

# An explanatory machine learning framework for studying pandemics: The case of COVID-19 emergency department readmissions

Behrooz Davazdahemami<sup>a,\*</sup>, Hamed M. Zolbanin<sup>b</sup>, Dursun Delen<sup>c,d</sup>

<sup>a</sup> Department of IT & Supply Chain Management, University of Wisconsin-Whitewater, United States

<sup>b</sup> Department of MIS, Operations & Supply Chain Management, Business Analytics, University of Dayton, United States

<sup>c</sup> Center for Health Systems Innovation, Spears School of Business, Oklahoma State University, United States

<sup>d</sup> School of Business, Ibn Haldun University, Istanbul, Turkey

## ARTICLE INFO

### Keywords:

Machine learning  
Pandemic  
COVID-19  
SHAP  
Deep learning  
Genetic algorithm

## ABSTRACT

One of the major challenges that confront medical experts during a pandemic is the time required to identify and validate the risk factors of the novel disease and to develop an effective treatment protocol. Traditionally, this process involves numerous clinical trials that may take up to several years, during which strict preventive measures must be in place to control the outbreak and reduce the deaths. Advanced data analytics techniques, however, can be leveraged to guide and speed up this process. In this study, we combine evolutionary search algorithms, deep learning, and advanced model interpretation methods to develop a holistic exploratory-predictive-explanatory machine learning framework that can assist clinical decision-makers in reacting to the challenges of a pandemic in a timely manner. The proposed framework is showcased in studying emergency department (ED) readmissions of COVID-19 patients using ED visits from a real-world electronic health records database. After an exploratory feature selection phase using genetic algorithm, we develop and train a deep artificial neural network to predict early (i.e., 7-day) readmissions (AUC = 0.883). Lastly, a SHAP model is formulated to estimate additive Shapley values (i.e., importance scores) of the features and to interpret the magnitude and direction of their effects. The findings are mostly in line with those reported by lengthy and expensive clinical trial studies.

## 1. Introduction

Hospital readmission bears significance in both institutional quality and cost of care [1]. Prediction of rehospitalization enables early intervention, which is crucial to preventing more serious or life-threatening events [2]. Readmissions also comprise a significant portion of total medical expenses [3]. Recent studies show that as much as 27% of hospital readmissions are potentially preventable [4], costing Medicare alone \$26 billion annually [5]. In recent years, and especially after the initiative by the U.S. Centers for Medicare and Medicaid Services (CMS) that penalizes hospitals for avoidable readmissions, readmission rates have increasingly been used as a quality benchmark for assessment of hospitals and health systems in certain contexts [6]. According to a former director of CMS' quality measurement and health assessment group, early rehospitalizations are mainly driven by "defects

in care, medication errors, failure to plan for necessary equipment, and shortcomings in the preparation of the patient and family for his or her care outside of the hospital" [7].

While cost and quality of care have been the main drivers of interest among scholars in hospital readmissions, the recent global COVID-19 pandemic has revealed a lesser-known problem of healthcare that can be aggravated by high hospital readmission rates: The unusual decision many doctors and healthcare providers had to make to prioritize intensive care unit (ICU) facilities for patients with higher chances of survival [8–10]. A similar situation holds in the emergency department (ED), where physicians have to prioritize patients with higher risks of developing complications to be hospitalized and discharge low- to moderate-risk patients. This poses another perspective to the study of hospital and ED readmissions: To determine the principal factors that lead to readmissions among patients of novel viruses (including SARS-

\* Corresponding author at: Department of IT & Supply Chain Management, University of Wisconsin-Whitewater, 809 W Starin Rd., Hyland Hall 1222, Whitewater, WI 53190, United States.

E-mail addresses: [davazdab@uww.edu](mailto:davazdab@uww.edu) (B. Davazdahemami), [hmzolbanin@udayton.edu](mailto:hmzolbanin@udayton.edu) (H.M. Zolbanin), [dursun.delen@okstate.edu](mailto:dursun.delen@okstate.edu), [dursun.delen@ihu.edu.tr](mailto:dursun.delen@ihu.edu.tr) (D. Delen).

<https://doi.org/10.1016/j.dss.2022.113730>

Received 27 December 2020; Received in revised form 21 August 2021; Accepted 10 January 2022

Available online 18 January 2022

0167-9236/© 2022 Elsevier B.V. All rights reserved.

CoV-2) to minimize the ethical and professional burden on hospitals and doctors. Such an endeavor is crucial for two reasons. First, when a pandemic is caused by a novel pathogen, our understanding of the disease's primary readmission risk factors is minimal or anecdotal. Second, even though we learn more about these risk factors as the number of affected citizens increases, we need to establish a reliable framework to identify the risk factors as early as possible such that the adverse effects of hospital readmissions can be minimized in an ongoing or a future pandemic. This study seeks to tackle these concomitant problems. Specifically, we use artificial intelligence (AI) and data analytics to implement a clinical decision support system (DSS) to predict readmission and to discover the prominent factors among COVID-19 patients.

Our study aims to make two main contributions. From the methodological perspective, we use evolutionary algorithms, deep artificial neural networks (DNN), and eXplanatory Artificial Intelligence (XAI) to develop a hybrid data analytics (i.e., an exploratory/predictive/explanatory) framework. This framework can be used to accelerate the development of clinical DSS, which in turn facilitates the process of understanding novel diseases. To evaluate the proposed framework from a practical perspective, we identify the risk of readmission in early COVID-19 patients using their electronic health records. Particularly, we aim at providing the practitioners with a tool to gain insight about novel diseases within a reasonable time frame (as opposed to running lengthy clinical trials) and to enable them to adjust their treatment protocols accordingly. While a large number of studies have been published on the COVID-19 pandemic, we argue that a majority of those studies have mainly focused on only one of the aspects incorporated in our proposed framework (i.e., either exploration, prediction, or explanation). Specifically, explanatory studies on COVID-19 have mostly relied on lengthy clinical trials that used limited samples, and therefore, led to hardly generalizable results. The Shapley Additive eXplanations (SHAP) approach employed in our proposed framework, however, not only provides more generalizable insights at the aggregate (population) level, but also enables the practitioners to analyze the specific risk factors that are associated with each individual subject.

We organize the remainder of the manuscript into four sections. In the following section, we review the recent work on the use of data analytics for the analysis and identification of risk factors that lead to ED or hospital readmissions. Next, we describe our data processing and data engineering steps. Subsequently, we propose our framework and evaluate it by building a clinical DSS to predict hospital readmissions in COVID-19 patients where the related data is obtained from Cerner's HealthFacts data warehouse. Finally, we discuss our findings and conclude the manuscript with a summary of our contributions and avenues for future research.

## 2. Prior work

Two approaches have generally been used to predict hospital readmissions. In real-time models, predictors are available on or shortly after the index hospitalization, whereas in retrospective studies, features used to predict rehospitalization are mostly known after the patient is discharged [11]. Since most electronic medical records (EMR) aggregate data on various aspects of hospital encounters, including variables on administration, demographics, visit, laboratory, medication, and discharge processes, models built using such data are considered to belong to the retrospective category [12].

Various retrospective studies in the literature have tried to predict hospital readmissions; however, a majority of them do not divide the data into training and validation sets for an objective evaluation and/or do not evaluate the prediction models using appropriate assessment metrics. As a result, these studies have a descriptive perspective rather than a predictive focus [13]. Logistic regression and survival analysis have been the most common modeling methods in these descriptive studies [14]. A smaller number of studies have adopted the predictive approach and used machine learning techniques for the prediction of

hospital readmissions.

While predictive analytics approaches have been widely used for studying hospital readmissions, ED readmission studies have primarily focused on determining explanatory variables instead of building predictive models. Nevertheless, it would be especially important for the emergency clinicians to identify particularly high-risk patients who may not only come back in a more critical state, but also may die even before returning to the hospital. Obermeyer et al. [15] examined Medicare claims from 2007 to 2012 and found that among patients discharged from the ED, 0.12% died within seven days, representing 10,093 cases per year nationally. Gunnarsdottir and Rafnsson [16], who found a mortality rate of 208.5 per 100,000 ED visits within eight days after discharge, attribute this rate to a misjudgment of patients' conditions at the time of discharge. Another study asserts that 3-day ED readmissions account for around 30% of unexpected intensive care unit (ICU) admissions [17].

Table 1 provides a summary of the relevant predictive studies on ED and hospital readmissions.

A review of the studies listed in Table 1 reveals three major gaps. First, all of these studies focus on known, chronic conditions, and therefore, their main focus is on improving the performance of the predictive models through a combination of novel modeling techniques and more effective data engineering processes. COVID-19, however, is an acute condition for which the identification of the risk factors and their significance is as equally important as the prediction accuracy. In other words, the descriptive and prescriptive components of the analytical models are as critical as their predictive angle. Second, most of the extant work in the literature focuses on 30-day readmission of patients; however, early readmissions (i.e., 7-day) are better indicators of quality of care [18] and are more likely to be preventable and amenable to hospital-based interventions [19]. Prediction of early readmissions is especially important for hospitals dealing with COVID-19 patients because it can inform clinical practice, discharge disposition decisions, and health care planning to ensure the availability of resources needed for acute and follow-up care of discharged patients [20]. Third, ED readmissions are relatively understudied from a predictive analytics point of view. Since prior research argues that a considerable fraction of ED discharges each year leads to death or early readmission of critically ill patients, COVID-19 patients are expected to be at a higher risk due to the novelty and less-known nature of the disease. Our goal in this research effort is to address these three gaps for COVID-19 patients simultaneously.

Our attempt is guided by the existing descriptive studies on COVID-19 rehospitalizations. Recent studies on hospital readmissions that are related to COVID-19 have identified various risk factors. While these studies differ on sample characteristics - such as size and geographic location - they point out certain factors as common risk factors for hospital readmission among COVID-19 patients. For instance, certain underlying diseases such as diabetes, COPD, diseases of the genitourinary system (e.g., chronic kidney disease), hypertension, cancer, and liver diseases, as well as older age (typically higher than 65) are found to be more common among those who were re-hospitalized due to post-discharge COVID-19 complications. Table 2 provides a summary of these studies. We draw on these findings to guide our data preprocessing and feature selection in the following section.

## 3. Methods and materials

### 3.1. Data

Emergency department (ED) visits recorded in the "Cerner Health Facts" data warehouse between November 30, 2019 (around the date when the first signs of the novel virus were noticed in China), and June 9, 2020, were used. We excluded visits by non-adult (< 18 years of age) patients. The resulting raw data set contained 27,215 visits made by 22,963 unique patients. We further refined our data set by only keeping

**Table 1**  
Predictive studies on hospital and ED readmissions.

Study	Readmission facility	Disease(s) <sup>a</sup>	Sample size (patients)	Period	Best performing method	Metric	Value
[68]	Hospital	COPD	106	30 days	Random Forest	AUC	0.720
[69]	Hospital	Multiple	92,530	30 days	LASSO & SVM	AUC	0.680
[70]	Hospital	PN	40,442	30 days	Deep Neural Network (DNN)	AUC	0.734
		COPD	31,457				0.711
		CHF	25,941				0.676
		AMI	29,060				0.649
		THA/TKA	23,128				0.638
[71]	Hospital	HF	4210	30 days	SVM	AUC	0.660
		AMI	2379				0.650
		PN	2825				0.630
[72]	Hospital	CHF	1641	30 days	PSO-SVM	Accuracy	0.784
						Sensitivity	0.973
						Specificity	0.086
[73]	Hospital	Multiple	64,912	3 days	Ensemble	AUC	0.666
				7 days			0.681
				15 days			0.700
				30 days			0.720
[74]	Hospital	CHF	4840	30 days	CHAID Decision Tree	AUC	0.707
[75]	Hospital	Diabetes	Not Given	30 days	Recurrent NN	AUC	0.800
[76]	Hospital	Multiple	32,718	30 days	DNN	AUC	0.780
[77]	Hospital	Multiple	304,888	30 days	Ensemble	AUC	0.771
[78]	Hospital	Lupus	9457	30 days	DNN	AUC	0.700
[2]	Hospital	Multiple	700	30 days	DNN	AUC	0.730
[12]	Hospital	CHF	32,350	30 days	Random Forest	AUC	0.742
		COPD	31,070				0.754
[79]	Hospital	COPD	111,992	30 days	Gradient Boosting Tree	AUC	0.653
[80]	Hospital	Multiple	38,597	30 days	DNN	AUC	0.714
[81]	ED	Multiple	279,611	30 days	Decision Tree	Accuracy	0.772
						Sensitivity	0.402
						AUC	0.732
[82]	ED	Multiple	330,631	3 days	Gradient Boosting	Sensitivity	0.16
						Precision	0.75
						AUC	0.76
[83]	ED	Multiple	120,000	3 days	DNN	Accuracy	0.680
						Sensitivity	0.679
						AUC	0.755
[48]	ED	Multiple	290,000	3 days	Ensemble	Accuracy	0.957
						AUC	0.610

\*AUC stands for Area Under the (Receiver Operating Characteristic) Curve.

<sup>a</sup> COPD: Chronic Obstructive Pulmonary Disease; PN: Pneumonia; CHF: Congestive Heart Failure; AMI: Acute Myocardial Infarction; THA/TKA: Total Hip/Knee Arthroplasty.

**Table 2**  
Risk factors for readmission among COVID-19 patients.

Study	Location	Sample size	Readmission rate	Risk factors
[84]	USA	279	6.8%	Comorbidity (hypertension, diabetes, COPD, liver disease, cancer, substance abuse)
[85]	South Korea	7590	4.5%	Gender (men), age, medical aid subscription, comorbidity, chest radiographs, computed tomography (CT) scans, HIV antivirals
[20]	USA	106,543	9%	Discharge disposition, age, comorbidity (COPD, CHF, diabetes, chronic kidney disease, obesity)
[86]	Spain	1368	4.4%	Weakened immune system, having fever within 48 h prior to discharge
[87]	Turkey	154	7.1%	Malignant tumor, Hypertension
[88]	USA	1775	19.9%	Age

\*AUC stands for Area Under the (Receiver Operating Characteristic) Curve.

those patients who tested positive for COVID-19 in at least one of their visits to the ED. For this group of patients, we retained the first visit with a positive test, as well as all their following visits (i.e., 6620 unique patients who made 7373 ED visits). We excluded other patients and all

their visits from the data set.

In the next step, we performed one-hot encoding for every comorbidity/symptom diagnosed in at least 50 patients during their visits (based on the ICD-10 coding system). This resulted in 729 binary variables where each variable represented the existence of one of the comorbid conditions in a patient. Similarly, 1487 unique medications (each being taken by at least 50 patients) were identified and one-hot encoded. Additionally, patients' demographic and visit-specific information (i.e., age, gender, race, admission type, and payer type) were included for further analyses.

Finally, we derived a binary response variable indicating whether each visit was followed by another one (by the same patient) within a period of 7 days (i.e., response = 1) or not (i.e., response = 0). Our records showed that 1358 out of 7373 (18.4%) COVID-19 related visits were followed by another visit within 7 days of discharge. This rate is far away from the 3–4% average rate for 7-day readmission to the emergency department reported in the literature [21,22] but is partly anticipated due to the novelty and mysterious nature of COVID-19 (at the time of visits). Table 3 shows the demographics of the final data set.

### 3.2. Methodology

#### 3.2.1. Feature selection

Since very little is known about the mechanism and confounders of the novel coronavirus disease, we chose to employ an exploratory approach to select comorbidity/symptom and medication features that

**Table 3**  
Summary statistics of the patients.

Variable	Average (std dev) / proportion
Age	45.81 (17.03)
Race	White 39.8%
	Black 28.3%
	Latino 19.7%
	Other 12.2%
Gender	Male 53.1%
	Female 46.7%
	Other 0.02%

affected the likelihood of post-COVID 7-day ED readmissions. Even though we employed a fully exploratory approach and did not make the algorithm to include any specific feature in the final feature set, we made sure that the relevant risk factors already identified in the literature (see Table 2) are present in the initial feature set (even if they did not meet the minimum frequency threshold of 50 instances).

A forward feature selection genetic algorithm (GA) was developed to refine the feature set. GA is a heuristic, evolutionary-based random-search approach that mimics the process of natural selection. It begins by the random generation of a large number (i.e., population) of feasible solutions (i.e., chromosomes). Each solution will then be scored using a fitness function that quantifies the desirability of the solution. The fittest solutions from each generation will then mate (a.k.a., crossover) to produce offspring of the next generation. The process of reproducing new generations continues until the algorithm converges in terms of the fitness of the top solutions.

We developed a GA with a population size of 1000 solutions in each iteration (i.e., generation). Each solution in the initial population involved a random set including 200 to 220 features from the 2221 features contained in the initial data set (i.e., around 10%). In each iteration, using the selected set of features for that iteration, we trained a basic Random Forest (RF) model<sup>1</sup> with 80% of the data and tested the model on the remaining 20%. The Area Under the Receiver Operating Characteristic curve (AUC) of the trained RF model was used as the fitness function to identify the top feature set in each generation. We chose AUC as the fitness function since this measure indicates the distinctive power of each feature set in distinguishing visits that lead to readmissions from those that do not. A tournament selection strategy was defined for the algorithm such that at the end of each iteration, it keeps 30% top solutions, discards 30% worst solutions, and generates new solutions by performing crossover among the top 30% (with a crossover rate of 20%). Also, a 5% mutation rate was considered to prevent the algorithm from getting stuck in a limited area of the solution space. A maximum of 100 generations was set for the algorithm, with the possibility of an early stop in case of no AUC improvement in 10 consecutive generations.

In the end, the selected set of features associated with the highest AUC was retained to be used for further training and eventual testing of the predictive models.

### 3.2.2. Predictive modeling

A fully-connected multilayer perceptron (MLP) deep artificial neural network - a representation learning method [23] - was employed for building the predictive model. In addition to finding the mapping from features to the output, which is done by classic machine learning (ML) methods, representation learning techniques operate by learning and

<sup>1</sup> The reason we chose RF for feature selection was the existence of a large number of binary features in the data, which makes it an excellent candidate for tree-based algorithms as compared to other ML techniques. We also tried other tree-based models such as gradient boosted trees and simple decision trees; however, RF yielded the best prediction results with the whole set of features in the model validation process.

discovering the data features as well [24]. In other words, a deep network, through multiple hidden layers incorporated in its architecture, first derives complex features from simple concepts and then maps those advanced features to the output.

Regarding the network elements (i.e., weight functions, transfer functions, etc.), MLP networks (a.k.a., deep feedforward networks) are the most similar type of deep networks to the typical neural network models used in classic ML [24]. They are called feedforward because no feedback connections (i.e., feeding outputs of a perceptron back to it as input) are allowed for the signals in their architecture [23,25]. A fully-connected (a.k.a., dense) MLP is one in which every unit (perceptron) from a layer is connected to all units from the preceding and succeeding layers. While MLP networks, technically, have no limitation in terms of the number of input features, incorporating more features requires providing the network with a large number of instances for decent training. Otherwise, the model may either not converge at all or overfit the training data quickly and perform poorly in classifying new, unseen cases. That is why we began with an exploratory feature selection approach to reduce the dimensionality by around 90%, yet trying to retain the most distinctive subset of features.

An MLP network with four hidden dense layers was developed to train the predictive model. The model (Fig. 1) involves a total of 89,793 trainable parameters (i.e., connection weights).

For each hidden dense layer, we used Rectified Linear Units (ReLU) as the activation (transfer) function. ReLU is shown to be highly efficient with deep learning (DL) applications since, unlike other common functions (e.g., linear or hyperbolic tangent), it does not suffer from the *vanishing gradient problem* [26]. This problem occurs in the gradient descent optimization process when the derivation of the transfer function for a given input becomes very small through the backpropagation of error. As a result, the weights of such inputs will not be updated during the training process. The literature suggests that He’s initialization schema [27] is the best parameter initialization method for hidden layers that involve the ReLU activation function [28] as it maintains the variance of weights high and compensates for the amount of variance decreased as a result of applying ReLU. Additionally, given the binary nature of the response variable, a sigmoid transfer function was used for the output layer.

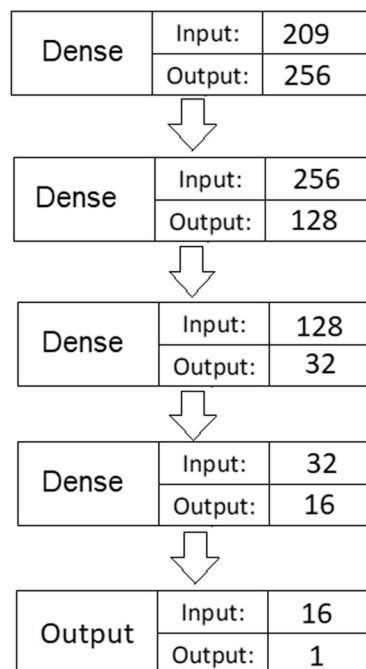


Fig. 1. Deep fully-connected MLP network architecture.

Regarding the relatively large number of parameters compared to the available instances for training the model, there were fairly high odds of overfitting the model to the training data and poor generalizability. To address that issue, we used a combination of three strategies suggested in the literature: learning rate decay, L1 regularization, and L2 regularization.

The learning rate decay strategy simply reduces the learning rate (i.e., the proportion of error that backpropagates through the network to tune up weights) after each epoch. This leads to faster convergence of the network weights at the beginning and then gradually slows down their changes as the model output becomes closer to the real value of the response variable. Otherwise, the weight optimization algorithm would quickly use all noises to tune up the weights and would create a model that would perfectly predict the outcome for all instances in the training data (i.e., overfitting). Various methods have been proposed for scheduling the decay, such as the exponential schedule [29,30], step-based schedule [31], and non-monotonic schedule [32] to obtain the best training performance in complex DL applications, such as image analysis or voice recognition. However, since our MLP network has a relatively simple architecture (compared with convolutional or recurrent neural networks, which are used for more complex tasks), we employed a simple linear decay schedule with a fixed decay rate per each epoch to avoid any additional parameterization.

L1 and L2 regularization strategies each add a penalty term to the loss function of the weight optimization algorithm. Specifically, L1 regularization adds a proportion of the *sum of absolute network weights* to the loss function, while L2 regularization adds a proportion<sup>2</sup> of the *sum of squared network weights* to that function. Generally speaking, they both prevent the algorithm from assigning large weights to any feature during the learning process, as large weights make the network unstable and result in large differences in the output by small changes in the input features [33]. This is particularly important when the trained model sees new data with somewhat different values for the features weighted heavily, leading to poor performance in predicting their outcomes. More technically speaking, L1 regularization (a.k.a., Lasso regression) encourages sparsity by making the weights to be zero when possible. Hence, it can be thought of as a feature selection approach that removes less important features by encouraging their weights (i.e., contribution) in calculating the outcome to be zero. L2 regularization (a.k.a., Ridge regression), on the other hand, is a more nuanced approach that encourages small weights for features by pushing them towards zero, but not necessarily making them be zero. For a more detailed discussion on DL regularization methods, see Goodfellow et al. [34].

Practical DL guidelines suggest using both L1 and L2 regularizations simultaneously to make a balance between the sparsity offered by L1 and the nuance offered by L2. Following these guidelines, we took advantage of both methods in training our network.

**3.2.2.1. Cost-sensitive learning.** Given the imbalanced nature of our data set (i.e., containing less than 20% positive cases) and the majority class's high degree of bias, using regular classification procedures did not yield acceptable per-class accuracy results. To address the data imbalance issue, several approaches are suggested in the literature, such as under- and over-sampling. Under-sampling generally leads to losing some information and may not be a good option in cases where the data is severely imbalanced, as is the case in the current study. Also, popular over-sampling approaches, such as SMOTE [35], typically require the majority of features to be continuous variables to perform well with minimum bias; hence, they are deemed improper choices for our data.

Another approach to address the data imbalance issue is to employ a cost-sensitive learning approach [36,37]. This approach is essentially designed for classification problems in which catching a specific

outcome correctly by the algorithm involves a much higher financial cost/benefit than incorrectly predicting the opposite outcome. As such, a higher cost is assigned to instances of the more important class in the data to make the algorithm produce a smaller number of false predictions for that class (i.e., the minority class). Similarly, for an imbalanced classification problem, we may give a higher weight to the minority class instances in the loss function to compensate for the bias caused by different class proportions. Although several methods are suggested for calculating weights for different classes, we used a simplistic approach by assuming an equal cost for false negative and false positive predictions. That is, given the 1000:226 ratio of negative and positive instances in the data, we used the cost matrix shown in Table 4 to customize the classification algorithm's loss function.

**3.2.2.2. Optimization of hyperparameters.** One of the challenges in working with deep networks is numerous hyperparameters that must be specified by the user. These include the network architecture hyperparameters (such as number of layers, number of neurons in each layer, and activation functions) and training hyperparameters (such as the optimizer algorithm, learning rate and its decay rate, batch size, number of epochs, loss function, and the regularization weights). A specific choice for each of these factors may work differently when combined with different values of other hyperparameters in the training phase of the network, and sometimes, slight changes can easily lead to over- or under-fitting of the predictive model. Hence, choosing a good combination of values for the hyperparameters is itself seen as a standalone optimization problem. To address this issue, a variety of approaches have been proposed in the literature, such as grid search [38], random search [39], Gaussian process [40], and sequential model-based optimization [41], just to name a few.

For this study, we employed a grid search approach to optimize the training hyperparameters (i.e., learning rate and its decay, batch size, epochs, and optimizer method), while keeping the network architecture hyperparameters fixed. We purposely did not consider varying every hyperparameter to keep the number of permutations in a reasonable range. This allowed the grid search to be completed in a reasonable time given the technical computing limitations. The specific grid search settings are elaborated in the Results Section.

### 3.2.3. Model interpretation

Machine learning techniques in general, and artificial neural networks in particular, have long been known as "black box" approaches with decent predictive power, but little to no interpretability. In recent years, however, there has been a surge of studies proposing various approaches, such as LIME [42], DeepLIFT [43], and Layer-Wise Relevance Propagation [44], to improve the interpretability aspect of ML.<sup>3</sup> Most of these methods aim at estimating the classic Shapley regression values [45] (i.e., feature importance for linear models in the presence of multicollinearity) for complex ML models using cooperative game theory equations [46]. Assuming each feature value of a given instance as a "player" and the prediction for that instance as the "payout" in a game, Shapley values determine how to fairly divide the payout among the features.

Lundberg and Lee [46] propose an intuitive approach called *SHaply*

**Table 4**  
Classification cost function weights.

	Actual negative	Actual positive
Predict negative	0	1
Predict positive	0.226	0

<sup>2</sup> Both of these proportions for L1 and L2 regularization methods are model hyperparameters and need to be specified by the user.

<sup>3</sup> See [89,90] for a comprehensive review of such methods.

Additive exPlanations (SHAP) to interpret complex predictive models, such as ensemble and deep learning models. SHAP combines the classic Shapley values approach with a couple of other agnostic methods (including LIME and DeepLIFT) and assigns each feature in the model an additive importance score. The importance score for each feature represents the change in the expected model prediction when conditioning on that feature.

To calculate the importance score of any given instance  $i$  of feature  $X$  (i.e.,  $X(i)$ ), the SHAP approach considers all feature subsets (not including  $X$  itself), then computes the effect on predictions (i.e.,

deviation from the average of all predictions) of adding  $X(i)$  to all those subsets. Essentially, one important difference between SHAP and the classic Shapley values approach is the “local accuracy” property involved in SHAP, which enables it to explain every instance of a factor in the data by calculating a single marginal contribution for that instance, whereas Shapley values just assign an importance score to the whole factor (and not to each instance of data) [46]. As a result, using SHAP we have an additive set of marginal contributions for each instance whose aggregation yields the prediction for that instance. Hence, in the specific context of this study, each patient’s readmission

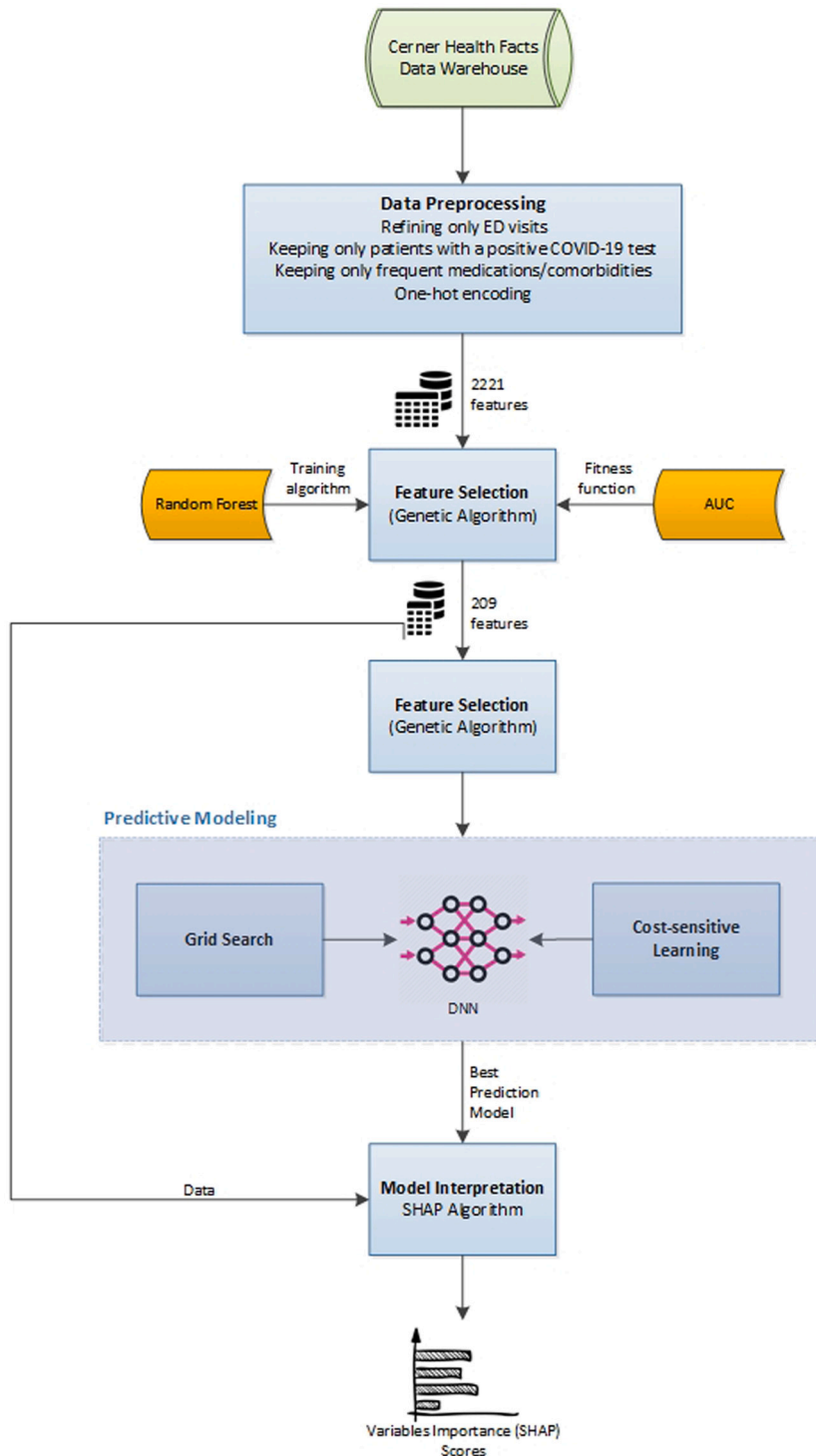


Fig. 2. Summary of the proposed framework.

probability (prediction) can be explained in terms of the additive marginal contributions of her corresponding health factors. Also, the overall importance score of each factor is simply the average of its marginal contributions across all instances (i.e., all patients in our context).

In the last stage of our proposed framework, and after developing a predictive model (stage 2) using the optimally selected features (stage 1), we employed SHAP to interpret the complex DL model we trained and to identify the major factors that contributed to one's readmission to the ED within a week after discharge. Fig. 2 summarizes our proposed framework.

## 4. Results

### 4.1. Feature selection

We ran the GA model on a computer with i9 2.90 GHz 8 Core processing power and a 64GB memory. The model converged after 42 generations, which took around 36 h of processing.<sup>4</sup> The optimized feature set involved a total of 209 variables and the basic random forest model trained by those features had an AUC of 0.812 (accuracy: 0.79; sensitivity: 0.61; F1-score: 0.54). From the selected features, 81 were comorbidity/symptom-related, 123 were medication indicator variables, and the remaining 5 were all the demographic features existing in the original data.

### 4.2. Prediction model

In this study, while holding the hyperparameters of the network architecture fixed (i.e., the number of hidden layers, the number of neurons, and activation functions), we employed a grid search approach to optimize the training hyperparameters such that they would work best with the selected network architecture and would minimize the *weighted binary cross-entropy* loss function. Table 5 indicates the values used for each of the training hyperparameters in the grid search.

For each permutation, the neural network was trained using 80% of the data and was validated on the remaining 20% (a random stratified sampling approach was used for partitioning). We used the TensorFlow package (with Keras backend) in Python to perform the model training. The grid search optimization was managed using the Talos package<sup>5</sup> in Python, which works seamlessly with TensorFlow.

We ran the algorithm on a workstation with two NVIDIA TITAN XP GPUs (working in parallel) and 64GB of memory. It took around 110 h for the algorithm to run all the 6480 permutations of the grid search (i.

**Table 5**  
Grid search settings for hyperparameter optimization.

Hyperparameter	Range of values	Number of values tested
Optimizer	[Adam, Nadam, RMSProp]	3
Learning rate	[0.005, 0.05]	10
Decay rate	[0.05, 0.30]	6
Batch size	[4, 8, 16, 32]	4
Epochs	[100, 200, 300]	3
Regularization weight	[0.0001, 0.001, 0.01]	3
	Total number of permutations	6480

<sup>4</sup> We performed the same procedure with 110 (instead of 220) as the target number of selected features as well. The GA took around 8 h to converge in this case. The results showed a roughly 5% decrease in the AUC of the model including the best feature set. While it suggested that including more than 220 features (e.g., 330) in the final data could have possibly improved the AUC of the RF model, we believe that the expected improvement is not worth the complexity added to the GA model, given the exponentially higher computational requirements resulted from that change.

<sup>5</sup> <https://github.com/autonomio/talos>

e., around 0.61 s per epoch). The (nearly) optimal hyperparameter settings obtained from the grid search are shown in Table 6.

Since a grid search strategy does not exhaust the entire solution space, after obtaining the nearly optimal set of hyperparameters, we performed a manual search in the vicinity of each of the hyperparameters (except the optimizer) by slightly changing them one at a time and observing the model outcome. We realized that a learning rate of 0.007 with a decay rate of 0.2 led to a relatively steady training process with better performance on the validation data set.

With a dichotomous target variable indicating whether a patient with a positive COVID-19 test returned to the ED within 7 days after discharge (1) or not (0), our best artificial neural network achieved an accuracy of 87.4% with an area under the Receiver Operating Characteristic (ROC) curve value of 0.883 (Fig. 3), a sensitivity of 71.9%, a specificity of 91.5%, and an F-1 measure of 70.4%.

For severely imbalanced classification problems, Sánchez-Hernández et al. [47] suggest G-mean, the geometric mean of true positive rate (TPR) and true negative rate (TNR) (Eq. (1)), as an objective measure of predictive power. They maintain that G-mean is more indicative of the performance of such classification models than F-measure or precision; two metrics that are basically designed for classification tasks with fairly balanced data. In our model, the G-mean measure was 81.2%.

$$G - mean = \sqrt{TPR * TNR} \quad (1)$$

While, to the best of our knowledge, the ED readmission prediction for COVID-19 patients has not been addressed in any prior study, we believe a comparison between our model with similar recent ED readmission studies in other contexts could demonstrate the utility of our proposed predictive approach. This comparison is given in Table 7.

Particularly, the difference between sensitivity and AUC of the present study with those of other studies is notable, suggesting the remarkably higher distinctive power of our proposed model. Even though the accuracy reported by Sarasa Cabezuelo [48] is considerably high, the AUC reported in that study suggests that the high accuracy is partially due to the use of a highly imbalanced data set (and therefore, a large gap between their model's sensitivity and specificity).

### 4.3. Model interpretation

The predicted probabilities by the best DNN model, along with the training and test data, were fed into the SHAP algorithm to assess the feature importance scores. The SHAP package<sup>6</sup> in Python was employed to perform these analyses. With a large number of features and instances, the runtime for the SHAP algorithm is considerably high. A suggested workaround to deal with this issue is to use a set of weighted k-means of instances (each weighted by the number of instances it represents) rather than the whole training data. Using this approach, we summarized the training data into 4 weighted k-means (using the *shap.kmeans()* method) and used them to train the SHAP algorithm.

Fig. 4 indicates the most important features in terms of SHAP scores, top medications, and top comorbid conditions, respectively. Bars shown

**Table 6**  
Optimal set of hyperparameters.

Hyperparameter	Optimal value
Optimizer	Adam
Learning rate	0.01
Decay rate	0.25
Batch size	8
Epochs	300
Regularization weight	0.0001

<sup>6</sup> <https://github.com/slundberg/shap>

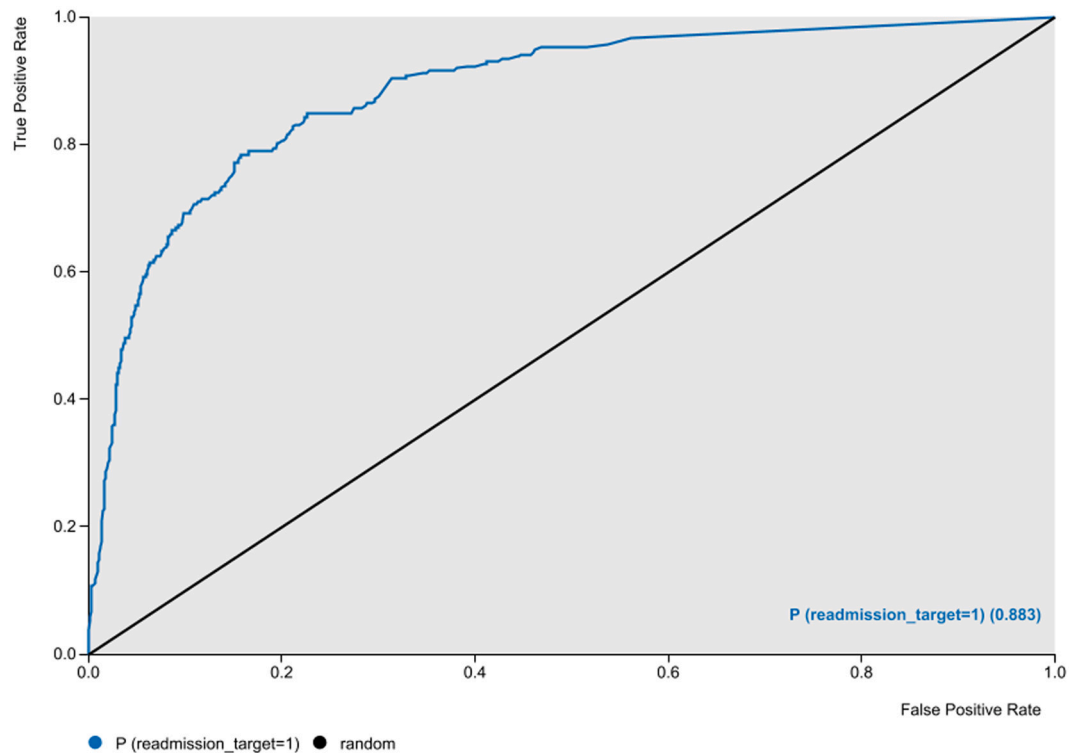


Fig. 3. The Receiver Operating Characteristic (ROC) curve of the best ANN model.

Table 7  
Prediction model results comparison with similar studies.

Article	Context	Readmission window	Approach	Model metrics
Present study	COVID-19	7-day	GA + DNN	Acc = 0.874 Sens = 0.719 Precision = 0.691 F1 = 0.704 G-mean = 0.812 AUC = 0.883
[81]	General	30-day	Decision Tree	Acc = 0.772 Sens = 0.402 F1 = 0.494 AUC = 0.732
[82]	General	3-day	Gradient Boosting	Sens = 0.16 Precision = 0.75 AUC = 0.76
[82]	General	9-day	Gradient Boosting	Sens = 0.23 Precision = 0.70 AUC = 0.75
[83]	General	3-day	DNN	Acc = 0.680 Sens = 0.679 AUC = 0.755
[48]	General	3-day	Ensemble	Acc = 0.957 AUC = 0.61

in green/red represent factors that decrease/increase the chances of readmission. As shown in Fig. 4, Enoxaparin, a drug typically used to prevent the formation of blood clots, turned out as the most important factor in decreasing the chances of readmission. Several studies have reported a suspiciously significant association between COVID-19 and Thromboembolism [49–52]. Our results suggest that, regardless of whether that association is causal, administering blood clot preventive medications could be effective in treating COVID-19 patients and

decreasing their chances of returning to ED.

Another medication shown to decrease the chances of readmission is Hydroxychloroquine (HCQ), a drug well-known for treating malaria. Since the beginning of the pandemic, HCQ has been one of the controversial treatments for COVID-19. Several clinical trials have been conducted to investigate its efficacy, with some reporting positive effects [53], while a larger group reporting no significant effect [54–57]. Mahase [58] argues that while there is no solid proof that HCQ is effective in treating COVID-19, some clinical trials report its efficacy in reducing severe symptoms of the disease.<sup>7,8</sup> We believe our findings confirm the results of those trials since ED visits are typically initiated after observing severe symptoms in the patients.

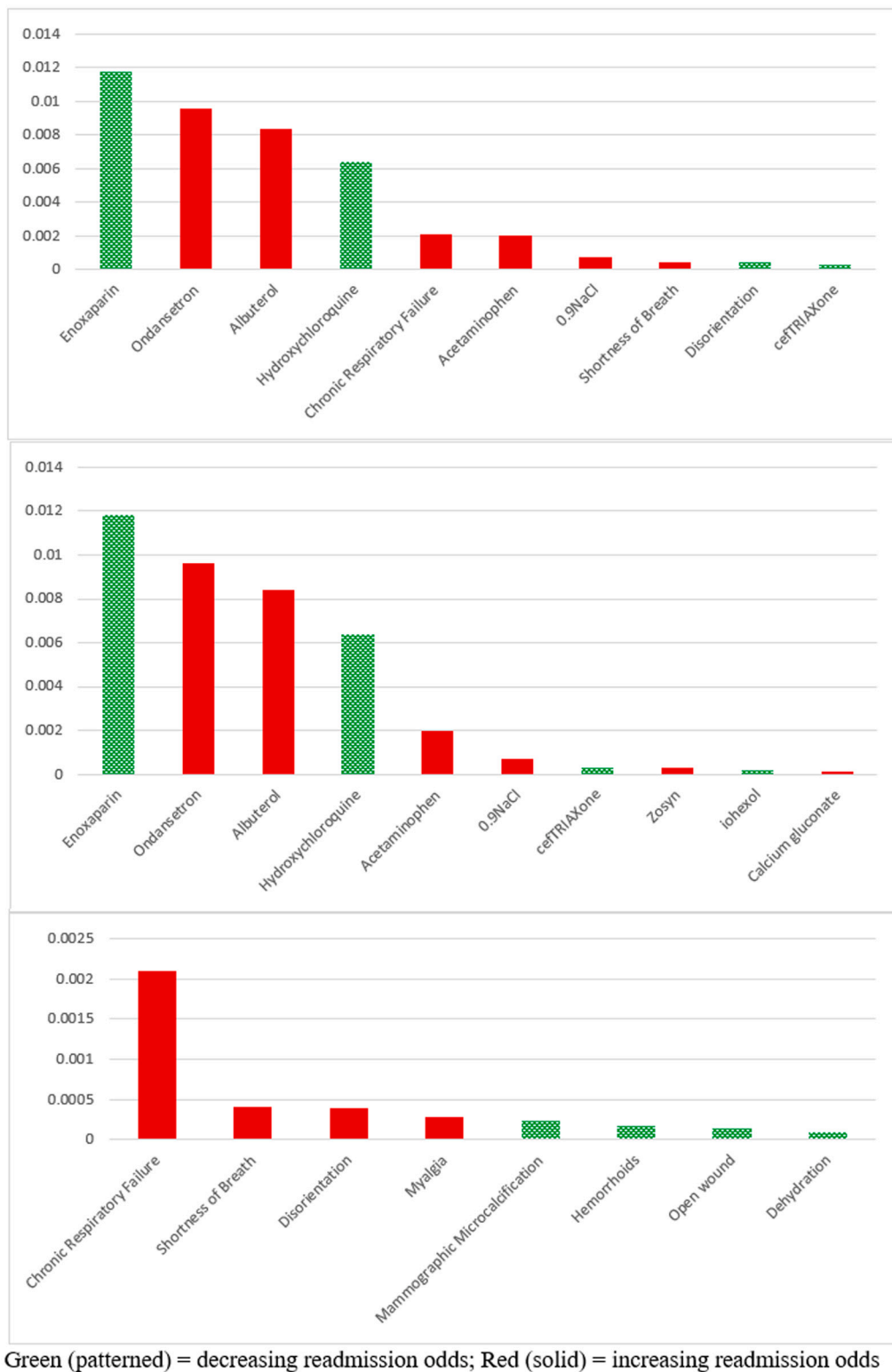
On the other hand, administering Ondansetron and Albuterol, two drugs that are typically used to treat nausea and shortness of breath, have turned out to increase the chances of returning to ED. The Centers for Disease Prevention and Control (CDC) has officially listed both conditions as typical symptoms of COVID-19. Since we could not find any notable clinical studies discussing these two drugs as possible treatments of COVID-19, it seems that they were mainly prescribed by doctors to alleviate nausea and dyspnea. However, according to our SHAP importance scores, it appears that these medications do not act as effectively on COVID-19 patients as they do on patients with other conditions.

In terms of comorbidity/symptom factors, the bottom chart in Fig. 4 indicates that chronic respiratory failure and shortness of breath are the most important reasons for patients' readmission to ED. While CDC has identified respiratory failure as a condition with a moderate risk of complication in COVID-19 patients [59], it is surprising to see that some patients with that condition had been discharged from ED after being tested positive for COVID-19. One explanation for this observation could

<sup>7</sup> <https://www.principletrial.org/news/principle-trial-rolled-out-across-uk-homes-and-communities>

<sup>8</sup> <https://www.nih.gov/news-events/news-releases/nih-begins-clinical-trial-hydroxychloroquine-azithromycin-treat-covid-19>





Green (patterned) = decreasing readmission odds; Red (solid) = increasing readmission odds

Fig. 4. SHAP importance scores of top features overall (top), medications (middle), and comorbidities (bottom). Green (patterned) = decreasing readmission odds; Red (solid) = increasing readmission odds.

be the limited capacity of health care organizations and the need to prioritize the allocation of hospital beds. This could also justify why high-risk comorbid conditions (e.g., COPD and chronic heart failure) did not turn out among the top factors in our study because patients with those conditions were most likely hospitalized once diagnosed with COVID-19. Our results suggest, however, that when hospital beds are

available, patients with COPD or heart diseases are better to be hospitalized because they have a higher probability of returning to the ED within a short time.

Since the outcome of SHAP analysis (as a post-hoc interpretation method) depends generally on the existing features in the model and because we employed an evolutionary approach (with some degree of

randomness) for feature selection, some concern may be raised about the robustness of our results. To address such concerns, we repeated the SHAP analysis with 50 and 110 features selected by GA in two separate runs. In both cases, the model converged before reaching the maximum number of generations. We observed that: 1) A majority of the features selected in these extra runs were among the 220 features used in the original run (88% and 87.2% for the 50- and 110-feature re-runs, respectively); 2) except “Ceftriaxone” and “open wound” which were not among the 50 selected features, all top features identified in Fig. 4 were present among the selected features in both new runs.; and 3) after calculating the average SHAP scores for the two new feature sets and sorting the features by their scores, we observed a Spearman’s rank-order correlation of 0.92 and 0.89 between the selected features in the two new runs (50 and 110 features, respectively) and the features from the original feature selection run (i.e., 220 features). Overall, we believe that these additional analyses corroborate the robustness of our results. We believe that the reasonable settings we considered for the feature selection step enabled the GA to search the solution space sufficiently and to go through enough iterations to converge.

In addition to the insights provided at the aggregate level, the additive nature of SHAP scores enables the practitioners to analyze the patient-specific risk factors at the individual level. In other words, sum of the features’ SHAP scores for each patient represents his or her corresponding deviation from the average (across the entire sample) probability of readmission. In addition, every single feature’s SHAP score for a specific patient represents the contribution of that feature to

the patient’s total deviation from the average readmission probability. The waterfall chart in Fig. 5 shows the contributing risk factors (and the extent of their contribution) to the probability of readmission for a sample patient. The column on the left shows the average probability of readmission across all patients, serving as a baseline for interpreting the additive SHAP scores. Fig. 5 only displays diagnosis- or symptom-related factors because these factors are present at the time of a patients’ initial visit and are the main criteria for physicians to decide on the hospitalization of the patient.

The figure shows the important factors specific to a female patient (63 years old) who returned to the ED within a week after her initial visit. The predicted probability by the model for the patient’s readmission was 0.842, which is 0.411 higher than the average probability across all patients. The length of the bar associated with each factor explains how much of that gap is attributable to that factor (red/green bars for increasing/decreasing effects). In line with the aggregate findings (Fig. 4), acute respiratory failure is a major factor for this patient’s return to the ED (this condition has increased her chances of readmission by around 0.27). This strongly suggests that patients experiencing this symptom at the time of their first visit to the ED should be given a high priority for hospitalization. Thus, we are able to confirm the findings of previous studies (e.g., [60]) in which respiratory failure at the time of admission was identified as a major mortality risk factor. Moreover, Fig. 5 suggests that diabetes mellitus type II, genetic susceptibility to diseases, and deficiency of the immune system are the other important risk factors leading this particular patient to be readmitted. A review of

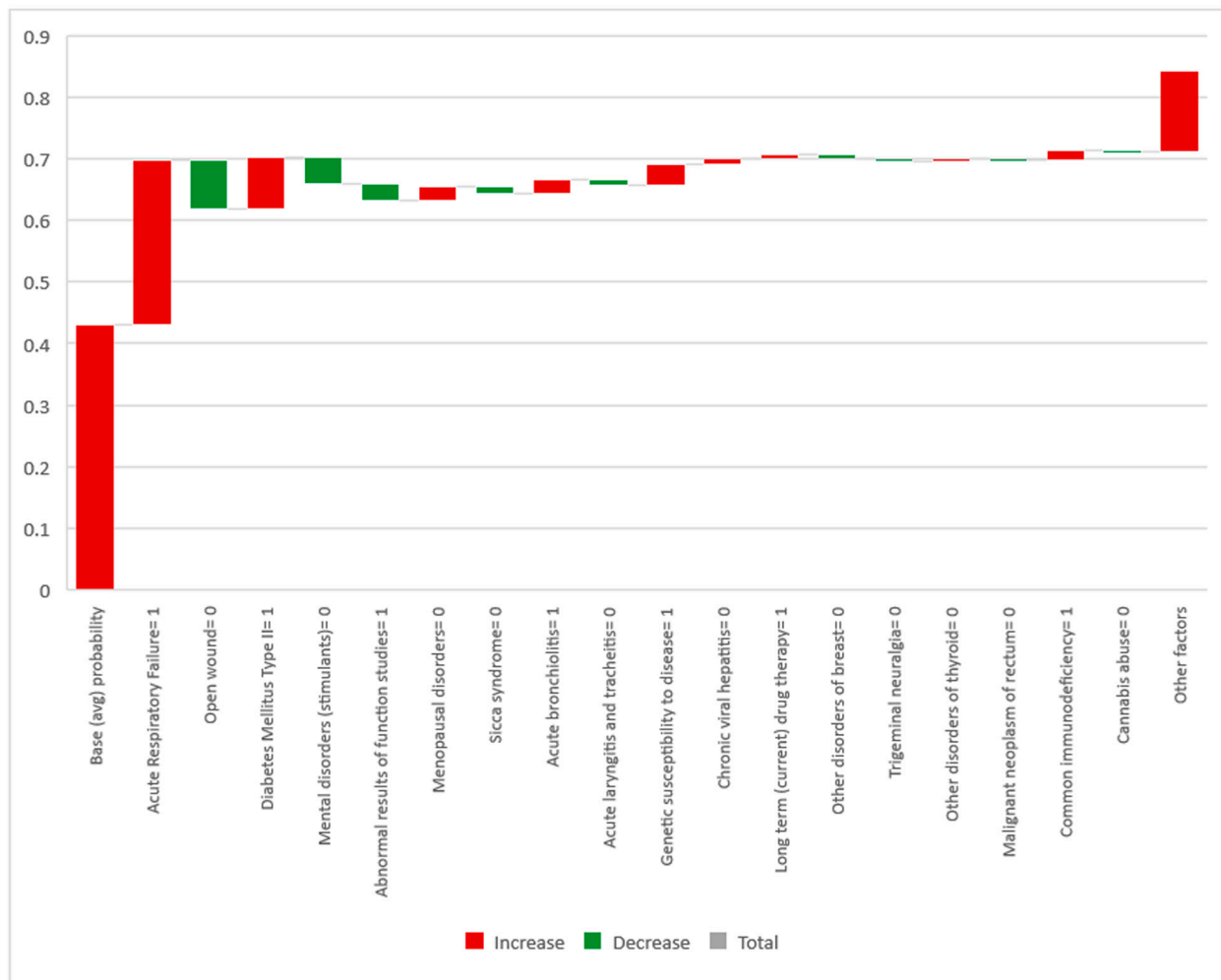


Fig. 5. SHAP feature importance scores of a sample patient (individual level).

the relevant literature reveals that these results confirm the findings by several clinical trial studies investigating the association between diabetes [61–63], genetic susceptibility [64,65], and immunodeficiency [66,67] with the severity of COVID-19.

Clearly, since the risk factors may differ from patient to patient, several factors identified in Fig. 4 (at the aggregate level) are not present in Fig. 5 for this particular patient. This sheds light on the importance of studying risk factors at the patient level in addition to the cohort level. Although the risk factors identified in cohort-level studies (especially clinical trials) are typically more common and generalizable, investigating those factors at the patient level may reveal somewhat rare risk factors as well as capture the effect of comorbidities (i.e., coexistence of multiple risk factors).

## 5. Summary and conclusion

In this study, we proposed an exploratory/predictive/explanatory machine learning framework that can be employed as a decision support tool to help clinicians identify high-risk patients and critical medical factors in a reasonable timeframe during pandemics.

The mysterious nature of novel viral diseases in their early stages of outbreak makes clinical predictive analytics a challenging activity mainly because such analyses rely on prior knowledge about the potential predictors of the outcome to limit the feature space and to reduce the dimensionality. In the absence of such knowledge, an *exploratory* effort is needed to identify the highly relevant factors. Additionally, a purely *predictive* machine learning approach barely provides insights for clinicians to gain a better understanding of the disease; hence, a predictive effort should ideally be followed by an *explanatory* one to simplify the complexity of predictive models and to provide intuitive guidelines. Our proposed framework offers such a holistic package to clinical decision-makers and enables them to make informative decisions even in the early stages of a pandemic.

While we tested our framework in the specific context of pandemic-related ED readmissions, we believe that it is highly generalizable to other, even non-pandemic, contexts where both prior knowledge and historical data are limited. We encourage future research to validate the applicability of this framework to other contexts.

## Declaration of Competing Interest

None.

## Acknowledgements

This work was conducted with data from the Cerner Corporation's HealthFacts datawarehouse of electronic medical records provided by the Oklahoma State University Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Cerner Corporation.

## References

- M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, E. Liu, K. Upadhyaya, Correction: Predicting all-cause risk of 30-day hospital readmission using artificial neural networks, *PLoS One* 13 (5) (2018), e0197793.
- H. Wang, Z. Cui, Y. Chen, M. Avidan, A. Ben Abdallah, A. Kronzer, Predicting hospital readmission via cost-sensitive deep learning, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15 (6) (2018) 1968–1978.
- B. Kauffman, Readmissions & Medicare: What's the Cost? National Investment Center, 2016.
- A.D. Auerbach, S. Kripalani, E.E. Vasilevskis, N. Sehgal, P.K. Lindenauer, J. P. Metlay, G. Fletcher, G.W. Ruhnke, S.A. Flanders, C. Kim, Preventability and causes of readmissions in a national cohort of general medicine patients, *JAMA Intern. Med.* 176 (4) (2016) 484–493.
- L. Wilson, MA Patients' Readmission Rates Higher Than Traditional Medicare, *Study Finds*, *HealthcareDive*, 2019, June.
- C. Fischer, H.F. Lingsma, P.J. Marang-van de Mheen, D.S. Kringos, N.S. Klazinga, E. W. Steyerberg, Is the readmission rate a valid quality indicator? A review of the evidence, *PLoS One* 9 (11) (2014), e112282.
- S. Klein, Focus: Preventing Unnecessary Hospital Readmissions, 2008 (Quality Matters).
- A. Beall, The Heart-wrenching Choice of Who Lives and Dies - BBC Future, BBC News, 2020.
- L.M. Monella, Coronavirus: Italy Doctors "Forced to Prioritize ICU Care for Patients With Best Chance of Survival", *Euronews*, 2020, March.
- Y. Mounk, The Extraordinary Decisions Facing Italian Doctors. *The Atlantic*, 2020.
- D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, S. Kripalani, Risk prediction models for hospital readmission, *JAMA* 306 (15) (2011) 1688, <https://doi.org/10.1001/jama.2011.1515>.
- H.M. Zolbanin, D. Delen, Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions, *Decis. Support. Syst.* 112 (2018) 98–110, <https://doi.org/10.1016/j.dss.2018.06.010>.
- G. Shmueli, O.R. Koppius, Predictive analytics in information systems research, in: *MIS Quarterly* vol. 35, Management Information Systems Research Center, University of Minnesota, 2011, pp. 553–572, <https://doi.org/10.2307/23042796>.
- A. Artetxe, A. Beristain, M. Grana, Predictive models for hospital readmission risk: a systematic review of methods, *Comput. Methods Prog. Biomed.* 164 (2018) 49–64.
- Z. Obermeyer, B. Cohn, M. Wilson, A.B. Jena, D.M. Cutler, Early death after discharge from emergency departments: Analysis of national US insurance claims data, *BMJ (Online)* 356 (2017), <https://doi.org/10.1136/bmj.j239>.
- O.S. Gunnarsdottir, V. Rafnsson, Death within 8 days after discharge to home from the emergency department, *Eur. J. Pub. Health* 18 (5) (2008) 522–526, <https://doi.org/10.1093/eurpub/ckn045>.
- J.S. Fan, W.F. Kao, D.H.T. Yen, L.M. Wang, C.I. Huang, C.H. Lee, Risk factors and prognostic predictors of unexpected intensive care unit admission within 3 days after ED discharge, *Am. J. Emerg. Med.* 25 (9) (2007) 1009–1014, <https://doi.org/10.1016/j.ajem.2007.03.005>.
- D.L. Chin, H. Bang, R.N. Manickam, P.S. Romano, Rethinking thirty-day hospital readmissions: shorter intervals might be better indicators of quality of care, *Health Aff.* 35 (10) (2016) 1867–1875.
- K.L. Graham, A.D. Auerbach, J.L. Schnipper, S.A. Flanders, C.S. Kim, E.J. Robinson, G.W. Ruhnke, L.R. Thomas, S. Kripalani, E.E. Vasilevskis, Preventability of early versus late hospital readmissions in a national cohort of general medicine patients, *Ann. Intern. Med.* 168 (11) (2018) 766–774.
- A.M. Lavery, L.E. Preston, J.Y. Ko, J.R. Chevinsky, C.L. DeSisto, A.F. Pennington, L. Kompaniyets, S.D. Datta, E.S. Click, T. Golden, Characteristics of hospitalized COVID-19 patients discharged and experiencing same-hospital readmission—United States, March–August 2020, *Morb. Mortal. Wkly Rep.* 69 (45) (2020) 1695.
- M.J. Press, Z. Meisel, M. Pesko, A. Ryan, Impact of public reporting of hospital readmission rates on emergency department admission decisions, *J. Gen. Intern. Med.* 28 (2013) S99–S100.
- L. Sivasundaram, N.N. Trivedi, C.-Y. Kim, J. Du, R.W. Liu, J.E. Voos, M. Salata, Emergency department utilization following elective hip arthroscopy, *Arthroscopy* 36 (6) (2020) 1575–1583.
- I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- H.M. Zolbanin, B. Davazdahemami, D. Delen, A.H. Zadeh, Data analytics for the sustainable use of resources in hospitals: predicting the length of stay for patients with chronic diseases, *Inf. Manag.* (2020), 103282, <https://doi.org/10.1016/j.im.2020.103282>.
- R. Sharda, D. Delen, E. Turban, *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*, 11th ed., Pearson, 2019.
- A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, *Proc. Icm1* 30 (1) (2013) 3.
- K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.
- D. Mishkin, J. Matas, All You Need is a Good Init, 2015 (ArXiv Preprint ArXiv: 1511.06422).
- W. An, H. Wang, Y. Zhang, Q. Dai, Exponential decay sine wave learning rate for fast deep neural network training, in: *2017 IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4, <https://doi.org/10.1109/VCIP.2017.8305126>.
- Z. Li, S. Arora, An Exponential Learning Rate Schedule for Deep Learning, 2019.
- R. Ge, S.M. Kakade, R. Kidambi, P. Netrapalli, The step decay schedule: a near optimal, geometrically decaying learning rate procedure for least squares, *Adv. Inform. Proc. Syst.* (2019) 14977–14988.
- S. Seong, Y. Lee, Y. Kee, D. Han, J. Kim, Towards Flatter Loss Surface via Nonmonotonic Learning Rate Scheduling, *UAI*, 2018, pp. 1020–1030.
- R. Reed, R.J. MarksII, *Neural Smoothing: Supervised Learning in Feedforward Artificial Neural Networks*, MIT Press, 1999.
- I. Goodfellow, Y. Bengio, A. Courville, Regularization for deep learning, *Deep Learn.* (2016) 216–261.
- N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (2002) 321–357.
- M. Kukar, I. Kononenko, Cost-sensitive learning with neural networks, *ECAI* 98 (1998) 445–449.
- B. Zadrozny, J. Langford, N. Abe, Cost-sensitive learning by cost-proportionate example weighting, in: *Third IEEE International Conference on Data Mining*, 2003, pp. 435–442.

- [38] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, Y. Bengio, An empirical evaluation of deep architectures on problems with many factors of variation, in: *Proceedings of the 24th International Conference on Machine Learning*, 2007, pp. 473–480.
- [39] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *J. Machin. Learn. Res.* 13 (1) (2012) 281–305.
- [40] C.E. Rasmussen, Gaussian processes in machine learning, in: *Summer School on Machine Learning*, 2003, pp. 63–71.
- [41] F. Hutter, H.H. Hoos, K. Leyton-Brown, Sequential model-based optimization for general algorithm configuration, in: *International Conference on Learning and Intelligent Optimization*, 2011, pp. 507–523.
- [42] M.T. Ribeiro, S. Singh, C. Guestrin, “Why should I trust you?” Explaining the predictions of any classifier, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.
- [43] A. Shrikumar, P. Greenside, A. Kundaje, Learning Important Features Through Propagating Activation Differences, 2017 (ArXiv Preprint ArXiv:1704.02685).
- [44] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, *PLoS One* 10 (7) (2015), e0130140.
- [45] L.S. Shapley, *Utility Comparison and the Theory of Games*, Cambridge Univ Pr, 1988.
- [46] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Adv. Neural Inform. Proc. Syst.* (2017) 4765–4774.
- [47] F. Sánchez-Hernández, J.C. Ballesteros-Herráez, M.S. Kraiem, M. Sánchez-Barba, M.N. Moreno-García, Predictive modeling of ICU healthcare-associated infections from imbalanced data. Using ensembles and a clustering-based undersampling approach, *Appl. Sci.* 9 (24) (2019) 5287.
- [48] A. Sarasa Cabezuelo, Application of machine learning techniques to analyze patient returns to the emergency department, *J. Personal. Med.* 10 (3) (2020), <https://doi.org/10.3390/jpm10030081>.
- [49] S. Cui, S. Chen, X. Li, S. Liu, F. Wang, Prevalence of venous thromboembolism in patients with severe novel coronavirus pneumonia, *J. Thromb. Haemost.* 18 (6) (2020) 1421–1424.
- [50] G.B. Danzi, M. Loffi, G. Galeazzi, E. Gherbesi, Acute pulmonary embolism and COVID-19 pneumonia: a random association? *Eur. Heart J.* 41 (19) (2020) 1858, <https://doi.org/10.1093/eurheartj/ehaa254>.
- [51] F. Grillet, J. Behr, P. Calame, S. Aubry, E. Delabrousse, Acute pulmonary embolism associated with COVID-19 pneumonia detected by pulmonary CT angiography, *Radiology* 296 (3) (2020) E186–E188, 201544.
- [52] V. Janardhan, V. Janardhan, V. Kalousek, COVID-19 as a blood clotting disorder masquerading as a respiratory illness: a cerebrovascular perspective and therapeutic implications for stroke thrombectomy, *J. Neuroimaging* 30 (5) (2020) 555–561.
- [53] X. Li, Y. Wang, P. Agostinis, A. Rabson, G. Melino, E. Carafoli, Y. Shi, E. Sun, Is hydroxychloroquine beneficial for COVID-19 patients? *Cell Death Dis.* 11 (7) (2020) 1–6.
- [54] S. Abd-El Salam, E.S. Esmail, M. Khalaf, E.F. Abdo, M.A. Medhat, M.S. Abd El Ghafar, O.A. Ahmed, S. Soliman, G.N. Serangaw, M. Alborai, Hydroxychloroquine in the treatment of COVID-19: a multicenter randomized controlled study, *Am. J. Trop. Med. Hyg.* 103 (4) (2020) 1635–1639.
- [55] Group, R. C., Effect of hydroxychloroquine in hospitalized patients with Covid-19, *N. Engl. J. Med.* 383 (21) (2020) 2030–2040.
- [56] P. Horby, M. Mafham, L. Linsell, J.L. Bell, N. Staplin, J.R. Emberson, M. Wiselka, A. Ustianowski, E. Elmahi, B. Prudon, Effect of Hydroxychloroquine in hospitalized patients with COVID-19: preliminary results from a multi-centre, randomized, controlled trial, *MedRxiv* (2020), <https://doi.org/10.1056/NEJMoa2022926>.
- [57] I. Torjesen, Covid-19: hydroxychloroquine does not benefit hospitalised patients, UK trial finds, *BMJ* (2020) 369.
- [58] E. Mahase, Hydroxychloroquine for covid-19: the end of the line? *Bmj* 369 (2020).
- [59] J. Cates, C. Lucero-Obusan, R.M. Dahl, P. Schirmer, S. Garg, G. Oda, A.J. Hall, G. Langley, F.P. Havers, M. Holodny, Risk for in-hospital complications associated with COVID-19 and influenza—Veterans Health Administration, United States, October 1, 2018–May 31, 2020, *Morb. Mortal. Wkly Rep.* 69 (42) (2020) 1528.
- [60] P. Santus, D. Radovanovic, L. Saderi, P. Marino, C. Cogliati, G. De Filippis, M. Rizzi, E. Franceschi, S. Pini, F. Giuliani, M. Del Medico, G. Nucera, V. Valenti, F. Tursi, G. Sotgiu, Severity of respiratory failure at admission and in-hospital mortality in patients with COVID-19: a prospective observational multicentre study, *BMJ Open* 10 (10) (2020), e043651, <https://doi.org/10.1136/bmjopen-2020-043651>.
- [61] L. Fang, G. Karakiulakis, M. Roth, Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir. Med.* 8 (4) (2020), e21.
- [62] R. Muniyappa, S. Gubbi, COVID-19 pandemic, coronaviruses, and diabetes mellitus, *Am. J. Physiol.* 318 (5) (2020) E736–E741.
- [63] Y. Zhang, Y. Cui, M. Shen, J. Zhang, B. Liu, M. Dai, L. Chen, D. Han, Y. Fan, Y. Zeng, Association of diabetes mellitus with disease severity and prognosis in COVID-19: a retrospective cohort study, *Diabetes Res. Clin. Pract.* 165 (2020) 108227.
- [64] D. Gemmati, B. Bramanti, M.L. Serino, P. Secchiero, G. Zauli, V. Tisato, COVID-19 and individual genetic susceptibility/receptivity: role of ACE1/ACE2 genes, immunity, inflammation and coagulation. Might the double X-chromosome in females be protective against SARS-CoV-2 compared to the single X-chromosome in males? *Int. J. Mol. Sci.* 21 (10) (2020) 3474.
- [65] K.J. Godri Pollitt, J. Peccia, A.I. Ko, N. Kaminski, C.S. Dela Cruz, D.W. Nebert, J.K. V. Reichardt, D.C. Thompson, V. Vasilou, COVID-19 vulnerability: the potential impact of genetic susceptibility and airborne transmission, *Hum. Genom.* 14 (2020) 1–7.
- [66] Y. Gao, Y. Chen, M. Liu, S. Shi, J. Tian, Impacts of immunosuppression and immunodeficiency on COVID-19: a systematic review and meta-analysis, *J. Infect.* 81 (2) (2020), e93.
- [67] G. Härter, C.D. Spinner, J. Roeder, M. Bickel, I. Krznic, S. Grunwald, F. Schabaz, D. Gillor, N. Postel, M.C. Mueller, COVID-19 in people living with human immunodeficiency virus: a case series of 33 patients, *Infection* (2020) 1.
- [68] B. Amalakuhan, L. Kiljanek, A. Parvathaneni, M. Hester, P. Cheriya, D. Fischman, A prediction model for COPD readmissions: catching up, catching our breath, and improving a national problem, *J. Commun. Hosp. Int. Med. Perspect.* 2 (1) (2012) 9915.
- [69] C. Walsh, G. Hripcsak, The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions, *J. Biomed. Inform.* 52 (2014) 418–426, <https://doi.org/10.1016/j.jbi.2014.08.006>.
- [70] J. Futoma, J. Morris, J. Lucas, A comparison of models for predicting early hospital readmissions, *J. Biomed. Inform.* 56 (2015) 229–238, <https://doi.org/10.1016/J.JBI.2015.05.016>.
- [71] S. Yu, F. Farooq, A. Van Esbroeck, G. Fung, V. Anand, B. Krishnapuram, Predicting readmission risk with institution-specific prediction models, *Artif. Intell. Med.* 65 (2) (2015) 89–96.
- [72] B. Zheng, J. Zhang, S.W. Yoon, S.S. Lam, M. Khasawneh, S. Poranki, Predictive modeling of hospital readmissions using metaheuristics and data mining, *Expert Syst. Appl.* 42 (20) (2015) 7110–7120.
- [73] D. Agrawal, C.-B. Chen, R.W. Dravenstott, C.T.B. Strömbld, J.A. Schmid, J. D. Darer, P. Devapriya, S. Kumara, Predicting patients at risk for 3-day postdischarge readmissions, ED visits, and deaths, *Med. Care* 54 (11) (2016) 1017–1023.
- [74] L. Turgeman, J.H. May, A mixed-ensemble model for hospital readmission, *Artif. Intell. Med.* 72 (2016) 72–82, <https://doi.org/10.1016/j.artmed.2016.08.005>.
- [75] C. Chopra, S. Sinha, S. Jaroli, A. Shukla, S. Maheshwari, Recurrent neural networks with non-sequential data to predict hospital readmission of diabetic patients, in: *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics*, 2017, pp. 18–23.
- [76] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, E. Liu, Predicting all-cause risk of 30-day hospital readmission using artificial neural networks, *PLoS One* 12 (7) (2017), e0181173.
- [77] M. Mesgarpour, T. Chausalet, S. Chahed, Ensemble risk model of emergency admissions (ERMER), *Int. J. Med. Inform.* 103 (2017) 65–77, <https://doi.org/10.1016/j.ijmedinf.2017.04.010>.
- [78] B.K. Reddy, D. Delen, Predicting hospital readmission for lupus patients: an RNN-LSTM-based deep-learning methodology, *Comput. Biol. Med.* 101 (2018) 199–209.
- [79] X. Min, B. Yu, F. Wang, Predictive modeling of the hospital readmission risk from patients’ claims data using machine learning: a case study on COPD, *Sci. Rep.* 9 (1) (2019) 1–10.
- [80] K. Huang, J. Altosaar, R. Ranganath, Clinicalbert: Modeling clinical notes and predicting hospital readmission, in: *The ACM Conference on Health, Inference, and Learning*, 2020.
- [81] J.R. Vest, O. Ben-Assuli, Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information, *Int. J. Med. Inform.* 129 (2019) 205–210, <https://doi.org/10.1016/j.ijmedinf.2019.06.013>.
- [82] W.S. Hong, A.D. Haimovich, R.A. Taylor, Predicting 72-hour and 9-day return to the emergency department using machine learning, *JAMIA Open* 2 (3) (2019) 346–352, <https://doi.org/10.1093/jamiaopen/ooz019>.
- [83] B. Davazdahemami, P. Peng, D. Delen, A deep learning approach for predicting early bounce-backs to the emergency departments, *Health Analytics* 2 (2022), 100018, <https://doi.org/10.1016/j.health.2022.100018>. In press.
- [84] E. Atalla, M. Kalligeros, G. Giampaolo, E.K. Mylona, F. Shehadeh, E. Mylonakis, Readmissions among patients with COVID-19, *Int. J. Clin. Pract.* 75 (3) (2020), e13700.
- [85] W.-H. Jeon, J.Y. Seon, S.-Y. Park, I.-H. Oh, Analysis of risk factors on readmission cases of COVID-19 in the Republic of Korea: using nationwide health claims data, *Int. J. Environ. Res. Public Health* 17 (16) (2020) 5844.
- [86] L.M. Parra, M. Cantero, I. Morrás, A. Vallejo, I. Diego, E. Jiménez-Tejero, E. Múñez, Á. Asensio, A. Fernández-Cruz, A. Ramos-Martinez, Hospital readmissions of discharged patients with COVID-19, *Int. J. Gen. Med.* 13 (2020) 1359.
- [87] O.A. Uyaroglu, N.Ç. BAŞARAN, L. Özişik, G.T. Dizman, İ. Eroglu, T.K. Şahin, Z. Taş, A.Ç. Inkaya, M.D. Tanriöver, G. Metan, Thirty-day readmission rate of COVID-19 patients discharged from a tertiary care university hospital in Turkey: an observational, single-center study, *Int. J. Qual. Health Care* 33 (1) (2020), mzaa144.
- [88] J.P. Donnelly, X.Q. Wang, T.J. Iwashyna, H.C. Prescott, Readmission and death after initial hospital discharge among patients with COVID-19 in a large multihospital system, *JAMA* 325 (3) (2020) 304–306.
- [89] D.V. Carvalho, E.M. Pereira, J.S. Cardoso, Machine learning interpretability: a survey on methods and metrics, *Electronics* 8 (8) (2019) 832.
- [90] L.H. Gilpin, D. Bau, B.Z. Yuan, A. Bajwa, M. Specter, L. Kagal, Explaining explanations: an overview of interpretability of machine learning, in: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 2018, pp. 80–89, <https://doi.org/10.1109/DSAA.2018.00018>.



**Behrooz Davazdahemami** is an Assistant Professor of Information Technology and Supply Chain Management (IT&SCM) at University of Wisconsin-Whitewater. He received his M.S. in Industrial Engineering from University of Tehran and his Ph.D. in Management Science and Information Systems from Oklahoma State University. His research interests include health analytics, business analytics, IT privacy, and technology addition. He is a member of the Association for Information Systems, The Institute for Operations Research and the Management Sciences, and the Decision Sciences Institute. Behrooz has published in journals such as *Journal of the American Medical Informatics Association (JAMIA)*, *Information & Management*, *Journal of Business Research*, *International Journal of Medical Informatics*, and *Expert Systems with Applications*, as well as IS proceedings such as the *Hawaii International Conference on System Sciences (HICSS)* and the *International Conference on Information Systems (ICIS)*.



**Hamed M. Zolbanin** is an assistant professor of information systems at the University of Dayton. Prior to this position, he served as the director of the business analytics program at Ball State University. He had several years of professional experience as an IT engineer prior to receiving his Ph.D. in Management Science and Information Systems from Oklahoma State University. His research has been published in such journals as *Decision Support Systems*, *Information & Management*, *Information Systems Frontiers*, *Communications of the Association for Information Systems*, and the *Journal of Business Research*. His main research interests are healthcare analytics, online reviews, sharing economy, and digital entrepreneurship.



**Dursun Delen** is the holder of Spears and Patterson Endowed Chairs in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University. He has authored/co-authored more than 120 journal papers and numerous peer-reviewed conference proceeding articles. His research has appeared in major journals including *Decision Sciences*, *Journal of Production Operations Management*, *Decision Support Systems*, *Communications of the ACM*, *Computers and Operations Research*, *Computers in Industry*, *Journal of the American Medical Informatics Association*, *Artificial Intelligence in Medicine*, *International Journal of Medical Informatics*, *Health Informatics Journal*, among others. He has published 11 books/textbooks in the broad area of Business Intelligence and Business Analytics. He is often invited to national and international conferences and symposiums for keynote addresses, and companies and government agencies for consultancy/education projects on data science and analytics related topics. Dr. Delen served as the general co-chair for the 4th International Conference on Network Computing and Advanced Information Management (held in Seoul, South Korea), and regularly chairs tracks and minitracks at various information systems and analytics conferences. He is currently serving as the editor-in-chief, senior editor, associate editor and editorial board member of more than a dozen academic journals.