Hybrid Modeling for Predicting Inpatient Treatment Outcome: COVID-19 Case

Simon D. Usoltsev[1[0000-0002-2317-2943]], Ilia V. Derevitskii[1[0000-0002-8624-5046]]

[1] ITMO University, Saint Petersburg, Russia
ivderevitckii@itmo.ru

Abstract. This study presents two methods to support the treatment process of inpatients with COVID-19. The first method is designed to predict treatment outcomes; this method is based on machine learning models and probabilistic graph models of patient clustering. The method demonstrates high quality in terms of predictive models, and the structure of the graph model is supported by knowledge from practical medicine and other studies. The method is used as a basis for finding the optimal intervention plan for severe patients. This plan is a set of interventions for patients that are optimal in terms of minimizing the probability of mortality. We tested the method for critically ill patients (item 4.5) and for 30 percent of all patients with lethal outcomes the methods found an intervention plan that leads to recovery as a treatment outcome as predicted. Both methods show high quality, and after validation by physicians, this method can be used as part of a decision support system for medical professionals working with COVID-19 patients.
Keywords: COVID-19, predict treatment outcome, optimal interventions, Bayesian Network.

1 Introduction

Coronaviruses are a group of viruses that can cause acute respiratory disease in humans and can also be transmitted between species and cause various diseases. A new coronavirus, SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus), which was designated COVID-19 by the World Health Organization (WHO) on February 11, 2020. It causes an acute and deadly illness with a global mortality rate of 1.68%. In Russia, the mortality rate is 2.96% [1]. The level of morbidity and mortality in Russia from Covid-19 and the complications it causes is quite high, which creates a serious burden on the health care system [2], coronavirus can cause severe effects on the heart, brain, kidneys, blood vessels, and other vital organs. Therefore, it is especially important for patients who are hospitalized with this disease to prevent the development of serious complications [3]. The risk of complications can be reduced with the right therapy, as well as enough time that medical professionals can devote to heavy patients, in a heavy workload, to reduce the burden on the medical staff requires special methods to assess the severity of the patient admitted to the hospital, optimal in terms of accuracy and necessary data for decision-making. This article presents a new hybrid approach for assessing the risk of hospital-acquired patient mortality, based on a highly

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_10

interpretable Bayesian network model, with the possibility of editing the model based on expert data as well as predictive machine learning models. There are many approaches in the literature for assessing the risk of covid mortality. The methods are based on several different approaches: statistical, simulation, and dynamic modeling. Using linear prediction, some researchers have used regression modeling to calculate the risk of 15-day mortality and determine predictors of mortality in critically ill patients with covid [4]–[6]. The use of this approach has the following advantages: interpretability, ease of implementation, and availability of many software tools. Among the disadvantages of this approach, it is important to highlight the impossibility of detecting nonlinear relationships. Disadvantages of linear methods are eliminated in the simulation modeling approach. Blagojevic A et al. described the classification of clinical severity of patients based on blood biomarkers, which were selected as biomarkers that have the greatest impact on the classification of patients with Covid-19 [7]. Another approach is to model the course of the disease as a dynamic process. Several studies have used Markov chains Monte Carlo to estimate the mortality rate [8].

2 Model

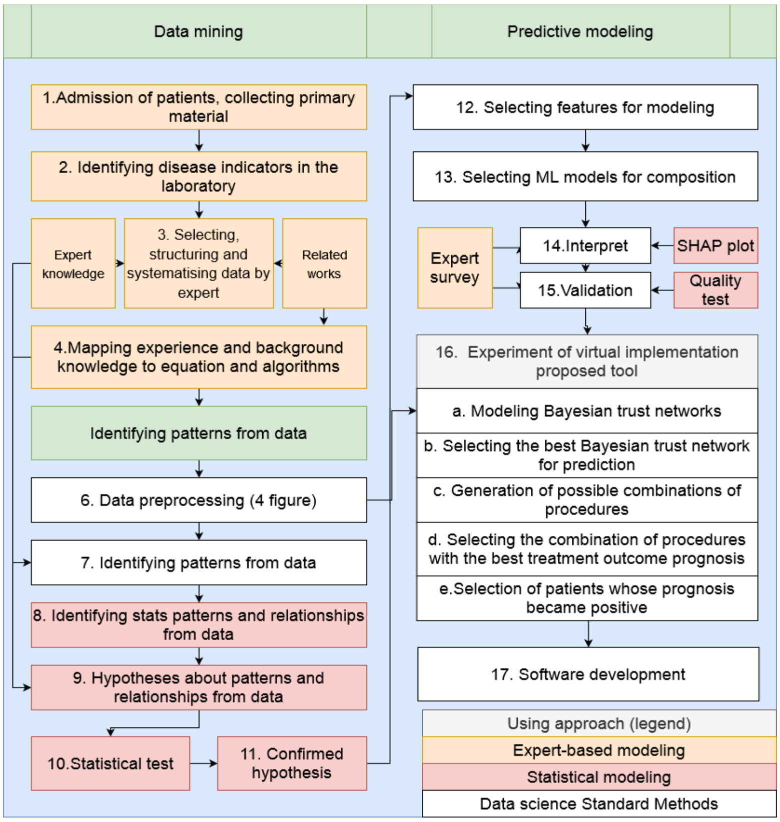

Fig. 1. Scheme of research.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_10

The first step is data mining; our team structures, selects and aggregates information to build hypotheses and models, then we create special scripts to convert information to convert information from medical experts into matrices. The rows are cases of diseases, the columns are important indicators of the course of the disease. At this stage we use only expert modeling. This stage includes steps 1-5. The second step is to identify patterns from the data (steps 6-10), then the scheme. This stage involves processing noise, outliers, removing/substituting gaps in the data, categorical feature coding using one-hot-label-encoding/dummy encoding methods, scaling and protocolization. In the next step, we create a Bayesian network model and machine learning on the resulting data, then use the Bayesian network to select the therapy that reduces the probability of lethality. In the next step, we interpret the models using additive Shapley explanations (SHAP) as well as knowledge from medicine and other studies. We then validate the methods using the classical quality metric of the predictive problem and comparing the results of using the recommended therapy with real-world results.

2.1 Data mining

The study was based on a data set including medical records of 2445 patients treated for covid-19 in the hospital of the V.A. Almazov National Research Center, St. Petersburg, Russia.

Table 1. Signs obtained during treatment in the data

Feature's group	Features
Measurements	Group includes height, weight, age, gender, SBP, DBP, heart rate, body mass index, hemoglobin, and other blood test data.
Chronic diseases	Group includes a history of chronic diseases in the patient
Time information	Group includes time of onset, time of tests, date of complications, and other.
Secondary attributes of the Covid-19	Group includes the presence of heaviness in the chest, lack of sense of smell and taste, fever, decreased consciousness, saturation, etc.
Signs of complications	The group includes the presence of pneumonia, anemia, hypothyroidism, bronchitis, and other.
drugs	The group includes 404 different drugs that were used to treat patients in the hospital. These drugs belong to the pharmacological group of glucocorticosteroids, insulins, anticoagulants, etc.
Procedures	The group includes procedures that were performed to maintain the patient's condition, such as blood transfusions, ventilation, oxygen therapy, and other

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_10

These medical indicators include information on medical history, test values, physical measurements, lifestyle, hospital department metadata, and medical professional opinions.

3 Experimental study

3.1 Mortality Prediction – Bayesian Networks

A Bayesian network is a graph-based probabilistic model, which is a set of variations and their Bayesian probability relations. The constructed Bayesian network was used to predict lethality. The network was constructed using the hill climb search approach with a k2 estimation metric; the approach implements a greedy local search that can be used to find the optimal network structure.

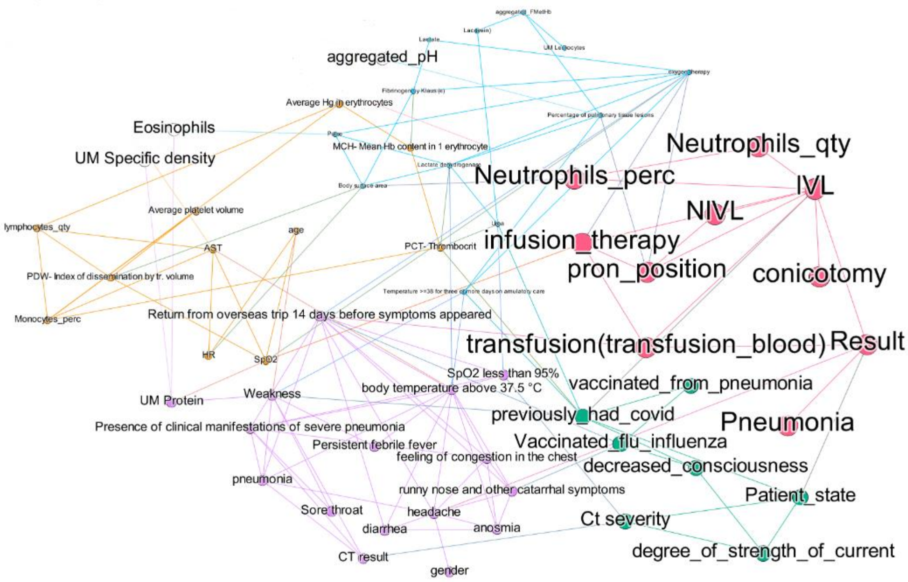

Fig. 8. The Bayesian Trust Network Hillclimb search approach

The model separates the metrics quite qualitatively according to groups of tests, which include attributes such as vein tests, blood test results, procedure correlation, and some patient characteristics. The red cluster localizes the traits that are related to the patient's procedures, the blue cluster localizes the area that is related to oxygen therapy. This area, through signs of weakness and fever above 37.5 is related to the purple area, in which headache and other general patient signs are localized. This area, through the sign of saturation, is related to the yellow sector, in which age, signs with the results of general blood tests and urine tests are localized.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_10

3.2 Prediction experiment with critically ill patients

For patients with a negative outcome, procedures were selected in which the prognosis of the model improved. For several patients and procedures, the model predicted a positive treatment outcome. The recommended procedures can be seen in Figure 10.

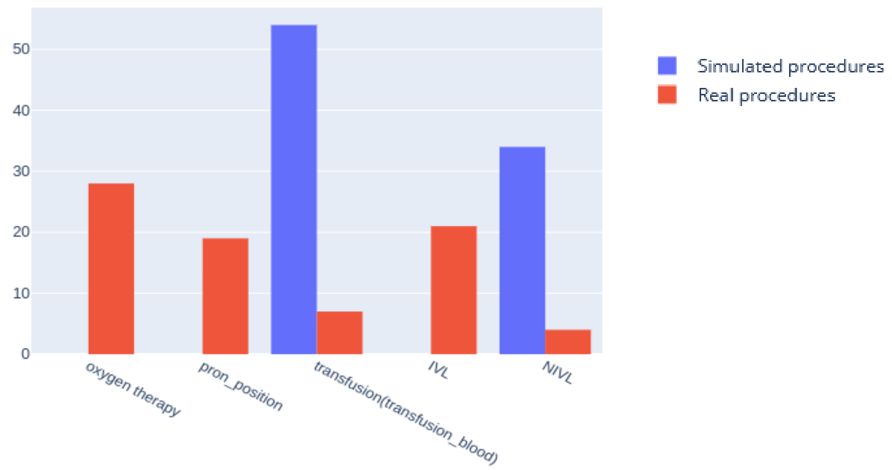

Fig. 9. Recommended procedures that improve prognosis for lethal patients

For these patients, the model recommends more blood transfusion and noninvasive lung ventilation. In several studies, blood transfusion has demonstrated high efficiency, for critically ill patients. On ventilator, after which some of them began to breathe on their own, also, the model recommends noninvasive lung ventilation, in some cases mortality of patients connected to invasive lung ventilation is rather high [9].

3.3 Predictive modeling with machine learning algorithms

Machine learning models were built based on data that were collected during the first 3 days of patients' stay in the hospital. The following algorithms were used: XGBoost, LGBM, SVC, Logistic Regression, Random Forest, Decision tree, KNN, SGD, Gaussian, Bayesian network.

Table 2. The obtained cross-validation accuracy for predictive models

Model	**Cross-validation-score, %**
XGBoost	96.9
LGBM	96.84
KNN	96.8

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_10

SVC	96.6
Logistic Regression	96.5
Decision tree	95.91
SGD	91.97
Bayesian network	91.0
Gaussian	90.2
Random forest	87.2

The XGBoost model showed the highest accuracy score, consider the results of the model. We interpret the results of the model using SHAP values; with their help we can understand the importance of a single trait and its impact on the assessment of the severity of the patient's condition.

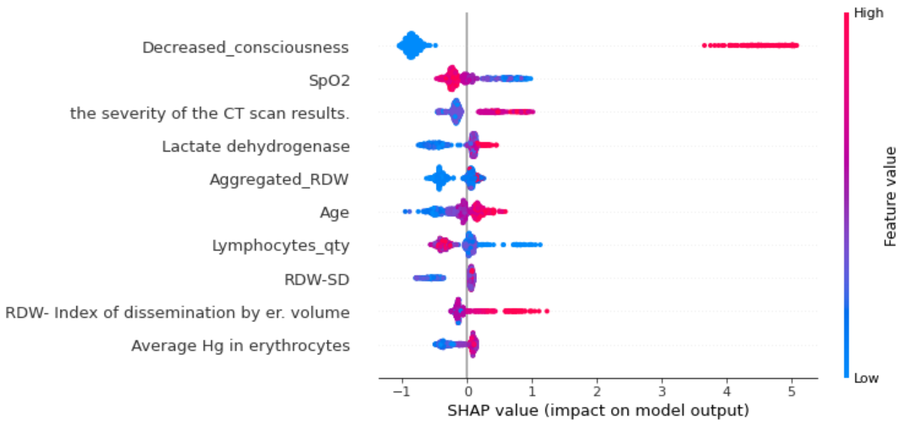

Fig. 11. Interpretation of obtained model values using SHAP values.

The graph shows the 10 indicators that most affect the assessment of the severity of the patient's condition. The most significant indicator for the model turned out to be the patient's decreased consciousness, which is often observed against the background of worsening lung condition. The second and the third significant indicator are the signs that are also related to the respiratory system: oxygen saturation level and CT scan results [10].

5. Conclusion and future work

This paper proposes a method for predicting treatment outcomes in critically ill patients and selecting optimal therapy. We used this method to create a practical tool for recommending intervention plans to reduce the probability of lethal outcome. We have tested the effectiveness of the method on critically ill patients with a negative treatment outcome, and according to the results of the model, for 30% of patients we can achieve

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_10

a change in the results of treatment to recovery. All models show high quality, so they can be used as part of a decision support system for medical professionals working with patients

Acknowledgments. This research is financially supported by The Russian Science Foundation, Agreement №17-71-30029 with co-financing of Bank Saint Petersburg.

References

[1] "Mortality Analyses - Johns Hopkins Coronavirus Resource Center." https://coronavirus.jhu.edu/data/mortality (accessed Jan. 21, 2022).

[2] P. K. Panda and I. K. Sharawat, "COVID-19 and/with dengue infection: A curse in an overburdened healthcare system," *Trop. Doct.*, vol. 51, no. 1, pp. 106–108, Jan. 2021, doi: 10.1177/0049475520975945.

[3] "Coronavirus and the Nervous System | National Institute of Neurological Disorders and Stroke." https://www.ninds.nih.gov/Current-Research/Coronavirus-and-NINDS/nervous-system (accessed Jan. 21, 2022).

[4] O. Y. Bello-Chavolla *et al.*, "Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico," *J. Clin. Endocrinol. Metab.*, vol. 105, no. 8, Aug. 2020, doi: 10.1210/clinem/dgaa346.

[5] N. N. Karyakin *et al.*, "Modernization of regression models to predict the number of deaths from the new coronavirus infection," *Sovrem. Tehnol. v Med.*, vol. 12, no. 4, pp. 6–12, 2020, doi: 10.17691/stm2020.12.4.01.

[6] S. Araç and M. Özel, "A new parameter for predict the clinical outcome of patients with COVID-19 pneumonia: The direct/total bilirubin ratio," *Int. J. Clin. Pract.*, vol. 75, no. 10, Oct. 2021, doi: 10.1111/ijcp.14557.

[7] A. Blagojević *et al.*, "Artificial intelligence approach towards assessment of condition of COVID-19 patients - Identification of predictive biomarkers associated with severity of clinical condition and disease progression," *Comput. Biol. Med.*, vol. 138, Nov. 2021, doi: 10.1016/j.compbiomed.2021.104869.

[8] D. Zhicheng *et al.*, "[Using Markov Chain Monte Carlo methods to estimate the age-specific case fatality rate of COVID-19]," *Zhonghua Liu Xing Bing Xue Za Zhi*, vol. 41, no. 11, pp. 1777–1781, Nov. 2020, doi: 10.3760/CMA.J.CN112338-20200609-00823.

[9] K. Duan *et al.*, "Effectiveness of convalescent plasma therapy in severe COVID-19 patients," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 17, pp. 9490–9496, Apr. 2020, doi: 10.1073/PNAS.2004168117/-/DCSUPPLEMENTAL.

[10] N. A. Chatterjee *et al.*, "Admission respiratory status predicts mortality in COVID-19," *Influenza Other Respi. Viruses*, vol. 15, no. 5, pp. 569–572, Sep. 2021, doi: 10.1111/IRV.12869.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_10