Karya J Health Sci. 2022; 3(3): 216-221
ARAŞTIRMA MAKALESİ/RESEARCH ARTICLE doi: 10.52831/kjhs.1117894 e-ISSN:2717-9540

KARYA JOURNAL OF HEALTH SCIENCE

journal homepage: www.dergipark.org.tr/kjhs

COVID19PREDICTOR: WEB-BASED INTERFACE TO DEVELOP MACHINE LEARNING MODELS FOR DIAGNOSIS OF COVID-19 BASED ON CLINICAL DATA AND ROUTINE TESTS

COVID19PREDICTOR: KLİNİK VERİLERE VE RUTİN TESTLERE DAYALI OLARAK COVID-19 TEŞHİSİ İÇİN MAKİNE ÖĞRENİMİ MODELLERİ GELİŞTİRMEYE YARAYAN WEB TABANLI ARAYÜZ

Volkan Kapucu[1], Sultan Turhan[1], Metin Pıçakçıefe[2], Eralp Doğu[1*]

[1]Department of Statistics, Faculty of Science, Muğla Sıtkı Koçman University, Muğla, Turkey

[2]Department of Public Health, Faculty of Medicine, Muğla Sıtkı Koçman University, Muğla, Turkey

ABSTRACT

Objective: The Covid-19 outbreak has become the primary health problem of many countries due to health related, social, economic and individual effects. In addition to the development of outbreak prediction models, the examination of risk factors of the disease and the development of models for diagnosis are of high importance. This study introduces the Covid19PredictoR interface, a workflow where machine learning approaches are used for diagnosing Covid-19 based on clinical data such as routine laboratory test results, risk factors, information on co-existing health conditions.

Method: Covid19PredictoR interface is an open source web based interface on R/Shiny (https://biodatalab.shinyapps.io/Covid19PredictoR/). Logistic regression, C5.0, decision tree, random forest and XGBoost models can be developed within the framework. These models can also be used for predictive purposes. Descriptive statistics, data pre-processing and model tuning steps are additionally provided during model development.

Results: Einsteindata4u dataset was analyzed with the Covid19PredictoR interface. With this example, the complete operation of the interface and the demonstration of all steps of the workflow have been shown. High performance machine learning models were developed for the dataset and the best models were used for prediction. Analysis and visualization of features (age, admission data and laboratory tests) were carried out for the case per model.

Conclusion: The use of machine learning algorithms to evaluate Covid-19 disease in terms of related risk factors is rapidly increasing. The application of these algorithms on various platforms creates application difficulties, repeatability and reproducibility problems. The proposed pipeline, which has been transformed into a standard workflow with the interface, offers a user-friendly structure that healthcare professionals with various background can easily use and report.

Key Words: Covid-19, Machine Learning, Shiny, Interface

ÖZ

Amaç: Covid-19 salgını sağlıkla ilgili, sosyal, ekonomik ve bireysel etkiler nedeniyle birçok ülkenin birincil sağlık sorunu haline gelmiştir. Salgın tahmin modellerinin geliştirilmesinin yanı sıra hastalığın risk faktörlerinin incelenmesi ve teşhise yönelik modellerin geliştirilmesi büyük önem taşımaktadır. Bu çalışma, rutin laboratuvar test sonuçları, risk faktörleri, birlikte var olan sağlık koşullarına ilişkin bilgiler gibi klinik verilere dayalı olarak Covid-19'u teşhis etmek için makine öğrenimi yaklaşımlarının kullanıldığı bir iş akışı olan Covid19PredictoR arayüzünü tanıtmaktadır.

Yöntem: Covid19PredictoR arayüzü, R/Shiny'de (https://biodatalab.shinyapps.io/Covid19PredictoR/) açık kaynaklı web tabanlı bir arayüzdür. Sistem içerisinde lojistik regresyon, C5.0, karar ağacı, rastgele orman ve XGBoost modelleri geliştirilebilir. Bu modeller aynı zamanda tahmin amacıyla da kullanılabilir. Model geliştirme sırasında ek olarak tanımlayıcı istatistikler, veri ön işleme ve model ayarlama adımları sağlanır.

Bulgular: Einsteindata4u veri seti, Covid19PredictoR arayüzü ile analiz edildi. Bu örnekle, arayüzün eksiksiz çalışması ve iş akışının tüm adımlarının gösterimi aktarıldı. Veri seti için yüksek performanslı makine öğrenme modelleri geliştirilmiş ve tahmin için en iyi modeller kullanıldı. Model başına vaka için özelliklerin analizi ve görselleştirilmesi (yaş, kabul verileri ve laboratuvar testleri) yapıldı.

Sonuç: Covid-19 hastalığını, ilgili risk faktörleri açısından değerlendirmek için makine öğrenimi algoritmalarının kullanımı, hızla artmaktadır. Bu algoritmaların çeşitli platformlarda uygulanması, uygulama zorlukları, tekrarlanabilirlik ve tekrar üretilebilirlik sorunları yaratmaktadır. Arayüz ile standart bir iş akışına dönüştürülen, tasarlanmış bu işlem zinciri, çeşitli geçmiş deneyimlere sahip sağlık uzmanlarının rahatlıkla kullanabileceği ve raporlayabileceği kullanıcı dostu bir yapı sunar.

Anahtar Kelimeler: Covid-19, Makine Öğrenmesi, Shiny, Arayüz

Makale Bilgisi/Article Info

Yükleme tarihi/Submitted: 18.05.2022, **Revizyon isteği/Revision requested:** 13.09.2022, **Son düzenleme tarihi/Last revision received:** 10.10.2022, **Kabul/Accepted:** 17.10.2022

***Sorumlu yazar/Corresponding author:** Muğla Sıtkı Koçman University, Faculty of Science, Department of Statistics, Mugla, Turkey
[4*]Email: eralp.dogu@mu.edu.tr,[1]Email: volkan@mu.edu.tr, [2]Email: sultantrhn@hotmail.com, [3]Email: metinpicakciefe@mu.edu.tr

INTRODUCTION

The epidemic, which started with the increase in suspected pneumonia cases in Wuhan, China on November 2019, was defined as a pandemic by the World Health Organization (WHO) on March 11, 2020 [1]. Coronaviruses (CoV) are a large family of viruses that cause illness ranging from the common cold to more severe diseases. A novel coronavirus (nCoV) is a new strain that has not been previously identified in humans. The new virus was subsequently named the "Covid-19 virus" [2]. The pandemic has caused millions of people worldwide to be infected and die. The Covid-19 pandemic has become the primary health problem of many countries due to its tremendous impact on social life and economy.

In addition to the development of epidemic prediction models, the examination of laboratory tests results, underlying risk factors and analyzing habits are of high importance for preventive measures. As each variant of the disease has the potential to have higher mortality and other effects, identifying patients as early as possible has become a major concern. The current practice to diagnose Covid-19 is to use Reverse Transcription-Polymerase Chain Reaction (RT-PCR) where, specimens collected from respiratory system of a patient is processed through the Ribonucleic Acid (RNA) technology [3]. Although RT-PCR is known as the gold standard for Covid-19 diagnosis, the need for special equipment, trained personnel and time limits its use. Moreover, false negative tests have been documented when specimen is processed in batches. Alternatively, Computed Tomography (CT) scans have been widely suggested as a complementary diagnosis tool to Polymerase Chain Reaction (PCR). There are some setbacks in using CTs as well, i.e. exposing to unnecessary radiation and occupying the limited capacity of the health system. As a consequence of these limitations, recently CT scans are not recommended for the initial diagnosis of Covid-19 [4].

It is critical that we as a global community continue to commit to the development of testing strategies to assist in Covid-19 public health efforts in all areas of the world if we are going to implement successful pandemic mitigation efforts [5].

Vaccination and preventive actions taken by the society help many patients develop mild symptoms, however, the risk of spreading the disease still exist. Thus, health professionals need complementary tools to early diagnose the disease and validate existing diagnosis. Machine learning models can be used to achieve this task. High performing models based on clinical data and routine tests can be used as an alternative, fast and cost effective diagnosis approach.

Various papers discussed machine learning to diagnose Covid-19 on different data sources, methods, performance criteria and reporting [6-21]. Systematic reviews emphasized the increasing use of machine learning for Covid-19 diagnosis [22-24]. Most of these studies focus on state of the art machine learning models, development and comparison of the models, feature extraction, analysing risk factors and data visualization of Covid-19 diagnosis. However, as many of these approaches does not provide ready to use, simple and open source software, health providers face difficulty in terms of developing models with their own data and deploying these methods in their daily decisions. Furthermore, many machine learning methods use advanced software pipelines that needs advanced skills of coding mostly lacking in the health care providers skillset.

The open source Shiny [25] application, Covid19PredictoR, proposed in this paper is an interface for reproducible modelling that includes a pipeline of pre-processing, training, tuning, testing, and also has reporting features. It is a platform that has a workflow where machine learning approaches can be used to predict Covid-19 test results based on clinical data. Researchers can report based on machine learning algorithms such as logistic regression [9-11,17,18,26], decision tree, C5.0 [10,15,27], random forest [9,10,12,14,15,19,26,28], and XGBoost [6,12,13,28,29], which are the most commonly reported for Covid-19 diagnosis. These models can be used to uncover the importance of laboratory test results, risk factors and co-existing diseases for diagnosis of Covid-19. Providing a reproducible, user-friendly and free platform for all researchers of various levels of knowledge will increase the widespread impact of the study. The interface is available on https://biodatalab.shinyapps.io/Covid19PredictoR.

METHOD

Einsteindata4u Dataset

The illustrative dataset used in this study was obtained from patients who underwent RT-PCR testing and additional laboratory tests were also conducted during their visit to the Israelite Albert Einstein Hospital in Sao Paulo, Brazil. There are 5644 cases and 111 variables in the dataset [30]. We used a randomly balanced subset of the original cohort and removed totally missing and zero variance features. Thus, the dataset was finalized with 25 features and 86 cases (43 patients with Covid positive and 43 patients for Covid negative PCR results). Table 1 shows the summary statistics for the dataset. Categorical features are summarized by the count and percentage and numerical features are summarized by the mean and standard deviation. In addition, after the Shapiro-Wilk normality test was applied to the data, t-test, Wilcoxon or chi-square tests were applied depending on the situation, and the obtained statistics and p-values were added to the table.

Proposed ML Framework

The steps of the ML workflow suggested for this study are given in Figure 1.

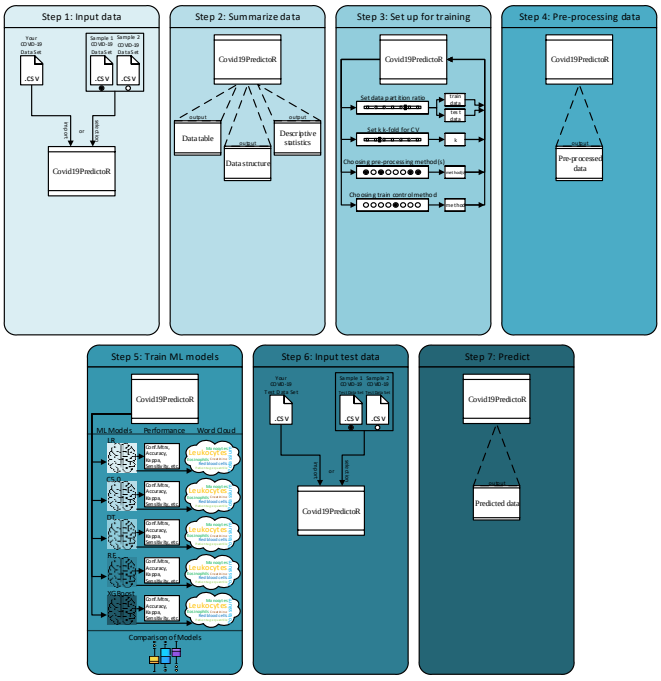

Figure 1. The ML workflow used in Covid19PredictoR.

Step 1. Input data: Here, the user uploads the dataset of patient specific data such as information on co-existing diseases, routine test results, gender etc. The first column of this csv file shows whether the patient has Covid (Code=1) or not (Code=0). We also provide two example datasets here for illustrative purposes.

Step 2. Summarize data: In the second step, data table, data structure, summary statistics and simple visual summaries are presented per feature.

Step 3. Set up for training: At this stage, global parameters that will be used during model development are selected. The training and validation ratio, the value for cross-validation, the data pre-processing methods (options include centering, scaling, zero variance feature

removal, Box Cox, Yeo Johnson and exponential transformations, K-Nearest Neighbors (knn), bagging and median imputation, and principal component and independent component analysis for dimension reduction) and the training control methods (options include bootstrap, cross validation, repeated cross validation, leave one out, out of bag, and adaptive versions of these methods) are chosen at this stage.

Step 4. Pre-Process data: Here, the data table of the pre-processed data is shown.

Step 5. Train ML models: The results of the developed models are displayed at this stage. In order to evaluate the performance results confusion matrix, performance summaries and the feature importance, the word cloud of the most important features, are shown in the results. Models can also be compared based on performance results such as Accuracy=((TP+TN)/((TP+TN+FP+FN)),Precision=((TP)/((TP+FP)),Recall=((TP)/((TP+FN)),F1=2(PrecisionxRecall)/(Precision+Recall),Kappa=(Relative observed agreement among raters-Hypothetical probability of chance agreement)/(1-Hypothetical probability of chance agreement), the user can determine the best model based on these results where TP is true positive, TN is true negative, FP is false positive and FN is false negative. There is also a separate tab where user can visually compare the results from all the models.

Step 6. Input test data: User uploads data for new patients where the class labels are unknown.

Step 7. Predict: In the last step, the test results are predicted based on a selected model.

Implementation

Implementation of above workflow was made in R, an open source software. The web based interface was created on Shiny, an open-source R package that allows creating a web interface using the R programming language. dplyr, stats, devtools, ggplot2, caret, e1071, plotly, DT, caTools, shinycssloaders, shinyWidgets, wordcloud, RColorBrewer, formattable, lime, shinydashboard and shiny packages were used during the study. Source code is available through https://github.com/infobiodatalab/Covid19PredictoR.

RESULTS

Input and Setup for Einsteindata4u Dataset

First, dataset containing routine blood tests and PCR results is uploaded on “Input data” tab. The settings were chosen through “set up for training” tab in Figure 2, the parameters required for training the model were set as follows: the data set was randomly separated as 75% training and 25% testing, the value was determined as 10 fold for Cross-Validation, from the methods required for pre-processing, center, scale and non-zero variance were selected and finally, cross validation method was chosen from the training control methods to be applied in the models.

In this first version of the interface, missing and outlier observations, class imbalance are ignored, since the model development stages are more important. Therefore, missing and outlier observations were removed from the original Einsteindata4u Dataset, which had 5644 observations and 111 variables. In addition, some observations were randomly removed to avoid class imbalance. Thus, the data set was finalized with a total of 86 patients, including 25 features, 43 Covid positive patients and 43 Covid negative PCR results.

Train and Compare Models for Einsteindata4u Dataset

The results of model development are displayed in the “Train ML models” step. Figure 3 shows the performance results obtained from random forest model along with the word cloud of the most important features. Random forest model provided a high accuracy in both training and testing phase.

Testing performance shows that the model is high performing in terms of various performance indicators; accuracy=0.90 (95% CI=[0.68 – 0.99]), kappa=0.8, precision=1.00, recall=0.83 and F1=0.91. Based on random forest model, feature importance matrix was converted into a word cloud where leukocytes, eosinophils and platelets are top predictors for Covid-19 diagnosis.

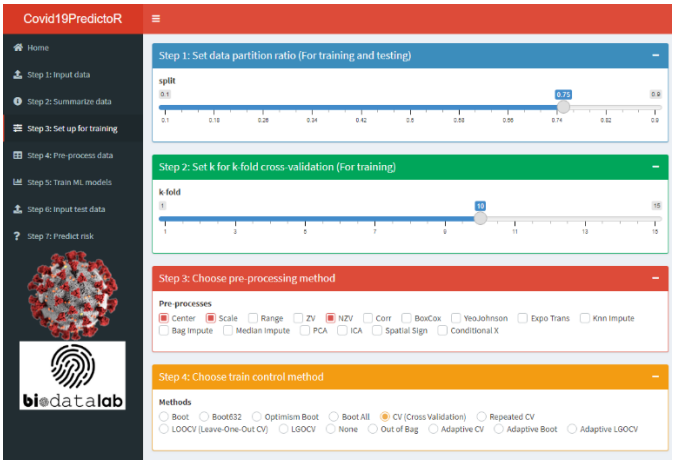

Figure 2. Input data and setting up training parameters

Similarly, the results (accuracy=0.85, 95% CI = [0.62 – 0.97], kappa=0.7, precision=0.90, recall=0.82 and F1 = 0.86) obtained from the XGBoost model are also promising. In terms of performance, the C5.0 model (accuracy=0.80 (95% CI=[0.56–0.94]), kappa=0.6, precision=0.90, recall=0.75 and F1=0.82) and the decision tree model (accuracy=0.75 (95% CI=[0.51–0.91]), kappa=0.5, precision=1.00, recall=0.67 and F1=0.80) is in the fourth place. The weakest performance was observed in logistic regression model (accuracy=0.65, kappa=0.3, precision=0.80, recall=0.62 and F1=0.70).

In the word cloud from the random forest model, it is clear that leukocytes and eosinophils are the best predictors for the diagnosis of Covid-19 (Figure 3).

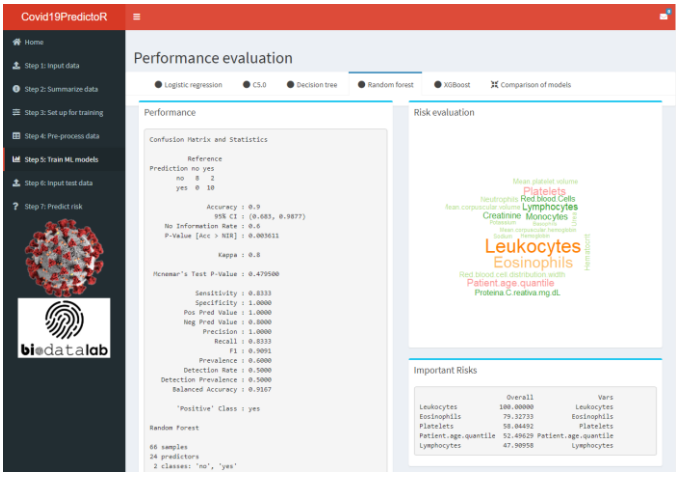

Figure 3. Performance of a trained model and word cloud of feature importance based on the test set. Case of random forest model.

A logistic regression model was created using the important features common to the random forest and XGBoost models obtained through the interface. The p values obtained for leukocytes (p=0.0009), eosinophils (p=0.0140), patient age quantile (0.0273), creatinine (0.0374), proteina C reativa (0.0266) and mean platelet volume (0.0491) as a result of logistic regression model are less than 0.05, which is considered significant. Also, as seen in Table 1, strong evidence of change in leukocytes, eosinophils, and patient age quantity is also seen in group comparisons.

Table 1. Summary statistics and group comparisons for Einsteindata4u dataset

Features		PCR negative	PCR positive	Test statistic	p-value
SARS-Cov-2 exam result		43 (50%)	43 (50%)		
Patient age quantile		11.77±5.33	14.42±3.61	W=638	**p=0.013**
Patient admitted to regular ward	No	40 (93%)	25 (58%)	X^2=12.349	**p<0.001**
	Yes	3 (7%)	18 (42%)		
Patient admitted to semi-intensive unit	No	39 (91%)	37 (86%)	X^2=0.11316	p=0.741
	Yes	4 (9%)	6 (14%)		
Patient admitted to intensive care unit	No	43 (100%)	38 (88%)	X^2=5.3086	p=0.053
	Yes	-	5 (12%)		
Hematocrit		0.16±0.94	0.24±0.84	W =900	p=0.836
Hemoglobin		0.14±0.92	0.25±0.84	t =-0.58915	p=0.556
Platelets		-0.18±0.75	-0.72±0.60	t=3.66	**p<0.001**
Mean platelet volume		0.09±1.07	0.16±0.76	W=832.5	p=0.429
Red blood Cells		-0.00±0.90	0.25±0.96	t=-1.2325	p=0.221
Lymphocytes		-0.24±0.80	0.14±0.86	t=-2.1329	**p=0.036**
Mean corpuscular hemoglobin concentration		-0.03±0.89	0.12±0.86	W=792	p=0.254
Leukocytes		-0.03±0.76	-0.83±0.45	W=1509	**p<0.001**
Basophils		-0.01±0.83	-0.15±0.64	W=972	p=0.680
Mean corpuscular hemoglobin		0.24±0.81	-0.01±0.68	t=1.5473	p=0.126
Eosinophils		0.08±0.92	-0.54±0.38	W=1376	**p<0.001**
Mean corpuscular volume		0.29±0.96	-0.07±0.71	W=1102.5	p=0.125
Monocytes		0.29±0.90	0.61±1.00	W=695	**p=0.048**
Red blood cell distribution width		-0.09±0.65	-0.16±0.75	W=998.5	p=0.525
Neutrophils		0.20±0.95	-0.15±0.89	t=1.792	p=0.076
Urea		-0.16±0.66	-0.18±0.57	t=0.1566	p=0.876
Proteina C reativa		-0.04±0.70	0.16±0.87	W=808.5	p=0.318
Creatinine		0.15±0.90	0.09±0.72	t=0.3051	p=0.761
Potassium		-0.05±1.00	-0.31±0.77	t=1.341	p=0.184
Sodium		-0.01±0.95	-0.22±0.98	t=1.0426	p=0.300

Prediction for New Patients and Interpretation

Using the best model, Covid19PredictoR can estimate the risk of Covid-19 for new patients using only the clinical observations. In Figure 4, prediction results for new patients number 1 to 3 are shown based on trained and tested random forest model. The results show that patient 2 and 3 have high risk of Covid-19 whereas patient 1 does not. Note that these particular patients were drawn from the Einsteindata4u dataset but were not used in model development or testing. The known labels of each patient are perfectly matched with the predictions. Plus, Covid19Predictor provides knowledge to interpret the predictions using model agnostics of the lime package. The patient 1 was classified as Covid negative because of the count of leukocytes and lymphocytes support the decision where count of eosinophils and creatinine and level of proteina C reativa does not support the decision to some extent. We see that the decisions were made for the patient 2 as Covid positive based on the count of leukocytes, eosinophils and lymphocytes. The major supporting decision rules applied to the patient 3 are the count of leukocytes, monocytes and platelets.

Most of these feature were found significant in our preliminary and later analyses and reported in literature.

Figure 4. Prediction step of Covid19PredictoR for random forest model

DISCUSSION

Our results show that a standard and reproducible software is helpful to analyse clinical parameters of Covid-19 to uncover the patterns based on clinical characteristics. Our framework is open source and easy to use which enables users to generate models based on their own data. The interface can be used as a stand-alone software. The source code also can be used to create pipelines on local servers.

Although other machine learning methods exist in literature such as Batista et al., Yang et al., Joshi et al., Kukar et al., Brinati et al., Tordjman et al., Cabitza et al., Gladding et al. most of these studies work with a certain dataset and report finding of this particular data [7,10-12,15,18-20]. Thus, user who wants to develop and test their data still needs to create a separate software pipeline where the standard workflow mentioned in these works might not be created easily resulting a reproducibility problem. Moreover, the available interfaces such as Kukar et al. and Brinati et al. relies on a certain training dataset which creates a model for a certain cohort represented in this data [12,15]. Covid-19 characteristics can change from population to population and to there is no global model that can be used for all patients. Our interface allows user to create independent models easily and user can use the interface for prediction purposes as well. This way, even small cohorts can be analysed and modelled.

The current literature mostly focuses on five machine learning algorithms namely, logistic regression, decision tree, C5.0, random forest and XGBoost. These methods are also available in Covid19PredictoR. Many machine learning algorithms have been tested in the creation of this interface. As a result of the trials, since it was seen that they were more successful in data sets with missing data, these algorithms were considered to be used in the interface. In addition, the selection of candidate algorithms with different analytical background is adopted in the selection of algorithms. For example; logistic regression was chosen to represent the statistical learning method, decision tree to represent the rule-based learning method, random forest to represent the bagging method, and XGBoost to represent the boosting method.

Unfortunately, the interpretation of these models is not well addressed in the literature. Alves et al. discussed the lack of interpretability and proposed tree and criteria graph-based model interpretation [31]. A software implementation for further use was not provided. Covid19PredictoR provides similar interpretation based on local model agnostics in the prediction phase. We believe this functionality is vital in terms on understanding the black box structure of the above algorithms.

Study Limitations

As a limitation of this work, the datasets are assumed to be balanced in the design of first version of the interface. Further research is needed for handling imbalanced datasets and additional pre-process steps for balancing needs to be integrated to the web interface.

Handling missing data is another limitation of this work. Although the pipeline does not include a missing data imputation method, the learning methods considered in this study can handle missing values. Further research needs to be done on integrating easy to apply data imputation methods. Moreover, handling noisy data might be another direction of research. This study doesn't include any methods to handle background noise. The next version of the interface needs to include summary statistics for such data processing issues.

The web interface now only covers five most commonly used machine learning algorithms including one bagging and one boosting approach and further research is needed to integrate other potential classification algorithms such as a stacked learning algorithm of trained models.

CONCLUSION

Effective diagnosis and information about the prognosis of Covid-19 are needed to relieve the burden on the health system and at the same time provide patients with the best possible care. Prediction models that combine risk factors and other measures or traits to predict the risk of people becoming infected or suffering a poor outcome from infection can assist medical personnel in classifying patients while allocating limited healthcare resources.

The evaluation of these factors can be done with machine learning methods. However, using various platforms, tools, data pre-processing, model training and tuning methods may lead to major differences among results. In order to support reproducible research on using machine learning methods on evaluating risk factors, we proposed Covid19PredictoR web interface. The interface is open source and uses a simple pipeline that is available for all levels of users including the ones with limited background of coding and machine learning. Our illustrative example showed that the platform can be used for model development and prediction.

Ethical Approval: *Ethics committee approval is not required for the study.*

Conflict of Interest: *The authors have no conflicts of interest to declare.*

Funding: *Sultan Turhan is supported via YOK 100/2000 doctorate program.*

Acknowledgements: *None.*

Author Contribution: Concept: *ED,VK;* ***Desing:*** *VK,ED;* ***Data collecting:*** *VK,ST,ED;* ***Statistical analysis:*** *ED,VK,ST;* ***Literature review:*** *VK,ED,ST,MP;* ***Writing:*** *VK,ED,ST,MP;* ***Critical review:*** *MP,ED,VK.*

REFERENCES

1. Taylor CHV, Johnson M. Wuhan 2019 Novel Coronavirus - 2019-nCoV. Materials and Methods. 2020;10.
2. WHO. Coronavirus disease (Covid-19) pandemic; 2021 [cited 2022 March 11]. Available from: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov.
3. Udugama B, Kadhiresan P, Kozlowski HN, et al. Diagnosing Covid-19: the disease and tools for detection. ACS Nano. 2020;14(4):3822-3835.
4. Rotzinger DC, Beigelman-Aubry C, Von Garnier C, Qanadli S. Pulmonary embolism in patients with Covid-19: time to change the paradigm of computed tomography. Thrombosis Research. 2020;190(C):58-59.
5. Morehouse ZP, Samikwa L, Proctor CM, et al. Validation of a direct-to-PCR Covid-19 detection protocol utilizing mechanical homogenization: A model for reducing resources needed for accurate testing. PLoS ONE. 2021;16(8):e0256316.
6. Li WT, Ma J, Shende N, et al. Using machine learning of clinical data to diagnose Covid-19: a systematic review and meta-analysis. BMC Medical Informatics and Decision Making. 2020;20(1):247.
7. Batista AFM, Miraglia JL, Donato THR, Chiavegatto Filho ADP. Covid-19 diagnosis prediction in emergency care patients: a machine learning approach. MedRxiv. 2020.
8. Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict Covid-19 infection. Chaos, Solitons and Fractals. 2020;140.
9. Soltan AA, Kouchaki S, Zhu T, et al. Rapid triage for Covid-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. Lancet Digit Health. 2021;3(2):e78-e87.
10. Yang HS, Hou Y, Vasovic LV, et al. Routine laboratory blood tests predict SARS-CoV-2 infection using machine learning. Clinical Chemistry. 2020;66(11):1396-1404.
11. Joshi RP, Pejaver V, Hammarlund NE, et al. A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. Journal of Clinical Virology. 2020;129.
12. Kukar M, Gunčar G, Vovko T, et al. Covid-19 diagnosis by routine blood tests using machine learning. Scientific Reports. 2021;11(1):10738.
13. Bayat V, Phelps S, Ryono R, et al. A severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) prediction model from standard laboratory tests. Clinical Infectious Diseases. 2021;73(9):e2901-e2907.
14. Wu J, Zhang P, Zhang L, et al. Rapid and accurate identification of Covid-19 infection through machine learning based on clinical available blood test results. MedRxiv. 2020.
15. 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. Journal of Medical Systems. 2020;44(8):135.
16. Tschoellitsch T, Dünser M, Böck C, Schwarzbauer K, Meier J. Machine learning prediction of SARS-CoV-2 polymerase chain reaction results with routine blood tests. Laboratory Medicine. 2021;52(2):146-149.

17. Shoer S, Karady T, Keshet A, et al. A prediction model to prioritize individuals for a SARS-CoV-2 Test Built from National Symptom Surveys. Med (NY). 2021;2(2):196-208.
18. Tordjman M, Mekki A, Mali RD, et al. Pre-test probability for SARS-Cov-2-related infection score: The PARIS score. PLoS ONE. 2020.
19. Cabitza F, Campagner A, Ferrari D, et al. Development, evaluation, and validation of machine learning models for Covid-19 detection based on routine blood tests. Clinical Chemistry and Laboratory Medicine. 2020;59(2):421-431.
20. Gladding PA, Ayar Z, Smith K, et al. A machine learning program to identify Covid-19 and other diseases from hematology data. Future Science OA. 2021;7(7).
21. Demirarslan M, Suner A. Development of a mobile application by using machine learning methods for the prediction of Covid-19 diagnosis with routine blood tests. Ege Journal of Medicine. 2021;60(4):384-393.
22. Alballa N, Al-Turaiki I. Machine learning approaches in Covid-19 diagnosis, mortality, and severity risk prediction: a review. Informatics in Medicine Unlocked. 2021;24.
23. Syeda HB, Syed M, Sexton KW, et al. The role of machine learning techniques to tackle Covid-19 Crisis: A systematic review. JMIR Medical Informatics. 2020;9(1).
24. Adadi A, Lahmer M, Nasiri S. Artificial Intelligence and Covid-19: A Systematic umbrella review and roads ahead. Journal of King Saud University – Computer and Information Sciences. 2021.
25. Shiny from R Studio. shiny: Web Application Framework for R. 2020 [cited 2020 September 7] Available from: https://cran.r-project.org/web/packages/shiny/index.html.
26. Schwab P, Schütte AD, Dietz B, Bauer S. predCovid-19: A systematic study of clinical predictive models for Coronavirus Disease 2019. Journal of Medical Internet Research. 2020;22(10).
27. Feng C, Wang L, Chen X, et al. A Novel triage tool of artificial intelligence-assisted diagnosis aid system for suspected Covid-19 pneumonia in fever clinics. MedRxiv. 2021.
28. AlJame M, Ahmad I, Imtiaz A, Mohammed A. Ensemble learning model for diagnosing Covid-19 from routine blood tests. Informatics in Medicine Unlocked. 2020;21.
29. Goodman-Meza D, Rudas A, Chiang JN, et al. A machine learning algorithm to increase Covid-19 inpatient diagnostic capacity. PLoS ONE. 2020;15(9):e0239474.
30. Kaggle. Diagnosis of Covid-19 and its clinical spectrum. 2021 [cited 2021 February 16] Available from: https://www.kaggle.com/einsteindata4u/covid19.
31. Alves MA, Castro GZ, Oliveira BAS, et al. Explaining machine learning based diagnosis of Covid-19 from routine blood tests with decision trees and criteria graphs. Computers in Biology and Medicine. 2021;132.