



## Original Article

# Performance Analysis of Selected Decision Tree Algorithms for Predicting Drug Adverse Reaction among COVID-19 Hospitalized Patients

Raof Nopour<sup>1</sup>, Mehrnaz Mashoufi<sup>2</sup>, Morteza Amraei<sup>3</sup>, Nahid Mehrabi<sup>4</sup>, Alireza Mohammadnia<sup>5</sup>, Abdollah Mahdavi<sup>2</sup>, Nader Mirani<sup>6</sup>, Mojgan Saki<sup>7</sup>, Mostafa Shanbehzadeh<sup>8,\*</sup>

<sup>1</sup>Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

<sup>2</sup>Department of Health Information Management, Ardabil University of Medical Sciences, School of Medicine and Paramedical Sciences, Ardabil, Iran

<sup>3</sup>Department of Health Information Technology, School of Allied Medical Sciences, Lorestan University of Medical Sciences, Khorramabad, Iran

<sup>4</sup>Department of Health Information Technology, Aja University of Medical Sciences (AJAUMS), Tehran, Iran

<sup>5</sup>Department of Health information Management, School of Medicine, Ardabil University of Medical Sciences, Ardabil, Iran

<sup>6</sup>Vice Chancellor for Treatment, Head of Medical Tourism, Zanjan University of Medical Sciences, Zanjan, Iran

<sup>7</sup>Department of Operating Room, Faculty of of Allied Medical Sciences, Lorestan University of Medical Sciences, Khorramabad, Iran

<sup>8</sup>Department of Health Information Technology, School of Paramedical, Ilam University of Medical Sciences, Ilam, Iran

## ARTICLE INFO

### Article history

Received: 2021-11-28

Received in revised: 2022-01-11

Accepted: 2022-01-14

Manuscript ID: [JMCS-2111-1352](#)

Checked for Plagiarism: **Yes**

Language Editor:

[Ermia Aghaie](#)

Editor who approved publication:

[Dr. Ali Delpisheh](#)

DOI:10.26655/JMCHMSCI.2022.4.7

## KEYWORDS

COVID-19

Coronavirus

Artificial intelligence

Machine learning

Drug therapy

Adverse effects

## ABSTRACT

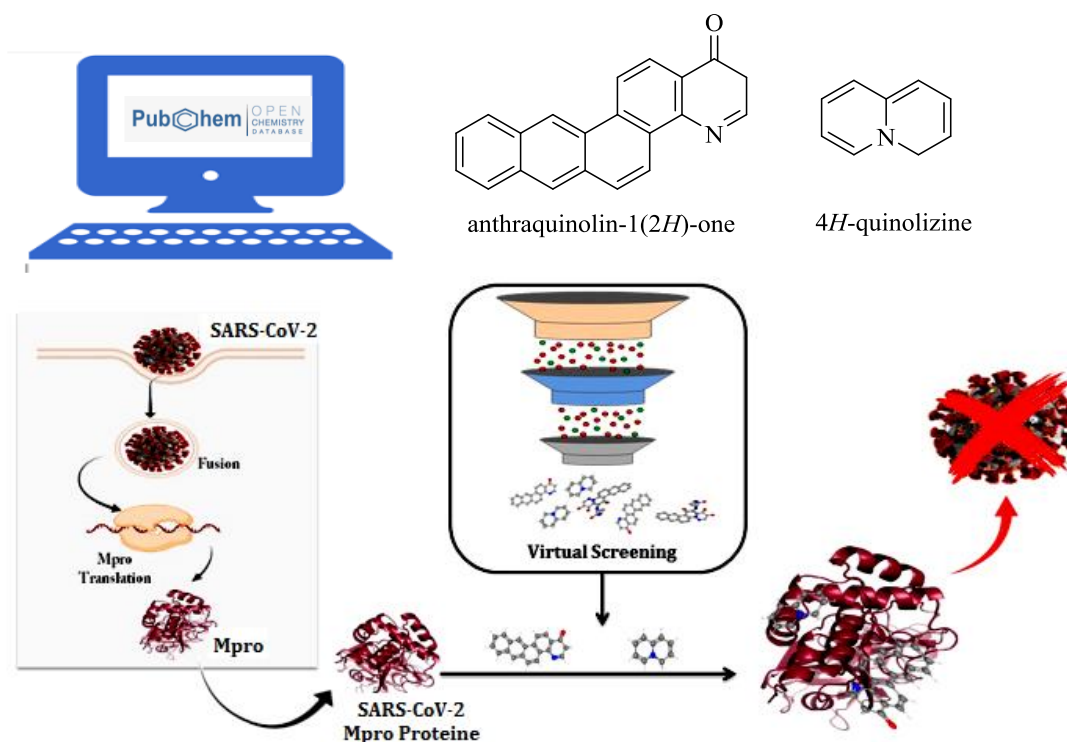
Increase in drug allergies and unpleasant adverse effects caused by COVID-19 medication therapies has doubled the need for computing technologies and intelligent systems for predicting poor medication outcomes. This study aimed to construct machine learning (ML) based prediction models to better predict adverse drug effects among COVID-19 hospitalized patients. In this retrospective and single-center study, 482 hospitalized COVID-19 patients were used for analysis. First, the Chi-square test was employed to determine the most critical factors predicting adverse drug effects at  $P < 0.05$ . Second, the four selected decision tree (DT) algorithms were applied to implement the model. Finally, the best DT model was acquired for predicting adverse drug effects using various performance criteria. This study showed that the 18 variables gained the Chi-square at  $P < 0.05$  as the most important factors predicting adverse drug reactions. Besides, comparing the performance of selected algorithms demonstrated that generally, the J-48 algorithm with F-Score=94.6% and AUC=0.957 was the best classifier predicting adverse drug reactions among hospitalized COVID-19 patients. Finally, it found that the J-48 algorithm enables a reasonable level of accuracy in predicting the risk of harmful drug effects among COVID-19 hospitalized patients. It potentially facilitates identifying high-risk patients and informing proper interventions by the clinicians.

\* Corresponding author: Mostafa Shanbehzadeh

✉ E-mail: Email: [mostafa.shanbehzadeh@gmail.com](mailto:mostafa.shanbehzadeh@gmail.com)

© 2022 by SPC (Sami Publishing Company)

## GRAPHICAL ABSTRACT



## Introduction

Today, Coronavirus 2019 (COVID-19), also known as Acute Respiratory Syndrome, is commonly dubbed as SARS-COV-2, which, in turn, has affected millions of people worldwide. A diverse and multidimensional clinical picture determines the disease and its incubation period is 14 days, and on average, symptoms appear within 4-5 days after exposure [1-3]. The disease manifestation can be ranged from asymptomatic to severe pneumonia, acute respiratory distress syndrome (ARDS), and even death [4]. Reportedly, approximately 80% of COVID-19-stricken patients suffer from asymptomatic or mild to moderate symptoms, and about 15% of patients present with severe symptoms and are referred for critical care units [5, 6]. It is reported that the patient's condition becomes critical and may require hospitalization in intensive care units (ICU) in 5% of cases. Despite widespread vaccination in the population, the prevalence of new emerging disease cases characterized by much more contagious species is still ever-increasing [7]. It is said that aging, male gender, obesity (BMI > 40), underlying diseases, and hypoxemia are essential factors that considerably exacerbate the condition. The severe or acute stage of the disease is characterized by severe

complications such as ARDS, cytokine syndrome, and multiple-system organ dysfunction (MOF) [5, 6]. In the meantime, many hospitalized patients with advanced stages need drug therapy to prevent patient deterioration and reduce respiratory complications [8]. So far, several drugs have been proposed to avoid severe complications and mortality caused by COVID-19. However, unfortunately, no approved drug therapy has been discovered to treat COVID-19 [9]. Many drugs have lacked significant effects due to the complex, unknown and mutable nature of the disease [10]. On the other hand, physicians have reported problems in predicting the COVID-19 drug adverse effects [11]. This requirement is more pronounced, especially concerning the increase number of drug adverse reactions and current unpredictability of the disease behavior and courses [11,12]. Therefore, designing predictive models of COVID-19 adverse effects possibility can increase drug therapy quality by reducing severe drug allergies and interactions. To deal with this problem, the design and implementation of clinical decision support systems (CDSS) based on artificial intelligence (AI) will be of great importance, accordingly [13-16]. Machine learning (ML) algorithms, a subset of artificial intelligence, are extensively utilized to

screen, diagnose, prognosis, and predict COVID-19 outcomes [17, 18]. These mathematics models could quickly combine and analyze large volumes of data. Besides, ML algorithms are applied to generate predictive models that can be used to support and improve clinical decision-making for a wide range of outcomes [19-22]. It should be noted that a great deal of ML-based models was developed to estimate the COVID-19 severity, patient deterioration [17,23], ICU admission [23-27], and mortality [24,25,28-34] in previous studies. Unfortunately, few studies have been conducted on applying ML techniques to predict

adverse drug effects among COVID-19 hospitalized patients.

Therefore, the present study was performed to establish and compare several ML-based predictive models to estimate adverse drug reactions among COVID-19 hospitalized patients.

### Material and Methods

This study aimed to predict the adverse drug effect among COVID-19 hospitalized patients and was carried out in four stages as follow:

**Table 1:** Characteristics related to drug adverse effect prediction

Inputs	Factors	Variables
	Demographical	Age (year), height (cm) <sup>1</sup> , weight (Kg) <sup>2</sup> , blood type, and gender
	Laboratory data	Blood creatinine (mg/dL)[3], red cell count (Cells/L)[4], white cell count (Cells/L)[4], hematocrit (L/L)[5], hemoglobin rate (g/dL)[6], platelet count (Cells/ $\mu$ L)[7], absolute lymphocyte count ( $10^3$ Cells/ $\mu$ L)[7], absolute neutrophil count (Cells/ $\mu$ L)[7], blood calcium (mg/dL)[3], blood sodium (mEq/L)[8], magnesium (mEq/L)[8], blood phosphor (mg/dL), blood potassium (mEq/L), blood urea nitrogen (mg/dL)[3], total bilirubin (mg/dL)[3], aspartate aminotransferase (units/L)[9], alanine aminotransferase (units/L)[9], serum albumin (g/dL)[10], blood glucose (mg/dL)[3], lactate dehydrogenase (Units/L)[9], activated partial thromboplastin time (S)[11], prothrombin time (S)[11], alkaline phosphatase (Units/L)[9], C-reactive protein (mg/L)[12], erythrocyte sedimentation rate (mm/hr)[13], and hypersensitive troponin (n/L)[14].
	Clinical manifestations	Cough, contusion, nausea, vomit, headache, gastrointestinal symptoms, muscular pain, chill, fever, temperature, pneumonia, dyspnea, loss of taste, loss of smell, rhinorrhea, pleural fluid, consolidation, and sore of the throat.
	Prescription	Current drug therapy (COVID-19 drug), drug dose, drug history (prior), drug allergy
	Epidemiological	Smoking, alcohol consumption
	History of diseases	Blood pressure, diabetes, cardiac diseases, and other underline diseases.
	Hospitalization	Length of hospitalization and ICU hospitalization.
Output	Class	Modes
	drug side / harmful effects	NO: 0 and Yes: 1

1- Centimeter, 2- Kilogram 3-Milligram per deciliter, 4-Number of cells per liter, 5-Number red cells per liter per number of cells per liter, 6-Grams per deciliter, 7- Number of cells per microlitre, 8- Miliequivalents per liter, 9- Units per liter, 10-Grams per deciliter, 11-Seconds, 12- Milligrams per liter, 13-Millimeters per hour, 14- Nanograms per liter

### Collectin Data

In this study, a COVID-19 hospital-based Electronic Medical Record (EMR) database from Mostafa Khomeini medical center, affiliated with Ilam University of medical sciences, West of Iran,

was investigated retrospectively, February 9, 2020, to July 20, 2021. During this period, a total of 2854 suspected cases with COVID-19 had been referred to this center, of whom 853 cases were detected as positive COVID-19, and their clinical

data were recorded in the EMR database based on six categories and 60 features. These features were categorized as laboratory (25 variables), demographic (five variables), clinical manifestations (18 variables), prescription (four variables), history of diseases (four variables), epidemiologic (two variables), and hospitalization (two variables), which are depicted in Table 1. The output variable was classified into two groups: hospitalized COVID-19 patients with adverse drug effects (code 1) and those who hadn't these conditions (code 0).

#### *Dataset Preparing and Analyzing*

In this study, first, two experienced health information managers (M.SH and R.N) investigated all the case records, consulting two infectious and internal specialists regarding the quantitative and qualitative attributes of medical documentation and appropriateness for statistical analysis. Therefore, the samples owning more than 70% missing values that had no substantial role in statistical analysis were excluded from the study. In the next step, for the cases with less than 70% missing values, we used the two methods of the replacement by averaged values and K-nearest neighborhood (KNN) with specific amounts of K in Rapid Miner V 7.1 software for embedding the missing quantitative and qualitative values, respectively. Finally, to obtain the most relevant features as the best variables for drug adverse effect prediction and reduce the dataset dimensions, we used the feature selection process in this respect. Due to the enormous amount of data with many unrelated attributes in databases, this process is essential in data science and data mining applications because of its capability to gain pertinent features and eliminate useless data elements (35-37). Some advantages of this process can be enumerated as 1- removing irrelevant attributes, 2- clustering dataset using more related features, 3- augmenting algorithms performance, 4- reducing training time, more understandable data mining results, and 5- preventing form overfitting (38-40). In this study, the independence Chi-square test ( $\chi^2$ ) has been considered for determining the highly associated features with the dependent variable (adverse

drug effect). The  $P < 0.05$  was considered the statistically significant level in this regard.

#### *Decision tree models*

For building the predictive models for drug adverse effect prediction among hospitalized COVID-19 patients, we have used four selected decision tree algorithms because of the high usage of these algorithms in recent articles with the best performance as follows:

J-48: This algorithm is also known as C4.5 and is considered as the expansion of the ID3 decision tree with the capabilities, such as perpending the missing values for classifying the samples, adjusting the tree size for pruning using the confidence factor, extracting the rules, and organizing the selections having continuous numerical value ranges. This algorithm also possesses a pleasant balance between accuracy and sample classification capabilities using the pruning characteristics in which samples are classified completely until the tree is complete. In other words, in this algorithm, the overfitting will be prevented, and rules are generated using the specialized cognizance generated using the dataset itself. The J-48 decision tree algorithm builds the tree using the entropy concept. Suppose that the training dataset includes samples ( $S = S_1 + S_2 + S_3 + \dots + S_n$ ) and every instance has a p-dimensional vector ( $X_{1,1}, X_{2,1}, X_{3,1}, \dots, X_{p,1}$ ), so in this regard,  $X_j$  demonstrates the feature that the  $S_i$  sample will be prolapsed. This algorithm uses elements with higher entropy differences than others for tree splitting. So it has samples categorizing capability with the high distinction that will be existed between the subtrees when having the classified samples with the highest frequency based on different output classes in their leave nodes [41-43].

Random-Forest: The random forest is a hybrid decision tree algorithm including various subtrees as classifiers with specified depths and nodes. This algorithm has reasonable flexibility for making the decision trees utilizing the multiple features for splitting the trees randomly. The accuracy of the random forest decision tree algorithm depends on each subtree's accuracy in predicting the classes. Its performance also depends on the number of the subtrees' votes

that existed between them; in other words, the model's performance will be considered the performance of the majority of the subtrees. This decision tree has common capabilities in classifying the samples with high performance, embedding the outliers and noisy values in features, and preventing the overfitting of the algorithm [44-47].

**Decision stump:** The decision stump owns only one layer, including the root node directly connected with the leaves node, in contrast with other decision tree algorithms having three-node layers (root, internal, and leave), and also the splitting process will be stopped after the first split in this structure. The most common application of this classifier is related to mining in large dimension datasets. Still, they also can be utilized in smaller dimensions of the dataset with binary splitting. It is also known as single-rule because it predicts the output class with just the values in one variable as a predictor [48-50].

**Hoeffding tree:** This tree is an additive decision algorithm common for an extensive dataset. This algorithm is a primitive decision tree algorithm with inflexibility in data set dimension variations in massive datasets. This algorithm possesses a potential advantage for selecting the highly differentiating attribute with finite Hoeffding. The Hoeffding decision tree algorithm is based on the Hoeffding limitation; in other words, by considering enough attributes occurrence, the range of random variable changeability is the predictable amount known as the Hoeffding limitation. In this decision tree algorithm, the samples categorization process can be done using a specified number with a predetermined fitness [51-53].

#### *Analyzing decision tree algorithm's performance*

After developing selected decision tree algorithms using the most common technical parameters for recognizing the best algorithm to predict adverse drug reactions among hospitalized COVID-19 patients, the performance of these algorithms has been compared and evaluated. First, the confusion matrix (Table 2) was utilized to compare the algorithm's performance and sample classification strength. True Positive (TP) and True Negative (TN)

representing hospitalized COVID-19 patients with the drug side effects (P) and didn't have (N) and were correctly classified by algorithms, respectively. False Positive (FP) and False Negative (FN) are the positive and negative cases incorrectly classified by algorithms, respectively. Also, based on the confusion matrix, the TP-Rate, FP-Rate, Precision, Recall, F-Measure, and Area under the Receiver Operator Characteristics (ROC) of each selected decision tree algorithm have been calculated for measuring and evaluating the decision tree algorithm's performance. Also, the ten-fold cross-validation has been considered for embedding errors when measuring the algorithms' performance. Finally, the best decision tree algorithm has been obtained using these different evaluation criteria for predicting adverse drug effects and also was drawn. Afterward, the most important rules with the structure of IF-THEN have been extracted from the tree and then interpreted as the essential clinical knowledge for predicting adverse drug effects among hospitalized COVID-19 patients with the most frequency of classified samples.

#### **Result and Dissection**

After excluding the samples that owned 70% or higher missing, noisy, and abnormal values in their attributes and applying the exclusion criteria such as records belonged to patients less than 18 years old, discharged, or died in emergency departments, 371 records were excluded from the study finally. Therefore, 482 records have remained for statistical analysis. Among them, 176 (36.5%) records have belonged to hospitalized COVID-19 patients with drug side effects, and 306 (63.5%) of them were associated with patients who hadn't them. The 227 (47.1%) records have belonged to men with the mean age of  $50 \pm 12.5$  years, and 255 (52.9%) of them were associated with women with the mean age  $52 \pm 11.7$  years. In Table 2, the results of using the independence Chi-square test ( $\chi^2$ ) for determining the most important factors predicting the drug side/ adverse effect among hospitalized COVID-19 patients at  $P < 0.05$  have been represented.

**Table 2:** The most critical factor predicting drug side/ adverse effect at  $P < 0.05$ 

NO	Variable	Type	Frequency or Mean $\pm$ SD	$\chi^2$	P (Level)
1	Length of hospitalization	Numeric	5.03 $\pm$ 2.188	89.758	<0.01
2	Cough	Binominal	Yes (353) No (129)	13.538	<0.01
3	Nausea	Binominal	Yes (145) No (337)	9.505	<0.01
4	Vomiting	Binominal	Yes (130) No (352)	4.074	<0.05
5	Contusion	Binominal	Yes (180) No (302)	10.130	<0.01
6	Oxygen therapy	Binominal	Yes (437) No (45)	45.817	<0.01
7	Dyspnea	Binominal	Yes (442) No (40)	6.813	<0.01
8	Loss of taste	Binominal	Yes (124) No (358)	28.627	<0.01
9	Loss of smell	Binominal	Yes (137) No (345)	29.681	<0.01
10	rhinorrhea	Binominal	Yes (280) No (202)	16.586	<0.01
11	Blood pressure	Binominal	Yes (189) No (293)	10.835	<0.01
12	Blood sodium	Numeric	138.271 $\pm$ 3.44	44.461	<0.01
13	Blood potassium	Numeric	5.491 $\pm$ 8	13.407	<0.05
14	Activated partial thromboplastin time	Numeric	35.453 $\pm$ 9.25	118.196	<0.01
15	Absolute lymphocytic count	Numeric	21.702 $\pm$ 12.01	10.684	<0.01
16	White blood cell count	Numeric	9684 $\pm$ 124.17	53.134	<0.01
17	Cardiac disease	Nominal	Yes (157) No (325)	3.514	<0.05
18	Pneumonia	Nominal	Yes (413) No (69)	17.775	<0.01

Based on the information represented in Table 3, the 18 variables obtained the specific Chi-square at  $P < 0.05$  as the final predictors. Also, the 15 variables had the  $\chi^2$  at  $P < 0.01$  in this respect. We obtained the hospitalization length ( $\chi^2=89.758$ ) ( $P < 0.01$ ), white blood cell count ( $\chi^2=53.154$ ) ( $P < 0.01$ ), and activated partial thromboplastin time (118.196) ( $P < 0.01$ ) as the best factors at  $P < 0.01$ . They were considered the best factors were predicting adverse drug effects among hospitalized COVID-19 patients. The results of classifying the samples for selected decision tree algorithms using the confusion matrix are presented in Table 3.

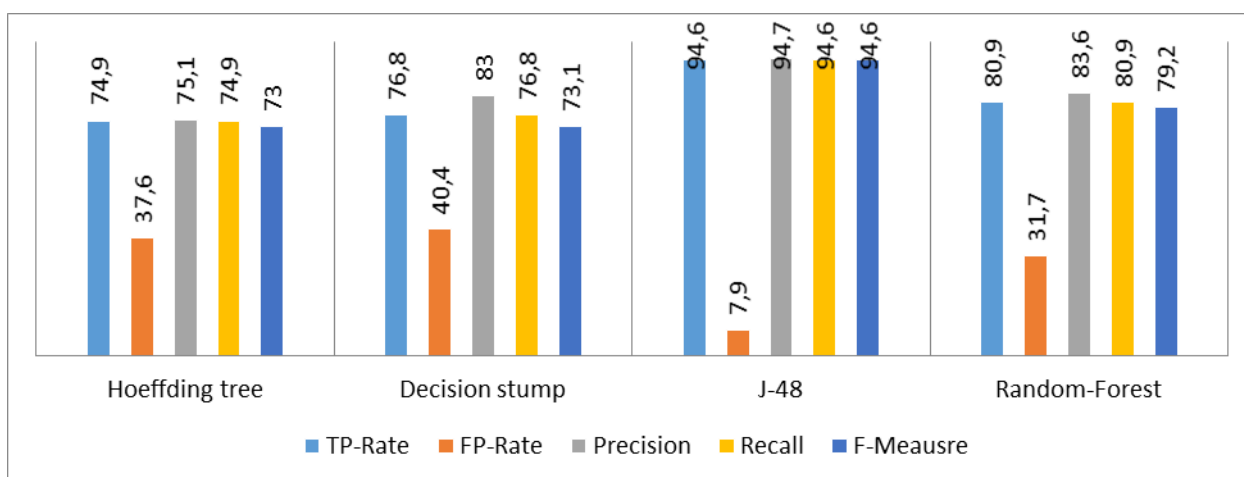
Based on the information given in Table 4, the Decision stump with  $TN=306$  and  $FP=0$  by

classifying all the negative samples (non-affected cases) obtained better performance than other decision tree algorithms. Also, the Random Forest and J-48 decision tree algorithms with  $TN=300$  and  $FP=6$  have gotten a pleasant performance with a meager difference rather than the Decision stump classifier. On the contrary, the J-48 decision tree algorithm with  $TP=156$  and  $FN=20$  could best classify the positive cases (having drug adverse effects) and performed considerably better than other algorithms. The results of comparing the decision tree algorithms performance using different evaluation criteria are demonstrated in Figure 1



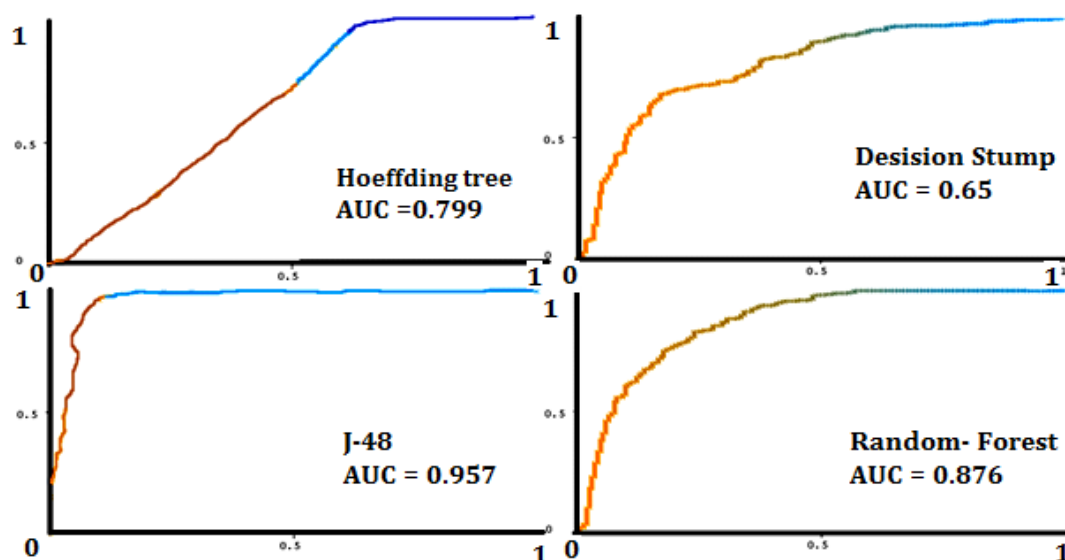
**Table 3:** Results of classifying samples using the confusion matrix

NO	Decision tree algorithms	Technical attributes	TP	FN	FP	TN
1	Hoeffding Tree	Grace period=200 Leaf prediction strategy= Naïve Bayes adaptive Hoeffding Tie threshold=0.05 Split confidence=1.0E-7 Split criterion=Info gain	80	96	25	281
2	Decision stump	Number of batch Size=100 Number of decimal places=2 Minimum number of instances =1	64	112	0	306
3	J-48	Confidence factor=0.2 Number of folds=3 Number of seeds=1 Binary splits=false	156	20	6	300
4	Random Forest	Maximum depth=3 Number of randomly attributes chosen=3 Number of iterations= 100 Break Tie randomly=false	90	86	6	300

**Figure 1:** Different indicators of algorithms performance evaluation

The results of comparing the decision tree algorithms performance using Figure 1 demonstrated that the J-48 decision tree algorithm with TP-Rate=94.6%, FP-Rate=7.9%, Precision=94.7%, Recall=94.6%, and F-Measure=94.6% acquired the best performance generally. Also, the

Hoeffding decision tree algorithm with TP-Rate=74.9%, Precision=75.1%, Recall=74.9%, and F-Measure=73% had the worst performance. In Figure 2, the ROC of all selected decision tree algorithms is depicted (the vertical and horizontal vertices demonstrate the TPR and FPR vertices, respectively).



**Figure 2:** The ROC of all selected decision tree algorithms

Comparing selected decision tree algorithms performance using the ROC diagram demonstrated that the J-48 algorithm with TP-Rate=94.6%, FP-Rate=7.9%, Precision=94.7%, Recall=94.6%, F-Measure=94.6%, and AUC=0.957 had the best capability than others and was considered as the best decision model for predicting the adverse drug effect among hospitalized COVID-19 patients. In conclusion, the J-48 decision tree algorithm has been drawn, and the essential clinical knowledge with the IF-THEN structure was extracted and interpreted with more detail. We pruned the tree using the confidence factor to reduce the tree's size for better understanding and extracting the clinical knowledge from the decision tree. We decreased it to 0.1 (Figure 3). The most important technical features for building the decision tree were batch size=100, binary splits=false, collapse tree=true, confidence factor=0.1, the minimum number of instances per leave=2, number of folds=3, number of seeds=1, use Laplace=false.

Some of the most important clinical rules with the most classified samples have been brought in as follow:

1. IF (Activated partial thromboplastin time  $\leq 31$ ) THEN (drug side/ adverse effect =1),
2. IF (Activated partial thromboplastin time  $> 31$ ) (Length of hospitalization  $\leq 6$ ) (Loss of taste =No) THEN (drug side/ adverse effect =0),

3. IF (Activated partial thromboplastin time  $> 31$ ) (Length of hospitalization  $> 6$ ) (White cell count  $> 8700$ ) THEN (drug side/ adverse effect =1).

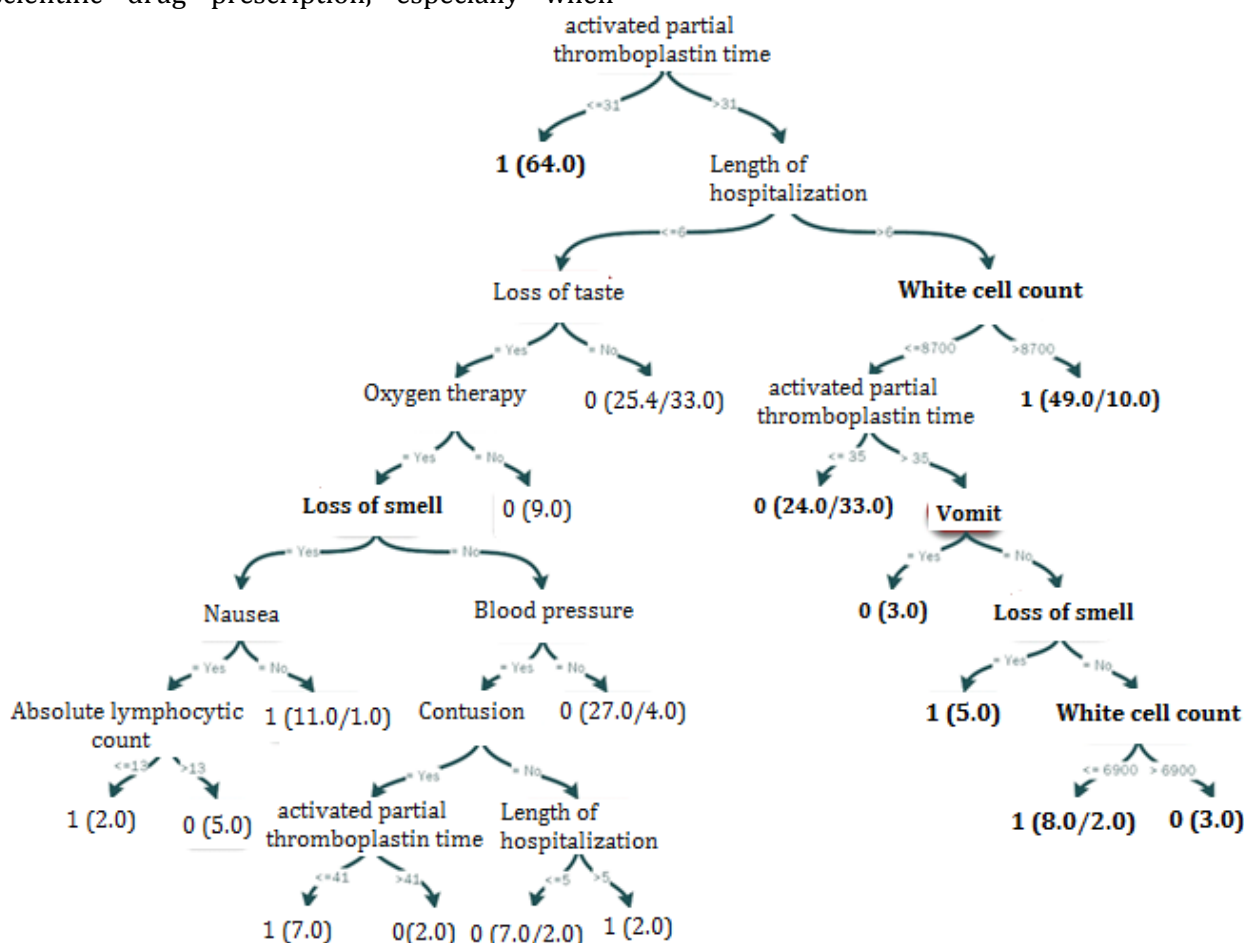
In the J-48 decision tree algorithm, activated partial thromboplastin time was considered the best variable for predicting drug side/ adverse effects with the highest info gain. Therefore, it was placed at the root node. Rule1 states that in hospitalized COVID-19 patients having activated partial thromboplastin time less than 31 seconds, the hospitalized COVID-19 presumably had the adverse drug effects, and 64 samples of the study have confirmed this pattern. In Rule 2, hospitalized COVID-19 patients with more than 31 Activated partial thromboplastin time seconds hospitalized less than six days and didn't lose their taste sensation, hadn't drug adverse effects with the probability of 87%. According to Rule 3, the negative drug effect existed among the 39 hospitalized COVID-19 patients. They have more than 31 activated partial time seconds and more than six days of hospitalization, and more than 8700 counts of white cells.

Given the wide range of clinical manifestations of COVID-19, it is crucial to develop intelligent models for predicting the likelihood of adverse drug effects using ML techniques [54]. Therefore, we examined four selected decision tree ML-based models on important parameters obtained from the independence test of Chi-square. The decision tree models used here included the J-48, Random forest, Hoeffding tree, and Decision stump, applied upon 482 confirmed RT-PCR



COVID-19 patients. Finally, our results showed that the J-48 classifier performed better than the other selected ML algorithms with an F-score of = 94.6% and AUC = 0.957. Necessarily, treating patients with COVID-19 requires informed and scientific drug prescription, especially when

hospitals are faced with an increasing number of patients and a shortage of care facilities [55, 56]. In this regard, physicians state that they encounter problems in predicting the likelihood of adverse drug effects [57].



**Figure 3:** The pruned J-48 decision tree algorithm

To deal with this problem thus, the design and implementation of CDSS based on AI will be precious for the optimal drug prescription and support for clinical decisions [17, 58]. For instance, ML-equipped CDSSs could assist clinicians in making clinical decisions by alerting caregivers and recommending interventions based on objective and generalizable empirical data [59]. This study showed that ML algorithms, especially the J-48 classifier, may predict the drug side and adverse effects in patients hospitalized with COVID-19.

To date, some studies have evaluated the application of ML techniques in predicting the poor and adverse outcomes of drug prescription among COVID-19. For example, Ganguli *et al.* (2021) developed an intelligent system based on

ML algorithms to predict medication error in hospitalized patients with COVID-19 using 1023 patient data. They reported the best performance for the J-48 algorithm with AUC = 0.84 [60]. Besides, Behery (2021) analyzed the data of 5643 negative and positive samples of COVID-19 to predict drug allergy in individuals using selected ML models. The results showed that the J-48 algorithm represented an acceptable detection power with 86% accuracy [61]. Accordingly, Lv *et al.* (2021) evaluated the performance of four ML algorithms to predict the adverse drug effects using information gathered from 3841 COVID-19 cases. Finally, the J-48 model with AUC = 0.92 was introduced as the most suitable algorithm [62]. Siqueira *et al.* (2021) evaluated four ML algorithms to predict the likelihood of patient

deterioration of patients with COVID-19 after drug therapy. Ultimately, the J-48 model with the best AUC of 0.92 was introduced as the superior algorithm [8]. In the present study, the results showed that the J-48 decision tree algorithm with F-Score = 64.6% and AUC = 0.957 enjoys the best capacity for early prediction of drug side and adverse effects in hospitalized patients with COVID-19. The high predictive indices reported by our J-48 model showed that this algorithm can differentiate between high-risk and low-risk patients.

The main advantage of the present study is that we predicted the possibility of prescription side effects based on the most appropriate variables derived from the independence test of Chi-square. Nevertheless, the present study also had some limitations. First, we analyzed a retrospective and single-center data set with a limited sample size. Second, continuous changes in some crucial variables should be thoroughly observed to accurately identify patients at higher risk of poor outcomes on time. Finally, the selected data set lacked clinically essential variables such as radiological indicators. In the future, if one intends to develop more ML techniques in a more significant, multicenter, and futuristic data set equipped with more quantitative and reliable data, the accuracy of the model performance and its generalizability will increase accordingly.

## Conclusions

In this research study, the data recorded in the selected hospital database were analyzed. Then, ML models were developed and tested to predict the possible drug side and adverse effects considering 18 clinical features. The results revealed the acceptable performance of the J-48 decision tree model. Therefore, the developed predictive models can be demanding in providing quality of care, diminishing the workload of the care team, minimizing prescription errors, increasing the quality of care, and rendering patient-centered treatments.

## Acknowledgments

This article is extracted from a research project supported by the Ilam University of Medical

Sciences (Ethical code: IR.MEDILAM.REC.1399.294). We also thank the Ilam University of Medical Sciences Research Deputy for financially supporting this project.

## Funding

This research received financial support from Ilam University of Medical Sciences.

## Authors' contributions

All authors contributed toward study designing, data mining, statistical analysis, reporting the results and agreed to be responsible for all the aspects of this work.

## Conflict of Interest

The author(s) declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

## ORCID

Mostafa Shanbehzadeh:

<https://www.orcid.org/0000-0002-3419-1947>

## References

- [1]. Huang C., Wang Y., Li X., Ren L., Zhao J., Hu Y., Zhang L., Fan G., Xu J., Gu X., *lancet.*, 2020, **395**:497 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [2]. Bolourani S., Brenner M., Wang P., McGinn T., Hirsch J.S., Barnaby D., Zanos T.P., *J. Med. Internet Res.*, 2021, **23**:e24246 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [3]. Arian M., Valinejadi A., Vakilian F., *Iran. J. Public Health*, 2021, **50**:46 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [4]. Da Rosa Mesquita R., Francelino Silva Junior L.C., Santos Santana F.M., Farias de Oliveira T., Campos Alcântara R., Monteiro Arnozo G., Rodrigues da Silva Filho E., Galdino dos Santos A.G., Oliveira da Cunha E.J., Salgueiro de Aquino S.H., *Wien. Klin. Wochenschr.*, 2021, **133**:377 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [5]. Zhou F., Yu T., Du R., Fan G., Liu Y., Liu Z., Xiang J., Wang Y., Song B., Gu X., *lancet*, 2020, **395**:1054 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [6]. Chow N., Fleming-Dutra K., Gierke R., Hall A., Hughes M., Pilishvili T., *MMWR Morb Mortal Wkly*

- Rep., 2020, **69**:382 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [7]. Bath C., Heger U., Petrovsky N., Senanayake S., Frydenberg J. *New vaccine*, 2021 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [8]. De Siqueira Santos S., Torres M., Galeano D., del Mar Sánchez M., Cernuzzi L., Paccanaro A., 2021, 100396 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [9]. Pawar A.Y., *Int. J. Antimicrob. Agents.*, 2020, **56**:105984 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [10]. Brimacombe K.R., Zhao T., Eastman R.T., Hu X., Wang K., Backus M., Baljinnyam B., Chen C.Z., Chen L., Eicher T., *BioRxiv*, 2020 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [11]. Li X., Liao H., Wen Z., *Appl. Soft Comput.*, 2021, **99**:106879 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [12]. Aljouie A.F., Almazroa A., Bokhari Y., Alawad M., Mahmoud E., Alawad E., Alsehawi A., Rashid M., Alomair L., Almozaai S., *J. Multidiscip. Healthc.*, 2021, **14**:2017 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [13]. Hawkins A., Stapleton S., Rodriguez G., Gonzalez R.M., Baker W.E., *West. J. Emerg. Med.*, 2021, **22**:678 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [14]. Hassanzad M., Orooji A., Valinejadi A., Velayati A., *Electron. Physician*, 2017, **9**:5974 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [15]. Askari-Majdabadi H., Valinejadi A., Mohammadpour A., Bouraghi H., Abbasy Z., Alaei S., *Acta Inform. Med.*, 2019, **27**:311 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [16]. Sadoughi F., Valinejad A., Vakilmofrad H., Mohammadhassanzadeh H., Bouraghi H., *Inf. Sci. Technol.*, 2012, **27** [[Google Scholar](#)], [[Publisher](#)]
- [17]. Assaf D., Gutman Y., Neuman Y., Segal G., Amit S., Gefen-Halevi S., Shilo N., Epstein A., Mor-Cohen R., Biber A., *Intern. Emerg. Med.*, 2020, **15**:1435 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [18]. Booth A.L., Abels E., McCaffrey P., *Mod. Pathol.*, 2021, **34**:522 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [19]. Yan L., Zhang H., Goncalves J., Xiao Y., Wang M., Guo Y., Sun C., Tang X., Jin L., Zhang M., 2020 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [20]. Wynants L., Van Calster B., Collins G.S., Riley R.D., Heinze G., Schuit E., Bonten M.M., Dahly D.L., Damen J.A., Debray T.P., De Jong V.M., *BMJ*, 2020, **369** [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [21]. Alipour J., Safari Lafti S., Askari Majdabadi H., Yazdiani A., Valinejadi A., *IIOAB Journal*. 2016:119 [[Google Scholar](#)], [[Publisher](#)]
- [22]. Soleiman-Beigi M., Arzehgar Z., *J. Ilam Uni. Med. Sci.*, 2013, **21**:1 [[Google scholar](#)], [[Publisher](#)]
- [23]. Arzehgar Z., Ahmadi H., *J. Chin. Chem. Soc.*, 2019, **66**:303 [[Crossref](#)], [[Google scholar](#)], [[Publisher](#)]
- [24]. Ryan L., Lam C., Mataraso S., Allen A., Green-Saxena A., Pellegrini E., Hoffman J., Barton C., McCoy A., Das R., *Ann. Med. Surg.*, 2020, **59**:207 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [25]. Zhao Z., Chen A., Hou W., Graham J.M., Li H., Richman P.S., Thode H.C., Singer A.J., Duong T.Q., *PloS one*, 2020, **15**:e0236618 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [26]. Allenbach Y., Saadoun D., Maalouf G., Vieira M., Hellio A., Boddaert J., Gros H., Salem J.E., Resche Rigon M., Menyssa C., *PloS one*, 2020, **15**:e0240711 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [27]. Pan P., Li Y., Xiao Y., Han B., Su L., Su M., Li Y., Zhang S., Jiang D., Chen X., Zhou F., Ma L., Bao P., Xie L., *J. Med. Internet Res.*, 2020, e23128 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [28]. Gao Y., Cai G.Y., Fang W., Li H.Y., Wang S.Y., Chen L., Yu Y., Liu D., Xu S., Cui P.F., *Nat. Commun.*, 2020, **11**:1 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [29]. Hernandez-Suarez D.F., Ranka S., Kim Y., Latib A., Wiley J., Lopez-Candales A., Pinto D.S., Gonzalez M.C., Ramakrishna H., Sanina C., *Cardiovasc. Revasc. Med.*, 2021, **22**:2228 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [30]. Parchure P., Joshi H., Dharmarajan K., Freeman R., Reich D.L., Mazumdar M., Timsina P., Kia A., *BMJ Support. Palliat. Care*, 2020 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [31]. Vaid A., Jaladanki S.K., Xu J., Teng S., Kumar A., Lee S., Somani S., Paranjpe I., De Freitas J.K., Wanyan T.F., *JMIR Med. Inform.*, 2021, **9**:e24207 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [32]. Yadaw A.S., Li Y.C., Bose S., Iyengar R., Bunyavanich S., Pandey G., *Lancet Digital Health*,

- 2020, **2**:e516 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)].
- [33]. Yan L., Zhang H.T., Goncalves J., Xiao Y., Wang M., Guo Y., Sun C., Tang X., Jing L., Zhang M.A., *Nat. Mach. Intell.*, 2020, **2**:283 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [34]. Pham Q.H., Chupradit S., Widjaja G., Alhassan M.S., Magizov R., Mustafa Y.F., Surendar A., Kassenov A., Arzehgar Z., Suksatan W., *Mater. Today Commun.*, 2021, **29**:102909 [[Crossref](#)], [[Google scholar](#)], [[Publisher](#)]
- [35]. Cai J., Luo J., Wang S., Yang S., *Neurocomputing*, 2018, **300**:70 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [36]. Khalid S., Khalil T., Nasreen S., *IEEE*; 2014, 372 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [37]. Sheikhpour R., Sarram M.A., Gharaghani S., Chahooki M.A.Z., *Pattern Recognit.*, 2017, **64**:141 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [38]. Tusar F.A., *Tusar, F.A., Hybrid Feature Selection Method for Health Data Mining*, Doctoral dissertation, Daffodil International University, 2019 [[Google Scholar](#)]
- [39]. Inbarani H.H., Azar A.T., Jothi G., Computer methods programs in biomedicine, 2014, **113**:175 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [40]. Verma L., Srivastava S., Negi P., *J. Med. Syst.*, 2016, **40**:1 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [41]. Kaur G., Chhabra A., *Int. J. Comput. Appl.*, 2014, 98. [[Crossref](#)], [[Google Scholar](#)]
- [42]. Gulia A., Vohra R., Rani P., *International Journal of Computer Science Information Technologies*, 2014, **5**:5110 [[Google Scholar](#)], [[Publisher](#)]
- [43]. Abdar M., Kalhori S.R.N., Sutikno T., Subroto I.M.I., Arji G., *International Journal of Electrical Computer Engineering*. 2015, **5**:1569 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)].
- [44]. Valecha H., Varma A., Khare I., Sachdeva A., Goyal M., 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), *IEEE*; 2018, 1 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [45]. Liu Y., Wang Y., Zhang J., New machine learning algorithm. International Conference on Information Computing and Applications, Springer; 2012, **7473**:246 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [46]. Dai B., Chen R.C., Zhu S.Z., Zhang W.W., International Symposium on Computer, Consumer and Control (IS3C), *IEEE*, 2018, 449 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [47]. Ghatasheh N., *Int. J. Adv. Sci. Technol.*, 2014, **72**:19 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [48]. Chen H., Lin Z., Mo L., Tan C., *Anal. Lett.*, 2017, **50**:2608 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [49]. McClendon L., Meghanathan N., *Mach. Learn.*, 2015, **2**:1 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [50]. Nahar N., Ara F., *International Journal of Data Mining Knowledge Management Process*. 2018, **8**:01 [[Google Scholar](#)],
- [51]. Jelodar H., Mirabedini S.J., Harounabadi A., 2015 Fifth International Conference on Communication Systems and Network Technologies, *IEEE*, 2015, 1025 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [52]. Manju B., Joshuva A., Sugumaran V., *Int. J. Mech. Eng. Technol.*, 2018, **9**:1061 [[Google Scholar](#)]
- [53]. Rodríguez-Gonzálvez P., Rodríguez-Martín M.J.I.A., *IEEE Access*, 2019, **7**:14714 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [54]. de Fátima Cobre A., Stremel D.P., Noleto G.R., Fachi M.M., Surek M., Wiens A., Tonin F.S., Pontarolo R., *Comput. Biol. Med.*, 2021, **134**:104531 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)].
- [55]. Leclerc T., Donat N., Donat A., Pasquier P., Libert N., Schaeffer E., D'Aranda E., Cotte J., Fontaine B., Perrigault P.F., Michel F., Muller L., Meaudre E., Veber B., *Anaesth. Crit. Care Pain Med.*, 2020, **39**:333 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [56]. Tambone V., Boudreau D., Ciccozzi M., Sanders K., Campanozzi L.L., Wathuta J., Violante L., Cauda R., Petrini C., Abbate A., Alloni R., Argemi J., Argemí Renom J., De Benedictis A., Galerneau F., García-Sánchez E., Ghilardi G., Hafler J.P., Linden M., Marcos A., Onetti Muda A., Pandolfi M., Pelaccia T., Picozzi M., Revello R.O., Ricci G., Rohrbach R., Rossi P., Sirignano A., Spagnolo A.G., Stammers T., Velázquez L., Agazzi E.,

- Mercurio M., *Public Health Front.*, 2020, **8**:284 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [57]. Burdick H., Lam C., Mataraso S., Siefkas A., Braden G., Dellinger R.P., McCoy A., Vincent J.L., Green-Saxena A., Barnes G., *Comput. Biol. Med.*, 2020, **124**:103949 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [58]. Chin V., Samia N.I., Marchant R., Rosen O., Ioannidis J.P., Tanner M.A., Cripps S., *European Journal of Epidemiology*. 2020, **35**:733 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [59]. Haug P.J., Gardner R.M., Tate K.E., Evans R.S., East T.D., Kuperman G., Pryor T.A., Huff S.M., Warner H.R., *Comput. Biomed. Res.*, 1994, **27**:396 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [60]. Ganguli R., *Springer*, 2021, 229 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [61]. El-Beheri H., Attia A.F., El-Fishawy N., Torkey H., *Comput. Biol. Chem.*, 2021, **93**:107536 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [62]. Lv H., Shi L., Berkenpas J.W., Dao F.Y., Zulfiqar H., Ding H., Zhang Y., Yang L., Cao R., *Brief. Bioinformatics*, 2021, **22**:bbab320 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]

### HOW TO CITE THIS ARTICLE

Raoof Nopour, Mehrnaz Mashoufi, Morteza Amraei, Nahid Mehrabi, Alireza Mohammadnia, Abdollah Mahdavi, Nader Mirani, Mojgan Saki, Mostafa Shanbehzadeh. Performance Analysis of Selected Decision Tree Algorithms for Predicting Drug Adverse Reaction Among COVID-19 Hospitalized Patients, *J. Med. Chem. Sci.*, 2022, 5(4) 505-517  
DOI: 10.26655/JMCHMSCI.2022.4.7  
URL: [http://www.jmchemsci.com/article\\_143572.html](http://www.jmchemsci.com/article_143572.html)