Scheduling the COVID-19 vaccine distribution based on data-driven decision-making methods

Mohammad Alipour-Vaezi[1], Reza Tavakkoli-Moghaddam[1, *], Mina Samieinasab[1]

Abstract

Since human societies have endured massive financial disruptions and life losses after the outbreak of the COVID-19 pandemic, it is critical to eliminate this disease as soon as possible. Today, the invention of the COVID-19 vaccine made this objective more reachable. But unfortunately, the suppliant of the vaccines is limited. Hence, to prevent further lethal harms, it seems rational to use a scientific method for vaccine allocation. This study proposes a method for prioritizing the patients based on their level of life-threatening danger according to the proven risk factors (e.g., age, sex, pregnancy, and underlying diseases) of the COVID-19. That is a new data-driven decision-making method for patients' classification based on their health condition information using several machine learning algorithms. In this method, vaccine applicants are classified into four classes. The scheduling of vaccine distribution would be conducted based on the results of this classification. Furthermore, a real-life case study is also investigated through the proposed method for better illumination in this paper. The vaccine distribution schedule of the real-case study has been performed with 94% accuracy. It should be mentioned that the main achievement of this research is to design a new efficient method for a vaccine distribution schedule.

Keywords: data-driven decision-making; scheduling; COVID-19; pandemic preparedness; classification.

1. Introduction

Seconds become undeniable vital in health care systems, where people's lives depend on time more than anywhere else. Here scheduling methods help the decision-makers to optimize their strategies. When a disruption occurs, two main things that should be done are managing it scientifically and learning from it for possible future disruptions. COVID-19 is highly infectious, but the severity of this disease is not the same for everyone. Although it infects people of all ages, evidence suggests that two groups of people are at higher risk of being infected by the disease. Bruine de Bruin (2021) showed that older adults are associated with perceiving more significant life-threatening risks if getting COVID-19. Also, those with

* Corresponding author; tavakoli@ut.ac.ir
[1] School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

underlying medical diseases (e.g., cardiovascular disease, diabetes, chronic respiratory disease, and cancer) are more vulnerable to this pandemic (Mehta et al., 2020, Muniyappa and Gubbi, 2020). Li et al. (2020) proposed comprehensive research on COVID-19 risk factors, which has to be considered, including heart disease, hyperglycemia, high-dose corticosteroid, hypertension, and high lactate dehydrogenase level. Due to the limited availability of COVID-19 vaccines in Iran, all provinces should develop allocation plans that outline a phased distribution. Therefore, people who are at a higher risk level will have faster access to the vaccines.

COVID-19 pandemic, as a disruption, touched the world in many aspects. Many researchers have proposed valuable studies for this pandemic preparedness or management. Some are mentioned in this section.

Güler and Geçici (2020) used a mixed-integer programming (MIP) model on a hospital physician's scheduling problem during COVID-19 disruption. In their survey, the hospital established a new department for treating COVID-19 patients. They embed their MIP model into a spreadsheet-based decision support system (DSS) that schedules both regular and disrupted (COVID-19) shifts. Judson et al. (2020) addressed the deployment of a self-triage and self-scheduling tool for COVID-19 patients. This tool checks the COVID-19 symptoms that were asked from the patients and categorizes the person into one of the four categories: emergent, urgent, non-urgent, and self-care. Patients in each category receive an appropriate level of healthcare services via direct scheduling or telephone hotline. Due to hospital limits in equipment and human force and reducing infection rate, Alipour-Vaezi and Tavakkoli-Moghaddam (2020) suggested using a location-allocation methodology based on multi-criteria decision-making (MCDM). The authors selected several hospitals in the city of Tehran, in which hospitals were chosen based on their qualification with the best-worst method (BWM) and the weighted aggregated sum product assessment (WASPAS) method.

Few studies referred to vaccine prioritization. Epstein et al. (2021) discussed that people with disabilities did not consider vaccine prioritization. To solve this problem, the Johns Hopkins Disability Health Research Center created the COVID-19 Vaccine Prioritization Dashboard for People with disabilities. Persad et al. (2021) studied the public perspective on COVID-19 vaccine prioritization. According to their study, most respondents agreed that health care workers should be vaccinated first, followed by medically vulnerable people and nursing home residents and staff. Some other papers have also studied the allocation of limited COVID-19 vaccine supply in different countries (Shim, 2021; Persad et al., 2020).

Data mining techniques have been used by many researchers in the COVID-19 pandemic preparedness and management field. Baralić et al. (2020) applied data mining techniques to assess the drug safety combination used for the COVID-19 treatment. With the help of the topological weighted centroid (TWC) algorithm, Buscema et al. (2020) tracked the COVID-19 data in Italy and gave insights on the geographical trend expansion of the epidemic. Sui et al. (2020) studied the changes in carbon emission patterns affected by the buses on the post-COVID-19 using data mining techniques. Harshavardhan et al. (2020) used generative adversarial networks (GAN) to diagnose COVID-19 patients based on medical images (e.g., X-Ray and CT Scan). Ren et al. (2020) used a data mining method to investigate the effectiveness of Chinese traditional medicines and treatments for COVID-19.

Due to the priceless studies conducted so far on the COVID-19 preparedness and management, today at the brink of the COVID-19 pandemic elimination, with the hopeful news of inventing the COVID-19 vaccine, an optimized model for vaccine distribution is necessary. The main features of this study are listed below:

- Proposing a data-driven decision-making method for scheduling the vaccine distribution. According to this method, people get classified and prioritized based on

their health information and the proven risk factors of the COVID-19. The vaccination scheduling process would take place based on this priority.

- Examining several data mining classifiers algorithms to select the most effective algorithm to deploy.
- Providing two real-life data sets. One for model fitting (train and test) and one as a case study to investigate the proposed model's application.
- Determining the benefit of this method over some other traditional scheduling methods.

The rest of this paper organizes as follows. In Section 2, the problem considered in this paper is solved. Section 3 is dedicated to the proposed method and describes used machine learning algorithms briefly. Section 4 gives models' experiments and selects the best model based on their accuracy, precision, recall, and F1-score. For a better understanding of the proposed method, Section 5 investigates a real-life case study. A comprehensive discussion is addressed in Section 6. Also, this section provides some valuable managerial insights. Final this paper reaches its end with an illuminating conclusion in Section 7.

2. Problem description

As mentioned before, the risk level of COVID-19 is not the same for everyone. For instance, it may become a more pernicious disease for people with underlying diseases. Also, other features (e.g., sexuality, age, and pregnancy) may cause higher or lower risk. There are several developing and testing candidates for the COVID-19 vaccine to immunize people against this pandemic virus. For the distribution of these vaccines, we need an optimized scheduling program that considers the risk factors of this pandemic and the hazard level of the disease on each person. In other words, to save more lives, it sounds reasonable to prioritize people based on their risk level for COVID-19.

Since there is a limited amount of vaccine, people, who want to receive it, should sign in their names and wait. Usually, the vaccine will be given to the ones who have enrolled earlier; however, the applicant's age is important in some cases. As the health condition of the applicant will not be considered most of the time, we find this method insufficient.

To address the mentioned problem, this paper proposes a data-driven decision-making method using various data mining techniques to classify and prioritize vaccine applicants. This method also benefits from experts' opinions in the preprocessing stage to enrich its accuracy. To stabilize the proposed methodology, this paper assumes that there is only one proofed vaccine ready for distribution. In other words, we exclude vaccine selection, which can be a good suggestion for future studies.

In this research, we assume that the vaccine applicants report their health condition data to the vaccine distribution center. After the data collection process, the decision-makers schedule the vaccination.

3. Methodology

3.1. Data collection

The data set used in this paper is collected from the "Sherkat Naft" Hospital in Tehran. Our study automated the manual COVID-19 cases' files of this hospital for the first time and presented the required information into a newly developed dataset. It has 1117 instances and four classes: low risk, medium risk, high risk, and very high risk. High-risk patients are those people, who are urgent for receiving the vaccine and very high-risk patients are emergent for vaccination, and for these people, COVID-19 can cause death. We remove 23 cases with missing values, and 1094 cases remained in the dataset. The developed dataset captures the

essential attributes of the virus's mortality rate based on expert opinions. More than 20 features are used to specify the main characteristics of patients selected from the manual records, and nine attributes of the overall attributes are extracted to be used in the prediction model based on combining some features and making new ones (see Table 1). We also use the One Hot Encoding to convert each underlying disease into a new column and assign a 1 or 0 value to the column.

Table 1. Data description

Attribute	Type
ID	Numerical
Age	Numerical
Pregnancy status	Boolean
Blood pressure	Boolean
Heart disease	Boolean
Kidney disease	Boolean
Diabetes	Boolean
Liver disease	Boolean
Lung disease	Boolean
Immune Deficiency disease	Boolean
Risk factor	categorical

3.2. Classification algorithms

Experts labeled the data set based on patients' priority in accessing the vaccine in our proposed work. In this paper, we used several machine-learning algorithms, including support vector machine (SVM), logistic regression, decision tree, bagging (bagged decision trees and extra trees), boosting (Gradient Boost, XgBoost, and CatBoost), and voting. In the voting algorithm, XgBoost, SVM, Gradient Boost, and extra trees classifier are used with the same weight. Then, the most effective algorithm is selected based on metrics described in Section 3.3. The concept of these methods is defined below.

3.2.1. Support Vector Machine

A support vector machine (SVM) is a supervised machine learning algorithm that uses in classification problems. It finds the best line to separate the possible outputs and can move data into a new space according to their predefined categories to be sorted and categorized linearly (or by hyperplane). Then, by finding the support vectors (or support planes in multi-dimensional space), the algorithm tries to find the linear equation that creates more distance between the two categories (Wang, 2005).

3.2.2. Bagging

The bagging approach is an ensemble method for improving weak learners by training the based classifiers in parallel. In bagging, a subset of the main data set is given to each of the classifiers. Each classifier views a portion of the data set and must build its model based on that portion of the data that is provided to it. The individual classifiers are combined with a majority vote in their decisions. The class selected by the largest number of classifiers is the ensemble's final decision (Schapire, 1990). Bagged decision trees, random forest, and extra trees are among the most common bagging algorithms that are also used in this study. Extra trees (extremely randomized trees) perform very well in accuracy and recall in most classification problems. It is a tree-based ensemble method that builds completely randomized trees and uses the whole learning sample to grow the trees (Geurts et al., 2006).

3.2.3. Boosting

Boosting is an ensemble model that generates multiple models for prediction. Schapire (1990) first used the boosting algorithm, proving that a weak classifier can become a strong classifier when placed in the probably approximately correct (PAC) format. In this method, the bias is reduced alongside the variance. This algorithm uses the entire data set to train each classifier, but it focuses more on the hard data to classify it correctly after each training session. This repetitive method is training with a greater focus on samples that have not previously been properly classified. Some of the most common boosting algorithms used in this study are Gradient Boosting, XGboost, and CatBoost. XGBoost and Gradient Boosting Machines (GBMs) are both ensemble tree-based methods that apply the principle of gradient descent. However, XGBoost has more parameters to be optimized and may enhance the prediction in this way (Friedman, 2001). CatBoost is a new gradient boosting approach with algorithmic advances in implementing ordered boosting and an innovative algorithm for processing categorical features (Prokhorenkova et al., 2017).

3.3. Metrics for Evaluation performance

In this section, metrics used to describe the performance of different ensemble algorithms with selected features are introduced. In all of these metrics, four parameters play a vital role, including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The TP and TN refer to correctly labeled positives and negatives by the algorithm, respectively. Also, the FP and FN are the counts of mislabeled samples.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Eq. (1) represents the Accuracy score, which is equal to the summation of true positives and true negatives divided into all true or false positives or negatives.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Eq. (2) calculates the Recall score that represents the model's ability to correctly predict the positives out of actual positives.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Eq. (3) is proposed to address the Precision score that represents the model's ability to correctly predict the positives out of all the positive predictions it made.

$$F1\ score = 2 * \frac{precision * Recall}{precision + Recall} \tag{4}$$

Eq. (4) provides the F1-score. This is a useful measure of the model in situations, in which both precision and recall are important.
It has to be mentioned that using an accuracy score as a defining metric for a model is necessary. However, it is always advisable to use Precision and Recall too. There might be other situations where the accuracy score is very high; however, the Precision or Recall score is low. Ideally, it is more favorable to avoid situations where the patient is at high risk for the

virus. Still, our model classifies it as not having the priority for the vaccine. So, recall is as critical as accuracy in our case.

The main advantage of using this method is to consider the COVID-19's risk factors. It leads the distribution schedule more reliable since it is based on the hazard level of vaccine applicants. The more accurate, we become in determining the applicants' hazard level, the more efficient the vaccine distribution schedule will be, and therefore, the fewer lives will be lost.

4. Experiments

The experiment is done using an Asus Corei7 Laptop with 2.50 GHz and 8 GB RAM to solve test problems and the open-source Python programming language, which is one of the most popular programming languages widely used in data mining and machine learning.

Table 2 lists the numerical results of applying classification algorithms. As mentioned before, in this study, both the Precision and the Recall scores are equally important. However, the recall of high-risk and very high-risk classes is more important because it is critical in this study to identify high and very high-risk patients to the virus more accurately. In this table, Precision, Recall, and F1-score are calculated based on their averages on all four classes. According to this table, both Voting and Bagged Decision Trees methods simultaneously perform better on accuracy than other ensemble methods. However, when it comes to the recall score for "high-risk" and "very high-risk" classes, Extra Trees performs better. As the Accuracy scores are close to each other, Extra Trees is chosen as the best prediction algorithm.

Table 3 shows the algorithm's Recall for each class. As discussed before, detecting 'high-risk' and 'very high-risk' patients to the COVID-19 virus is critical, and it should be paid special attention. The Extra Trees algorithm has a higher recall for these two classes alongside high accuracy.

Confusion Metrix for all algorithms used in this study is given in Figure 1. The Confusion Matrix (CM) is also determined to more clearly understand the performance of each algorithm. The performance is more accurate when the elements other than the diagonal elements are zero or near to negligible values in the CM. Classes 0 to 3 belong to low-risk, medium-risk, high-risk, and very high-risk patients.

Table 2. Performances evaluation metrics for classification algorithms

Classification algorithms	Precision	Recall	F1Score	Accuracy
Decision Tree	0.81	0.80	0.81	0.90
SVM	0.90	0.89	0.90	0.94
Random Forest	0.90	0.86	0.88	0.94
CatBoost	0.91	0.88	0.90	0.94
Logistic Regression	0.89	0.78	0.82	0.92
Bagged Decision Tree	0.90	0.90	0.90	0.95
Gradient Boosting	0.90	0.89	0.90	0.95
XgBoost	0.90	0.86	0.88	0.94
Voting Ensemble	0.90	0.90	0.90	0.95
Extra Trees	**0.83**	**0.91**	**0.86**	**0.94**

Table 3. Recall metric for different classes

Classification algorithms	Low-risk	Medium-risk	High-risk	Very high-risk
Decision Tree	0.60	0.94	0.94	0.75
SVM	0.76	0.98	0.94	0.87
Random Forest	0.76	0.98	0.96	0.75
CatBoost	0.70	0.98	0.96	0.87
Logistic Regression	0.52	1	0.92	0.69
Bagged Decision Tree	0.70	0.98	0.96	0.93
Gradient Boosting	0.70	0.98	0.94	0.94
XgBoost	0.70	0.98	0.96	0.81
Voting Ensemble	0.70	0.98	0.94	0.94
Extra Trees	**0.77**	**0.94**	**0.96**	**1**

According to Figure 1, it can be interpreted that Extra Trees can predict 50 patients from the high-risk group, correctly and just one high-risk patient predicted as very high-risk. The Extra Trees algorithm also predicts all of the 16 patients from very high risk correctly. Based on these results, this algorithm is very beneficial in this problem.

Comparison between different algorithms' accuracies both in the training and testing part is depicted in Figure 2. There is no significant difference in both the testing and training parts, so it is inferred that overfitting does not happen in this study.
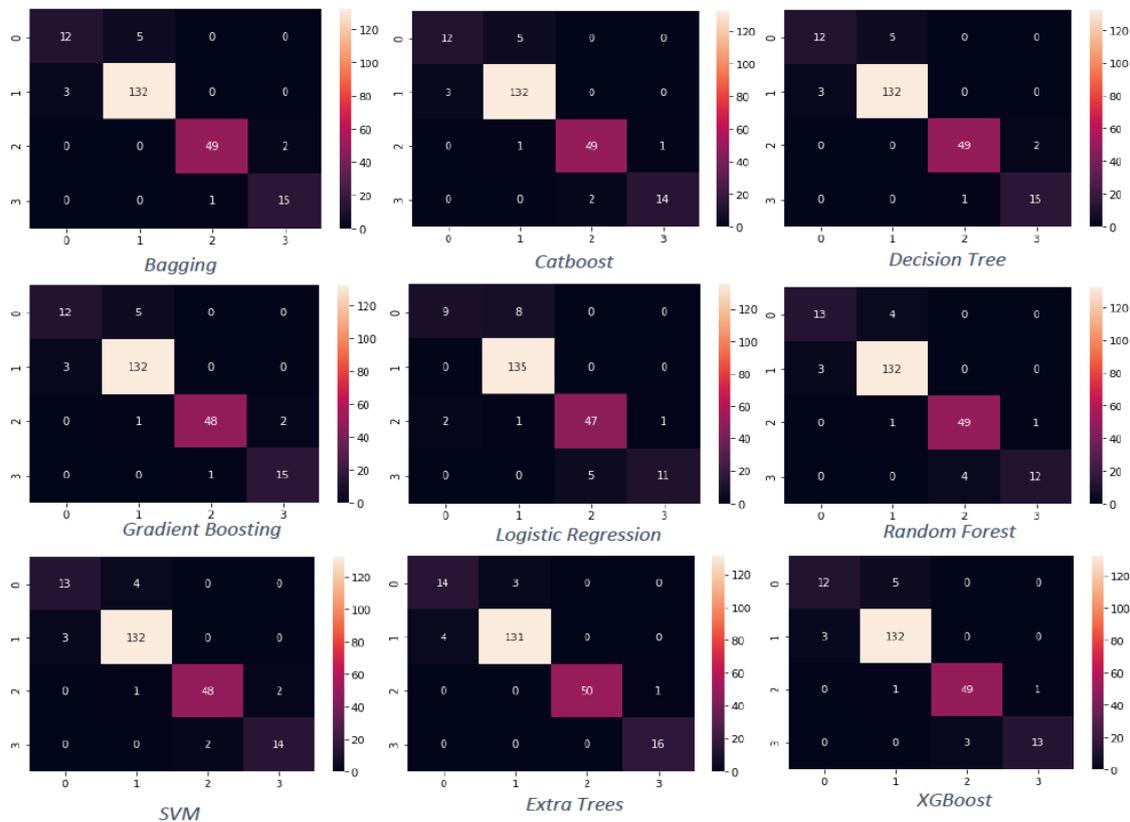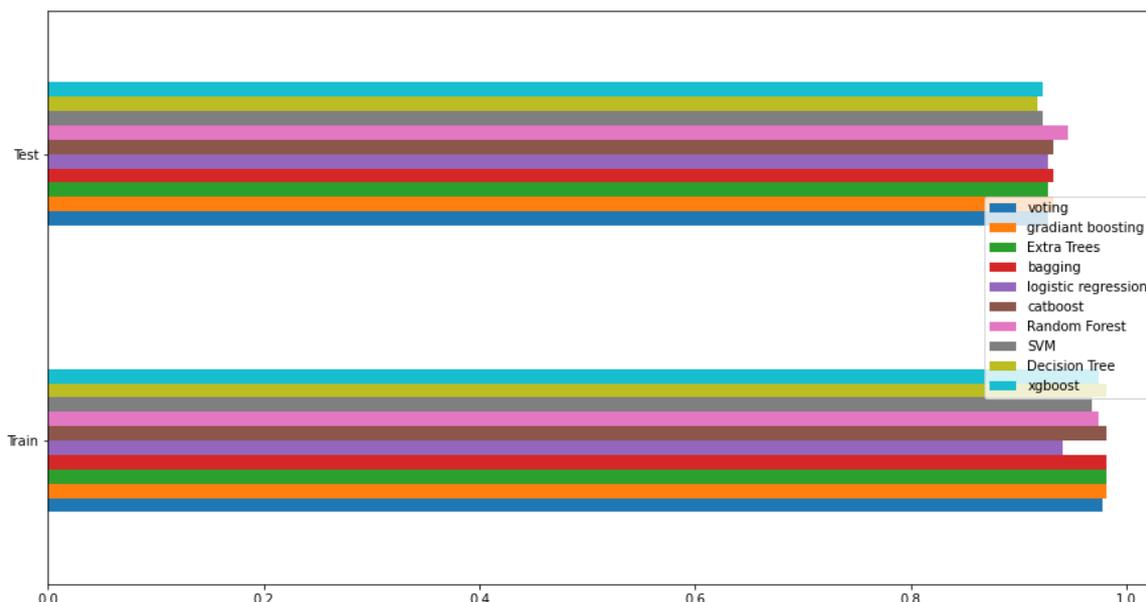

Figure 1. Confusion matrix

Article in Press

Figure 2. Overall comparison based on the accuracy

5. Case study

In Section 4, we have already ratified the performance of the proposed method. For better illumination of the proposed method, this section proposes a real-life case study, which represents a daily fraction of people, who are intended to receive the COVID-19 vaccine from Javad-al-Aemeh hospital in Bojnord in Iran. This case includes 27 patients. Medical information of these patients is extracted and relevant attributes to the risk factor of getting the severe coronavirus are chosen.

Numerical data are converted to binary attributes based on the expert's opinion. In this study, we assume that there are only 20 vaccines per hour in this hospital for injection, and these 27 patients have registered to receive the vaccine as their names are on the dataset.

In this case, we implement the Extra Trees classifier and labeled these 27 patients in four classes, including low-risk, medium-risk, high-risk, and very high-risk, which are represented by distinctive numbers from 0 to 3, respectively. The classification is based on patients' risk level to the virus. The results of the proposed method on this data set are given in figure 3. We have 12 high-risk and very high-risk patients, who should be vaccinated as soon as possible. In the next priority, moderate-risk patients will be vaccinated.

A vaccine supply is limited and logistical challenges in vaccine delivery persist, prioritizing vaccines based on the suggested method, offers a means of ensuring that vaccines reach those most at risk of getting a severe version of the COVID-19 virus.

Without this prediction framework, the decision-maker may use different dispatching rules for sequencing patients. For example, by using the first-come-first-served (FCFS) dispatching rule, patients 0 to 19 should first be vaccinated, and patients 20 to 26 should then be vaccinated due to the limitation. In this approach, the vaccination of four people in emergencies, who have a higher priority for vaccination will be deferred to the next time.

Article in Press

	label	Age	Pregnancy status	Blood pressure	Heart disease	Kidney disease	Diabetes	Liver disease	Lung disease	Immune Deficiency disease
0	1	34	1	0	0	0	0	0	0	0
1	1	64	0	0	0	0	0	0	0	0
2	1	44	0	0	0	0	0	0	0	0
3	2	29	0	0	0	0	0	0	0	1
4	1	58	0	0	0	0	0	0	0	0
5	2	56	0	0	0	0	1	0	0	0
6	2	48	0	0	0	0	0	1	0	1
7	0	23	0	0	0	0	0	0	0	0
8	2	26	0	0	0	1	0	0	0	0
9	2	21	1	0	1	0	1	0	0	0
10	3	48	0	0	1	1	1	0	0	0
11	0	26	0	0	0	0	0	0	0	0
12	3	58	0	0	1	0	1	0	0	1
13	1	45	0	0	0	0	0	0	0	0
14	2	87	0	0	1	1	0	0	0	0
15	1	59	0	0	0	0	0	0	0	0
16	1	69	0	0	0	0	0	0	0	0
17	0	30	0	0	0	0	0	0	0	0
18	1	73	0	0	0	0	0	0	0	0
19	1	54	0	0	0	0	0	0	0	0
20	1	81	0	0	0	0	0	0	0	0
21	1	66	0	0	1	0	0	0	0	0
22	1	73	0	0	0	0	0	1	0	0
23	3	50	0	0	0	1	0	1	0	0
24	2	71	0	0	1	1	0	0	0	0
25	2	33	1	0	0	0	0	0	1	0
26	2	52	0	0	0	0	0	0	1	0

Figure 3. Result information of the case study

6. Discussion and managerial insights

An effective prioritization scheme to allocate the vaccines to the most needed population will be critical if the vaccination program is to succeed. Therefore, quick and accurate planning for vaccine injection in health centers is vital. Based on our findings, the Extra Trees classifier could be a practical ensemble algorithm for classification accuracy and recall. Also, this algorithm is much faster than other Tree-based algorithms because of selecting the split points randomly. So, it can be beneficial for large medical datasets. It can be inferred that applying this data-driven method can save the lives of people who are at risk of severe coronavirus and death by properly and quickly prioritizing people. In this method, upon arrival of a new entry into the system, one can decide on their priority and vaccinate them as soon as possible if they have a high priority. This method is a dynamic approach that is superior to many sequencing algorithms in this point of view.

According to the above discussion, here several points are mentioned as managerial insights. First, the comprehensive patient records information in different hospitals should be aggregated in an integrated system. Then, patients should be prioritized based on the method presented in this study. The number of hospitals with vaccines and nurses in each hospital should be calculated. Finally, patients will be allocated to these centers based on the calculated priority.

7. Conclusion

Human societies are on the brink of the elimination of the COVID-19 pandemic benefitting from several vaccines. Since the supplement of these vaccines are limited and there is a massive influx of vaccine applicants, it seems more reasonable to allocate available vaccines to people

with greater mortality risk. This paper proposed a new methodology that prioritized vaccine applicants based on their health conditions. Several machine learning algorithms have been applied to determine the risk level of vaccine applicants based on the proven risk factors of the COVID-19. Applied machine learning algorithms include decision tree, support vector machine, random forest, CatBoost, logistic regression, Extra trees, gradient boosting, XgBoost, voting ensemble, and bagged decision tree. The most efficient algorithm was selected based on four evaluating metrics: Accuracy, Precision, Recall, and F1-score. A real-life case study was proposed, which was investigated by the proposed method for better understanding.

There were some limitations in this study that can be the subject of future studies. The proposed model can be developed with more attributes, such as taking drugs. Moreover, some attributes (like blood pressure) are Boolean in this data set; however, it can be numerical for a more accurate prediction. As can be seen from experiments in Section 4, all of the considered algorithms are seen to generate acceptable classification performance rates. So, all algorithms may prove valuable and appropriate approaches for dealing with the prediction problem in prioritizing the patients for receiving the vaccine. However, we select the Extra Trees classifier, which provides 93% accuracy and 91% recall. If the prediction accuracy is low, we suggest using parameter tuning methods, such as Bayesian Optimization or meta-heuristic algorithm including grasshopper optimization algorithm (GOA). In a case with a high number of attributes, feature selection will be crucial to choose the most relevant attributes among all.

Acknowledgments

References

Alipour-Vaezi, M., and Tavakkoli-Moghaddam, R., (2020). "A New Methodology For Covid-19 Preparedness Centers Based On A Location-Allocation Platform", *Journal of Industrial and Systems Engineering,* Vol. 13, pp. 35-41.

Baralić, K., Jorgovanović, D., Živančević, K., Miljaković, E. A., Antonijević, B., Djordjevic, A. B., Ćurčić, M., and Đukić-Ćosić, D., (2020). "Safety assessment of drug combinations used in covid-19 treatment: in silico toxicogenomic data-mining approach", *Toxicology and Applied Pharmacology,* Vol. 406, pp. 115237.

Bruine De Bruin, W., (2021). "Age differences in covid-19 risk perceptions and mental health: evidence from a national us survey conducted in March 2020", *The Journals of Gerontology: Series B,* Vol. 76**,** E24-E29.

Buscema, P.M., Della Torre, F., Breda, M., Massini, G., and Grossi, E., (2020). "Covid-19 In Italy and extreme data mining", *Physica A: Statistical Mechanics and Its Applications,* Vol. 557, pp. 124991.

Epstein, S., Ayers, K., and Swenor, B.K., (2021). "Covid-19 vaccine prioritisation for people with disabilities", *The Lancet Public Health,* Vol. 6**,** E361.

Friedman, J.H., (2001). "Greedy function approximation: A gradient boosting machine. *Annals of Statistics***,** pp. 1189-1232.

Geurts, P., Ernst, D., and Wehenkel, L., (2006). "Extremely randomized trees", *Machine Learning,* Vol. 63, pp. 3-42.

Güler, M.G., and Geçici, E., (2020). "A decision support system for scheduling the shifts of physicians during Covid-19 pandemic"**,** *Computers & Industrial Engineering,* Vol. 150**,** 106874.

Harshavardhan, A., Bhukya, H., and Prasad, A.K., (2020). "Advanced machine learning-based analytics on Covid-19 data using generative adversarial networks", *Materials Today: Proceedings*.

Judson, T.J., Odisho, A.Y., Neinstein, A.B., Chao, J., Williams, A., Miller, C., Moriarty, T., Gleason, N., Intinarelli, G., and Gonzales, R., (2020). "Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for Covid-19", *Journal of The American Medical Informatics Association,* Vol. 27, pp. 860-866.

Li, X., Xu, S., Yu, M., Wang, K., Tao, Y., Zhou, Y., Shi, J., Zhou, M., Wu, B., and Yang, Z., (2020). "Risk factors for severity and mortality in adult covid-19 inpatients in Wuhan", *Journal of Allergy and Clinical Immunology,* Vol. 146, pp. 110-118.

Mehta, V., Goel, S., Kabarriti, R., Cole, D., Goldfinger, M., Acuna-Villaorduna, A., Pradhan, K., Thota, R., Reissman, S., and Sparano, J.A., (2020). "Case fatality rate of cancer patients with covid-19 in a new york hospital system", *Cancer Discovery,* Vol. 10, pp. 935-941.

Muniyappa, R., and Gubbi, S., (2020). "Covid-19 pandemic, coronaviruses, and diabetes mellitus", *American Journal of Physiology-Endocr`inology And Metabolism,* Vol. 318**,** E736-E741.

Persad, G., Emanuel, E.J., Sangenito, S., Glickman, A., Phillips, S., and Largent, E.A., (2021). "Public perspectives on covid-19 vaccine prioritization", *Jama Network Open,* Vol. 4, E217943-E217943.

Persad, G., Peek, M.E., and Emanuel, E.J., (2020). "Fairly prioritizing groups for access to covid-19 vaccines", *Jama,* Vol. 324, pp. 1601-1602.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., and Gulin, A., (2017). "Catboost: unbiased boosting with categorical features", *Arxiv Preprint Arxiv:1706.09516*.

Ren, X., Shao, X.-X., Li, X.-X., Jia, X.-H., Song, T., Zhou, W.-Y., Wang, P., Li, Y., Wang, X.-L., and Cui, Q.-H., (2020). "Identifying potential treatments of covid-19 from traditional chinese medicine (tcm) by using a data-driven approach", *Journal of Ethnopharmacology*, 112932.

Schapire, R.E., (1990). "The strength of weak learnability", *Machine Learning,* Vol. 5, pp. 197-227.

Shim, E., (2021). "Optimal allocation of the limited covid-19 vaccine supply in South Korea", *Journal of Clinical Medicine,* Vol. 10, pp. 591.

Sui, Y., Zhang, H., Shang, W., Sun, R., Wang, C., Ji, J., Song, X., and Shao, F., (2020). "Mining urban sustainable performance: spatio-temporal emission potential changes of urban transit buses in post-covid-19 future", *Applied Energy,* Vol. 280, 115966.

Wang, L., (2005). *Support Vector Machines: Theory And Applications*, Springer Science & Business Media.