. [3] ANN model is used to classify data and SMOTE algorithm is applied to balance training data similar to ERLX model. However, ERLX obtained considerable higher scores than ANN with SMOTE in all performance metrics. As expected, because ERLX is an ensemble-based classifier which combines multiple classification models to improve the prediction. In addition, ERLX uses iForest to eliminate outliers that also improves the classifier. The study in Ref. [22] tests different classification models, and results showed that the Bayes Net model achieved the best results. Even though, the proposed ERLX has better results than the Bayes Net in terms of accuracy, sensitivity and specificity as shown in Table 4. Further comparison, in the study [56] SVM recorded the best performance metrics in terms of AUC, sensitivity, and specificity. Nevertheless, Table 4 shows that ERLX significantly improves over SVM in AUC, sensitivity, and specificity. For additional COVID-19 diagnosing model blood tests based, the readers are referred to several accomplished works in this field [11,43,66,79]. Due to the unavailability of the COVID-19 dataset used in those studies, ERLX was not compared with them.

The results of ERLX demonstrate that the integration of KNNImputer, iForest, and SMOTE enhance the model performance compared to other models in the literature. Results reveal that imputing the missing values by applying KNNImputer helps in preventing loss of COVID-19 dataset records. Hence, the model trains with more qualified records which improve performance. In addition, the usage of iForest, and SMOTE raise the model robustness. Furthermore, the basic blocks of ERLX which is a two levels ensemble classification model yield to a robust prediction as it combines predictions from different classification models.

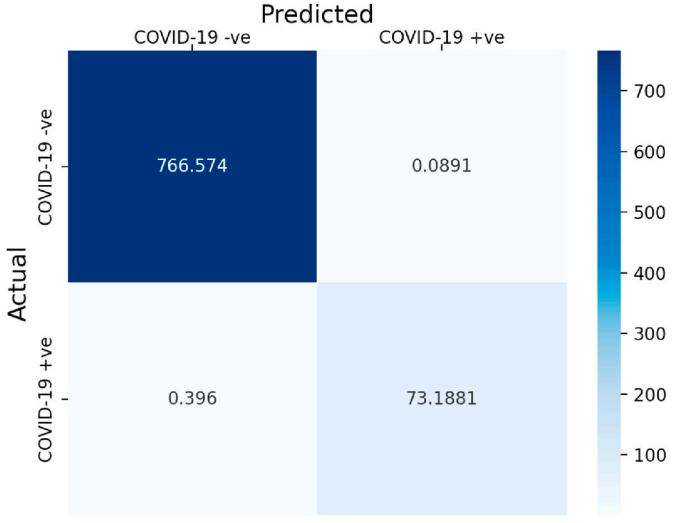

Fig. 4. Average confusion matrix obtained from 100 replications of ERLX with [71] features.

Table 4
Average performance metrics for each model.

Model	Accuracy	AUC	Sensitivity	Specificity
ERLX with [3] features	99.94% [95% CI: 99.8–100]	99.7% [95% CI: 98.7–100]	99.38% [95% CI: 97.3–100]	99.99% [95% CI: 99.9–100]
ANN with SMOTE [3]	87%	80% ± 0.09	43%	91%
ERLX with [22] features	99.94% [95% CI: 99.6–100]	99.69% [95% CI: 98.3–100]	99.93% [95% CI: 99.6–100]	99.96% [95% CI: 99.4–100]
Bayes Net [22]	95.159% ± 0.693	–	96.8% ± 0.007	93.6% ± 0.011
ERLX with [56] features	99.94% [95% CI: 99.6–100]	99.77% [95% CI: 98.3–100]	99.55% [95% CI: 96.5–100]	99.98% [95% CI: 99.9–100]
SVM [56]	–	85%	68%	85%

Consequently, results demonstrate the efficacy of ERLX in diagnosing COVID-19 patients from routine blood tests. The source codes of the proposed ERLX classification model are publicly available at https://github.com/Maryom/ERLX. Despite the great results of ERLX model, ERLX has some drawbacks, the main limitation is that all the results are based on single data source. Thus, the model suffers from a limited generalizability. Further, ERLX selected features manually without applying any of the well known features selection methods.

4.4. Feature importance

The clinical decisions made with the assistance of ML models in healthcare settings may affect the lives of patients in addition to other legal and ethical accountability. Therefore, in these applications, it is highly demanded to have prediction models that are both accurate and interpretable [55]. In the context of medical field, the model interpretability means that healthcare practitioners can understand how the model uses the input features to make predictions, be able to verify model outputs before acting on them, and defend care decisions based on the ML model [30]. Hence, causality-based feature importance estimates play a central role both for the interpretability and robustness of predictive models. In order to interpret the proposed ensemble model, we adopted the SHapley Additive exPlanations (SHAP) technique [51] to assess each feature importance in determining the predicted output.

SHAP interprets a model based on Shapley values, which explain the contribution of each feature to the prediction. Fig. 5 depicts a density scatter plot of SHAP values which integrates feature importance with feature effect regarding to positive cases of COVID-19. On the left side features are sorted according to their importance. The color on the right side represents feature value; the red color corresponds to a higher value whereas the blue color represents a lower value. Fig. 5 shows that monocytes contribute the most to the predictive model in determining COVID-19 positive cases. Common laboratory findings found in positive patients in clinical settings include low eosinophils, low platelets, high C-reactive protein and high aspartate transaminase [70], all of which show strong correlation values in our predictive model, as shown Fig. 5.

5. Conclusions

Early detection of COVID-19 patients is critical for timely intervention and prevention of the spread of the pandemic. Recent studies have revealed the use of routine blood tests for initial screening of COVID-19 patients supported by the fact that blood tests are relatively quick, less

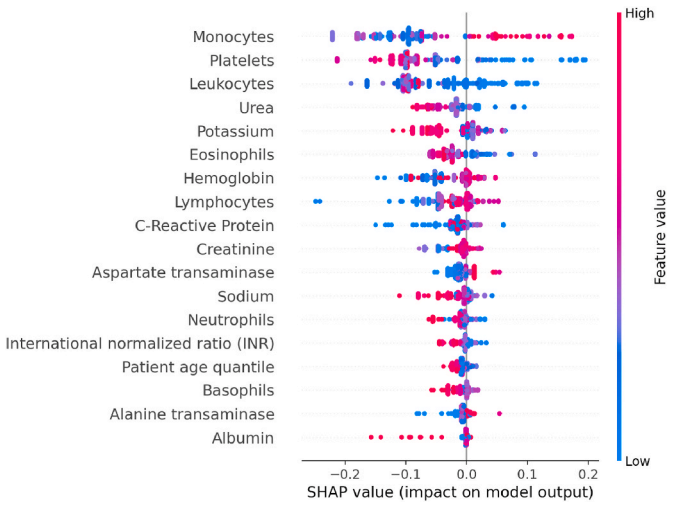

Fig. 5. Shap summary plot.

expensive, and readily available in all patient care locations. In this article, we first presented a review of the state-of-the-art techniques for COVID-19 detection using routine laboratory and clinical data to inspire the researchers in developing better prediction models to combat the disease. Then we developed an ensemble learning model called ERLX for diagnosing COVID-19 from routine blood tests. The proposed model utilized structural diversity by employing two level of classifiers, the prediction from the first level classifiers (extra trees, RF, LR) was fed to the second level classifier (XGBoost) to enhance the predictor capabilities. A number of data preparation steps were performed by using KNNImputer algorithm to handle null values in the dataset, isolation forest (iForest) to remove outlier data, and synthetic minority oversampling technique (SMOTE) to balance data distribution. The effectiveness and reliability of the ensemble learning model for diagnosing COVID-19 was demonstrated by comparing the results against existing state-of-the-art studies for a publicly available dataset from Albert Einstein Hospital in Brazil. The proposed prediction model performance gains were mainly from using an ensemble modeling approach that exploits the strength of a number of diverse classifiers and then combining their predictions via stacking. Ensemble models are very effective, robust, and extremely versatile in their performance since diversity is their key guiding principle to capture underlying structure of training data. To further interpret the proposed ensemble model, we adopted the SHAP technique to assess each feature importance in determining the predicted output.

Nevertheless, in order for machine learning models to make progress in automated and accurate COVID-19 diagnosis in clinical healthcare settings, some challenges need to be addressed including the availability of diverse high quality datasets, rigorous testing and external validation under the guidance from clinicians and health care providers [32], and construction of multi-modal machine learning models that can process and fuse information from many diverse sources of data [52] such as information from patient history, clinical signs and symptoms, physical examination, vital signs, X-rays medical imaging, epidemiological and clinical laboratory studies [20,44,48,72,80].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to thank the anonymous reviewers for their constructive comments and suggestions.

References

[1] Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Statistician 1992;46:175–85.
[2] Arnold DT, Attwood M, Barratt S, Elvers K, Morley A, McKernon J, Oates A, Donald C, Noel A, MacGowan A, et al. Blood parameters measured on admission as predictors of outcome for covid-19; a prospective UK cohort study. medRxiv URL, https://doi.org/10.1101/2020.06.25.20137935; 2020.
[3] Banerjee A, Ray S, Vorselaars B, Kitson J, Mamalakis M, Weeks S, Mackenzie LS. Use of machine learning and artificial intelligence to predict sars-cov-2 infection from full blood counts in a population. Int Immunopharm 2020. 106705.
[4] Bao FS, He Y, Liu J, Chen Y, Li Q, Zhang CR, Han L, Zhu B, Ge Y, Chen S, et al. Triaging moderate covid-19 and other viral pneumonias from routine blood tests. 2020. arXiv preprint arXiv:2005.06546.
[5] Bao J, Li C, Zhang K, Kang H, Chen W, Gu B. Comparative analysis of laboratory indexes of severe and non-severe patients infected with covid-19. Clin Chim Acta 2020. https://doi.org/10.1016/j.cca.2020.06.009. URL.
[6] Bayat V, Phelps S, Ryono R, Lee C, Parekh H, Mewton J, Sedghi F, Etminani P, Holodniy M. A covid-19 prediction model from standard laboratory tests and vital signs. Available at: SSRN 3594614 URL, https://doi.org/10.2139/ssrn.3594614; 2020.
[7] Born J, Brändle G, Cossio M, Disdier M, Goulet J, Roulin J, Wiedemann N. Pocovid-net: automatic detection of covid-19 from a new lung ultrasound imaging dataset (pocus). 2020. arXiv preprint arXiv:2004.12084.
[8] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory; 1992. p. 144–52.
[9] Bradley AP. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recogn 1997;30:1145–59.
[10] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[11] Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of covid-19 infection from routine blood exams with machine learning: a feasibility study. medRxiv URL, https://doi.org/10.1007/s10916-020-01597-4; 2020.
[12] Bullock J, Pham KH, Lam CSN, Luengo-Oroz M, et al. Mapping the landscape of artificial intelligence applications against covid-19. 2020. arXiv preprint arXiv:2003.11336.
[13] Burog A, Yacapin C, Maglente RRO, Macalalad-Josue AA, Uy EJB, Dans AL, Dans LF. Should igm/igg rapid test kit be used in the diagnosis of covid-19? Asia Pacific Center for Evidence Based Healthcare 2020;4:1–12.
[14] Chawla NV, Lazarevic A, Hall LO, Bowyer KW. Smoteboost: improving prediction of the minority class in boosting. In: European conference on principles of data mining and knowledge discovery. Springer; 2003. p. 107–19.
[15] Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016. p. 785–94.
[16] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, et al. Detection of 2019 novel coronavirus (2019-ncov) by real-time rt-pcr. Euro Surveill 2020;25:2000045.
[17] Döhla M, Boesecke C, Schulte B, Diegmann C, Sib E, Richter E, Eschbach-Bludau M, Aldabbagh S, Marx B, Eis-Hübinger AM, et al. Rapid point-of-care testing for sars-cov-2 in a community screening setting shows low sensitivity. Public health URL, https://doi.org/10.1016/j.puhe.2020.04.009; 2020.
[18] Dong D, Tang Z, Wang S, Hui H, Gong L, Lu Y, Xue Z, Liao H, Chen F, Yang F, et al. The role of imaging in the detection and management of covid-19: a review. In: IEEE reviews in biomedical engineering URL; 2020. https://doi.org/10.1109/RBME.2020.2990959.
[19] Fan BE, Chong VCL, Chan SSW, Lim GH, Lim KGE, Tan GB, Mucheli SS, Kuperan P, Ong KH. Hematologic parameters in patients with covid-19 infection. Am J Hematol 2020;95:E131–4.
[20] Feng C, Huang Z, Wang L, Chen X, Zhai Y, Zhu F, Chen H, Wang Y, Su X, Huang S, et al. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected covid-19 pneumonia in fever clinics. medRxiv URL, https://doi.org/10.1101/2020.03.19.20039099; 2020.
[21] Ferrari D, Motta A, Strollo M, Banfi G, Locatelli M. Routine blood tests as a potential diagnostic tool for covid-19. Clinical chemistry and laboratory medicine (CCLM). 2020. https://doi.org/10.1515/cclm-2020-0398. URL.
[22] de Freitas Barbosa VA, Gomes JC, de Santana MA, de Almeida Albuquerque JE, de Souza RG, de Souza RE, dos Santos WP. Heg. ia: an intelligent system to support diagnosis of covid-19 based on blood tests. medRxiv URL, https://doi.org/10.1101/2020.05.14.20102533; 2020.
[23] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001:1189–232.
[24] Gao Y, Li T, Han M, Li X, Wu D, Xu Y, Zhu Y, Liu Y, Wang X, Wang L. Diagnostic utility of clinical laboratory data determinations for patients with the severe covid-19. J Med Virol 2020. https://doi.org/10.1002/jmv.25770. URL.
[25] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63:3–42.
[26] Golino HF, Amaral LSdB, Duarte SFP, Gomes CMA, Soares TdJ, Reis LAd, Santos J. Predicting increased blood pressure using machine learning. Journal of Obesity 2014. https://doi.org/10.1155/2014/637635. 2014. URL.
[27] Gunčar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M, Notar M. An application of machine learning to haematological diagnosis. Sci Rep 2018;8:1–12.

[28] rekha Hanumanthu S. Role of intelligent computing in covid-19 prognosis: a state-of-the-art review. Chaos: Solitons & Fractals; 2020. 109947.
[29] Haykin S. Neural networks: principles and practice. Bookman 2001;11:900.
[30] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Min Knowl Discov 2019;9:e1312.
[31] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression, 398. John Wiley & Sons; 2013.
[32] Hu Y, Jacob J, Parker GJ, Hawkes DJ, Hurst JR, Stoyanov D. The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. Nature Machine Intelligence 2020:1–3.
[33] Ijaz MF, Alfian G, Syafrudin M, Rhee J. Hybrid prediction model for type 2 diabetes and hypertension using dbscan-based outlier detection, synthetic minority over sampling technique (smote), and random forest. Appl Sci 2018;8:1325.
[34] Imran A, Posokhova I, Qureshi HN, Masood U, Riaz S, Ali K, John CN, Nabeel M. Ai4covid-19: ai enabled preliminary diagnosis for covid-19 from cough samples via an app. 2020. https://doi.org/10.1016/j.imu.2020.100378. arXiv preprint arXiv: 2004.01275 URL.
[35] Ivanov I. Vecstack. 2016. https://github.com/vecxoz/vecstack.
[36] Joshi RP, Pejaver V, Hammarlund NE, Sung H, Lee SK, Furmanchuk A, Lee HY, Scott G, Gombar S, Shah N, et al. A predictive tool for identification of sars-cov-2 pcr-negative emergency department patients using routine test results. J Clin Virol 2020. 104502.
[37] Kaggle. Diagnosis of covid-19 and its clinical spectrum | kaggle. 2020. https://www.kaggle.com/einsteindata4u/covid19. Accessed on 07/18/2020.
[38] Kam HT. Random decision forest. In: Proceedings of the 3rd international conference on document analysis and recognition. Montreal: Canada; 1995, 278282. August.
[39] Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks. IEEE; 1995. p. 1942–8.
[40] Kermali M, Khalsa RK, Pillai K, Ismail Z, Harky A. The role of biomarkers in diagnosis of covid-19–a systematic review. Life Sciences; 2020. 117788.
[41] Khartabil T, Russcher H, van der Ven A, de Rijke Y. A summary of the diagnostic and prognostic value of hemocytometry markers in covid-19 patients. Crit Rev Clin Lab Sci 2020;1–17.
[42] Kohonen T. Essentials of the self-organizing map. Neural Network 2013;37:52–65.
[43] Kukar M, Gunčar G, Vovko T, Podnar S, Černelč P, Brvar M, Zalaznik M, Notar M, Moškon S, Notar M. Covid-19 diagnosis by routine blood tests using machine learning. 2020. arXiv preprint arXiv:2006.03476.
[44] Langer T, Favarato M, Giudici R, Bassi G, Garberi R, Villa F, Gay H, Zeduri A, Bragagnolo S, Molteni A, et al. Use of machine learning to rapidly predict positivity to severe acute respiratory syndrome coronavirus 2 (sars-cov-2) using basic clinical data URL. 2020. https://doi.org/10.21203/rs.3.rs-38576/v1.
[45] Latif S, Usman M, Manzoor S, Iqbal W, Qadir J, Tyson G, Castro I, Razi A, Boulos MNK, Weller A, et al. Leveraging data science to combat covid-19: a comprehensive review. IEEE Transactions on Artificial Intelligence; 2020.
[46] Lewis DD. Naive (bayes) at forty: the independence assumption in information retrieval. In: European conference on machine learning. Springer; 1998. p. 4–15.
[47] Li D, Wang D, Dong J, Wang N, Huang H, Xu H, Xia C. False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: role of deep-learning-based ct diagnosis and insights from two cases. Korean J Radiol 2020;21:505–8.
[48] Li WT, Ma J, Shende N, Castaneda G, Chakladar J, Tsai JC, Apostol L, Honda CO, Xu J, Wong LM, et al. Using machine learning of clinical data to diagnose covid-19. 2020. https://doi.org/10.1101/2020.06.24.20138859. medRxiv URL.
[49] Liang KH, Yao X, Newton C. Evolutionary search of approximated n-dimensional landscapes. Int J Knowl Base Intell Eng Syst 2000;4:172–83.
[50] Liu FT, Ting KM, Zhou ZH. Isolation forest. In: 2008 eighth IEEE international conference on data mining. IEEE; 2008. p. 413–22.
[51] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in neural information processing systems; 2017. p. 4765–74.
[52] Mei X, Lee HC, Diao Ky, Huang M, Lin B, Liu C, Xie Z, Ma Y, Robson PM, Chung M, et al. Artificial intelligence–enabled rapid diagnosis of patients with covid-19. Nat Med 2020:1–5.
[53] Meier L, Van De Geer S, Bühlmann P. The group lasso for logistic regression. J Roy Stat Soc B 2008;70:53–71.
[54] Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. Kaohsiung J Med Sci 2013;29:93–9.
[55] Molnar C. Interpretable machine learning. 2019. https://christophm.github.io/interpretable-ml-book/.
[56] de Moraes Batista AF, Miraglia JL, Donato THR, Chiavegatto Filho ADP. Covid-19 diagnosis prediction in emergency care patients: a machine learning approach. medRxiv URL, https://doi.org/10.1101/2020.04.04.20052092; 2020.
[57] Nan SN, Ya Y, Ling TL, Nv GH, Ying PH, Bin J, et al. A prediction model based on machine learning for diagnosing the early covid-19 patients. 2020. https://doi.org/10.1101/2020.06.03.20120881. medRxiv URL.
[58] Nguyen TT. Artificial intelligence in the battle against coronavirus (covid-19): a survey and future research directions. Preprint 2020;10. https://doi.org/10.13140/RG.2.2.36491.23846. URL.
[59] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[60] Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinf 2017;18:1–14.
[61] Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Alvarado-Arnez LE, Bonilla-Aldana DK, Franco-Paredes C, Henao-Martinez AF, et al. Clinical, laboratory and imaging features of covid-19: a systematic review and meta-analysis. Travel medicine and infectious disease; 2020, 101623.
[62] Rokach L. Ensemble-based classifiers. Artif Intell Rev 2010;33:1–39.
[63] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics 1991;21:660–74.
[64] Schölkopf B, Smola AJ, Bach F, et al. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press; 2002.
[65] Schuller BW, Schuller DM, Qian K, Liu J, Zheng H, Li X. Covid-19 and computer audition: an overview on what speech & sound analysis could contribute in the sars-cov-2 corona crisis. 2020. arXiv preprint arXiv:2003.11117.
[66] Schwab P, Schütte AD, Dietz B, Bauer S. predcovid-19: a systematic study of clinical predictive models for coronavirus disease 2019. 2020. arXiv preprint arXiv: 2005.08302.
[67] Scikit-learn. 2020. sklearn.ensemble.extratreesclassifier, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html. Online. accessed 19 July 2020.
[68] Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, Shen D. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. In: IEEE reviews in biomedical engineering URL; 2020. https://doi.org/10.1109/RBME.2020.2987975.
[69] Siatka C, Eveillard M, Nishimura J, Duroux C, Ferrandi G. Covid-19 screening, prognosis and severity assessment with biomarkers for management of patients. 2020. URL, https://hal.archives-ouvertes.fr/hal-02547315.
[70] Siordia Jr JA. Epidemiology and clinical features of covid-19: a review of current literature. J Clin Virol 2020. 104357.
[71] Soares F, Villavicencio A, Fogliatto FS, Rigatto MHP, Anzanello MJ, Idiart M, Stevenson M. A novel specific artificial intelligence-based method to identify covid-19 cases using simple blood exams. medRxiv URL, https://doi.org/10.1101/2020.04.10.20061036; 2020.
[72] Soltan AA, Kouchaki S, Zhu T, Kiyasseh D, Taylor T, Hussain ZB, Peto T, Brent AJ, Eyre DW, Clifton D. Artificial intelligence driven assessment of routinely collected healthcare data is an effective screening test for covid-19 in patients presenting to hospital. 2020. https://doi.org/10.1101/2020.07.07.20148361. medRxiv URL.
[73] Torgo L. Data mining with R: learning with case studies. CRC press; 2016.
[74] Ulhaq A, Khan A, Gomes D, Pau M. Computer vision for covid-19 control: a survey. 2020. arXiv preprint arXiv:2004.09420.
[75] WHO. Coronavirus disease (covid-19). 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed on 07/10/2020.
[76] Wolpert DH. Stacked generalization. Neural Network 1992;5:241–59.
[77] Wu G, Zhou S, Wang Y, Li X. Machine learning: a predication model of outcome of sars-cov-2 pneumonia. 2020. https://doi.org/10.21203/rs.3.rs-23196/v1. URL.
[78] Wu H, Yang S, Huang Z, He J, Wang X. Type 2 diabetes mellitus prediction model based on data mining. Informatics in Medicine Unlocked 2018;10:100–7.
[79] Wu J, Zhang P, Zhang L, Meng W, Li J, Tong C, Li Y, Cai J, Yang Z, Zhu J, et al. Rapid and accurate identification of covid-19 infection through machine learning based on clinical available blood test results. medRxiv URL, https://doi.org/10.1101/2020.04.02.20051136; 2020.
[80] Yan L, Zhang HT, Xiao Y, Wang M, Sun C, Liang J, Li S, Zhang M, Guo Y, Xiao Y, et al. Prediction of criticality in patients with severe covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in wuhan. MedRxiv URL, https://doi.org/10.1101/2020.02.27.20028027; 2020.
[81] Yang HS, Vasovic LV, Steel P, Chadburn A, Hou Y, Racine-Brzostek SE, Cushing M, Loda M, Kaushal R, Zhao Z, et al. Routine laboratory blood tests predict sars-cov-2 infection using machine learning. medRxiv URL, https://doi.org/10.1101/2020.06.17.20133892; 2020.