Predicting ICU Admissions for Hospitalized COVID-19 Patients with a Factor Graph-based Model

Yurui Cao, Phuong Cao, Haotian Chen, Karl M. Kochendorfer, Andrew B. Trotter, William L. Galanter, Paul M. Arnold, and Ravishankar K. Iyer

Abstract This paper presents a factor graph-based model that takes comorbidities and clinical measurements as inputs and predicts intensive care unit (ICU) admissions 3 days and 7 days in advance for hospitalized COVID-19 patients. We applied the proposed model on a COVID-19 cohort from a large medical center in Chicago (with records from March 2020 to August 2021). We used the first occurrence of the Delta variant in the U.S., February 2021, as the threshold to divide the dataset into pre-Delta data (533 patients) and post-Delta data (56 patients). Our model demonstrated 0.82 AUC on the pre-Delta data and 0.87 AUC on the post-Delta data in 7-day predictions. Our contribution is a model that (i) explains relationships between different clinical features and provides interpretations for ICU admissions, (ii) outperforms existing methods for 7-day predictions, and (iii) maintains more robustness than existing models in predictions under the influence of the Delta variant. The proposed model

Y. Cao (✉) · P. Cao · H. Chen · R. K. Iyer
University of Illinois Urbana-Champaign, Champaign, IL, USA
e-mail: yuruic2@illinois.edu

P. Cao
e-mail: pcao3@illinois.edu

H. Chen
e-mail: hc19@illinois.edu

R. K. Iyer
e-mail: rkiyer@illinois.edu

K. M. Kochendorfer · A. B. Trotter · W. L. Galanter
University of Illinois Hospital & Health Sciences System, Chicago, IL, USA
e-mail: kkoche1@uic.edu

A. B. Trotter
e-mail: trottera@uic.edu

W. L. Galanter
e-mail: billg@uic.edu

P. M. Arnold
Carle Foundation Hospital, Urbana, IL, USA
e-mail: paul.arnold@carle.com

A. Shaban-Nejad et al. (eds.), *Multimodal AI in Healthcare*, Studies in Computational Intelligence 1060, https://doi.org/10.1007/978-3-031-14771-5_17

could be used as a predictive tool in clinical practice to help clinicians in decision-making by predicting which patients will need ICU support in the future.

Keywords COVID-19 prognosis · Predictive biomarkers · Comorbidities · Factor graph · Probabilistic graphical model

1 Introduction

The spread of COVID-19 has recently been influenced by vaccines [1], which have stymied the spread of the disease, and the Delta variant [2], which has an altered pathology progression and has become the dominant variant of the virus [3]. These changes have presented new challenges in predicting COVID-19 disease progression among hospitalized patients, and prediction systems that support clinical decision-making will be crucial for managing patients' health to reduce fatality rates. In this paper, we present a probabilistic graphical model (PGM) for inferring ICU admissions in hospitalized COVID-19 patients. We make 3-day and 7-day predictions on ICU admissions based on a patient's comorbidities and clinical measurements (including laboratory tests and vitals). Specifically, the output labels are binary, indicating whether a patient will be admitted to the ICU or not.

Several prognostic models were developed during the early phases of the pandemic [4–7], including ones that used logistic regression, support vector machine, decision tree, and random forest approaches. Although these models performed well on the pre-Delta data, their performance on post-Delta data has not been evaluated. The altered pathology progression of the Delta variant is likely to affect the performance of models that were trained on pre-Delta data significantly. Moreover, the data available from the post-Delta period are limited, which poses additional challenges for training models with a large number of parameters.

To address those challenges, we developed a factor graph (FG) model [8], a type of PGM that has found success in a variety of applications, such as clinical diagnosis [9, 10] and cyber-security [11, 12]. FGs allow us to make predictions based on longitudinal data. Central to the FG model are factor functions (FFs), which are mathematical formulae that encode the relationships among clinical measurements, previous ICU admission status, and future ICU admission status. These relationships are inspired by domain knowledge and learned from statistical analysis, which can be reliably done with limited training data. In addition, FGs are more interpretable due to their graphical representation, which is beneficial in a clinical setting.

We evaluated our model with data from 589 patients hospitalized at the University of Illinois Hospital, a large academic hospital in Chicago. The data contain electronic health records (EHR) from March 2020 to August 2021, with 533 patients in the pre-Delta data and 56 patients in the post-Delta data. To demonstrate our model's performance in predicting ICU admissions and the model's robustness against the Delta variant, we evaluated the model on both the pre-Delta data and post-Delta data.

Because of the limited size of the post-Delta data, we trained the model only on the pre-Delta data. The key results are as follows:

- Our FG model outperforms state-of-the-art methods in predicting ICU admissions 7 days in advance (with 0.82 AUC on pre-Delta data, and 0.87 AUC on post-Delta data) and has comparable performance on 3-day predictions (with 0.81 AUC on pre-Delta data, and 0.73 AUC on post-Delta data).
- The proposed model's performance is more robust than the state-of-the-art methods to altered pathology progression in the post-Delta period. While the AUC and accuracy of most competing methods drop substantially on the post-Delta data, our model's AUC and accuracy changes for post-Delta data stabilize within small ranges (9.9%–13.5%).
- The FFs in our model explained the relationship between greater severity of comorbidities and higher risk of ICU admission for COVID-19 patients. Moreover, we also identified the change in predictive biomarkers for different prediction time windows.
- Our model can be used as a tool in clinical practice to suggest appropriate placement of patients, either in regular beds or the ICU. The flexibility of our model's FF constructions and its ability to work from limited training data allows it to be easily adapted for other diseases and new viruses with small data samples.

2 Model

In this section, we discuss how we constructed the FG model. We first provide an overview of the model's structure and define the variables. Then, we explain the methods used for variable selection and factor function construction. Finally, we explain how we perform prediction with the inference algorithms.

2.1 Model Overview

The proposed model (see Fig. 1) builds on FGs to predict the ICU admissions of COVID-19 patients d days in advance ($d = 3$ and $d = 7$). An FG [8] is represented by $G = (V \cup F, \mathcal{E})$, where $V = \{v_1, ..., v_n\}$ are the variable nodes, $F = \{f_1, ..., f_m\}$ are the factor nodes, and $\mathcal{E} = \{(v_p, f_q) | v_p \in V_q, f_q \in F\}$ are the edges connecting each factor f_q to its neighbors $v_p \in V_q$, where $V_q \subseteq V$. The FG is a bipartite graph between variables V and factors F, where the factor nodes $f_q \in F$ represent FFs and are non-negative. While the values of some nodes $E \subseteq V$ are observed (or provided in the data), the values of other nodes $S \subseteq V$ are hidden (and need to be inferred), where $V = E \cup S$.

When the above FG model is applied to ICU admissions for hospitalized COVID-19 patients, the observed events are $E = E_C \cup E_L$ (where $E_C = \{e_{c_i} | c_i \in C\}$ are

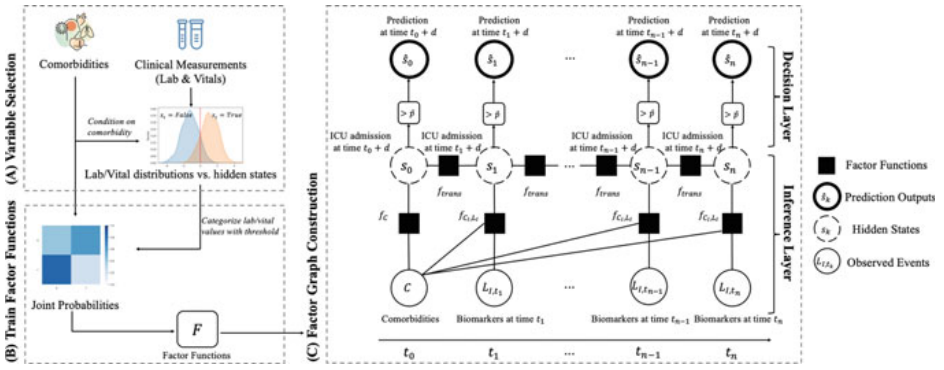

Fig. 1 Approach Overview. **a** For each comorbidity, lab/vital distributions are plotted with respect to the label (ICU/No ICU). The thresholds that best separate the distribution are computed. **b** Thresholds are used to categorize the lab/vital values and calculate the joint probabilities to construct FFs. **(C)** FGs are constructed by connecting FFs with observed events and hidden states. Predictions are made based on a threshold for the hidden state probabilities learned from the training data. $C = \{c_1, ..., c_6\}$ is a set of comorbidity nodes; $L_{I,t_k} = \bigcup_{i \in I} L_{i,t_k}$ is a union of the sets of predictive biomarkers L_i for the patient's comorbidities $c_i \in C_I$; $f_C = \{f_{c_1}, ..., f_{c_6}\}$ is a set of Comorbidity FFs; $f_{C_I,L_I} = \bigcup_{i \in I} (\bigcup_{l_j \in L_i} f_{c_i,l_j})$ is a union of the Bio FFs

comorbidities and $E_L = \{e_{l_j,t} | lj \in L, t \in T_n\}$ are lab/vital measurements), and the hidden states $S = \{s_t | t \in T_n\}$ are the future ICU admissions d days in advance. $T_n = \{t_1, ..., t_n\}$ are the timestamps of the observed events. (The intervals in this model have the granularity of a single day because labs/vitals are drawn on a daily basis.) The probabilities of the hidden states are estimated using the belief propagation (BP) algorithm [13] with a factorized joint probability distribution of E and S:

$$P(E, S) = \frac{1}{Z} \prod_a f_a(V_a), \tag{1}$$

where $V_a = (E_a \cup S_a)$, $V_a \subseteq E \cup S$, and $E \cup S = \bigcup_a V_a$. Z is a constant to normalize the product of factor functions and ensure that P is a valid probability distribution. While the BP algorithms only compute the probability of each hidden state, a threshold $\hat{p}$ was learned from the training data to achieve the highest accuracy for the prediction task. If the probability of ICU admission is greater than $\hat{p}$, then the future state is predicted to be ICU admission. Otherwise, it is predicted to be No ICU admission. The threshold $\hat{p}$ is the best value that separates the estimated posterior probability distributions of both the ICU class and the No ICU class, evaluated by the F1 score.

2.2 Variable Selection

The input features of the model consist of comorbidities, laboratory tests, and vitals. We chose 6 comorbidities that are common in the U.S. [14] as the inputs in our FG model, including hypertension, chronic obstructive pulmonary disease (COPD), type 2 diabetes, renal disease, heart failure, and obesity, represented as $C = \{c_1, ..., c_6\}$.

We converted lab/vital values into binary values to facilitate probabilistic model construction (see Fig. 1a). More specifically, we plotted the distributions of labs/vitals across the hidden states (ICU/No ICU) and found the threshold that best distinguishes the two distributions (that of the ICU states, and that of the No ICU states), when the F1 score is used as the evaluation metric. We then converted the lab/vital values to binary depending on whether their values were higher or lower than the thresholds. We introduced the influence of comorbidities on ICU admission by analyzing the lab/vital values conditioned on the comorbidity (i.e., we studied the lab/vital distributions for each comorbidity, and also for patients without comorbidities). For the purpose of dimensionality reduction, for each comorbidity, we selected only 3 labs/vitals to use in constructing the FGs; we called them *predictive biomarkers*. This process was developed in consultation with our clinical co-authors. The predictive biomarkers were the 3 labs/vitals with the highest F1 scores for each type of comorbidity; the 3 labs/vitals with the highest F1 scores were also selected for patients without comorbidities.

2.3 Factor Function Construction

We designed 3 kinds of FFs to explain the relationship among comorbidities, laboratory tests, vital measurements, and ICU admission after d days, including (i) the Transition FF, f_{trans}; (ii) the Comorbidity FFs, f_{c_i}; and (iii) the Bio FFs, f_{c_i,l_j}. The mathematical formulae for calculating the FFs are listed in Table 1.

Transition FF. The Transition FF $f_{trans}(s_{t_k}, s_{t_{k+1}})$ measures the probability that hidden state s_{t_k} will transition to hidden state $s_{t_{k+1}}$, $P(s_{t_k}, s_{t_{k+1}})$.

Comorbidity FFs. The Comorbidity FFs $f_{c_i}(e_{c_i}, s_{t_0})$ capture the joint probabilities of the patients' comorbidities and their initial hidden state s_{t_0}, $P(e_{c_i}, s_{t_0})$.

Bio FFs. To incorporate the impact of comorbidities in each step of inference, we designed the Bio FFs $f_{c_i,l_j}(e_{l_j,t}, s_t)$ to capture the joint probabilities of the patients' lab/vital values and their hidden state at time t given their comorbidities, $P(e_{l_j,t}, s_t|e_{c_i})$. $l_j \in L_i$. $e_{l_j,t}$ are the binary lab/vital values categorized using the thresholds chosen in Sect. 2.2. The FFs of the predictive biomarkers are independent of each other, so the model can tolerate missing values for some of the predictive biomarkers when inferring the hidden states.

Table 1 The formulas to construct factor functions (FFs)

FF(s)	Equation	Comments
Trans.	$f_{trans}(s_a, s_b) = \frac{\sum_{\Phi, t \in T_{n-1}} \mathbb{1}_{\{s_{t_k}=s_a, s_{t_{k+1}}=s_b\}}}{\sum_{\Phi, t_{k'} \in T_{n-1}, s'_a, s'_b \in B} \mathbb{1}_{\{s_{t_{k'}}=s'_a, s_{t_{k'+1}}=s'_b\}}}$	Each entry is the normalized frequency of a type of transition
Comorb.	$f_{c_i}(c, s) = \frac{\sum_{\Phi} \mathbb{1}_{\{e_{c_i}=c, s_{t_0}=s\}}}{\sum_{\Phi, c' \in B, s' \in B} \mathbb{1}_{\{e_{c_i}=c', s_{t_0}=s'\}}}$	Each entry is the joint probability that the patient would (or would not) have comorbidity c_i with initial hidden state s
Bio	$f_{c_i,l_j}(l, s) = \frac{\sum_{\Phi_{c_i}, t \in T_{n-1}} \mathbb{1}_{\{e_{l_j,t}=l, s_t=s\}}}{\sum_{\Phi_{c_i}, t \in T_{n-1}, l', s' \in B} \mathbb{1}_{\{e_{l_j,t}=l', s_t=s'\}}}$	Each entry is the joint probability that patients with comorbidity c_i will have labs/vitals l_j above or below the selected threshold with hidden state s

∗ Abbreviations: Trans., Transition; Comorb., Comorbidity; Φ, all patients; Φ_{c_i}, all patients with comorbidity c_i; $T_{n-1} = \{t_0, ..., t_{n-1}\}$; $B = \{True, False\}$. Constraints: $c, l, s, s_a, s_b \in B$

2.4 Inference Algorithms

Each patient's comorbidity information was first collected upon his or her admission to the hospital. Thus, the comorbidity nodes were first added as the initial events e_{c_i}, and a corresponding initial hidden state s_{t_0} was inferred with the Comorbidity FFs f_{c_i}. Then, during the patient's stay in the hospital, his/her labs/vitals were measured. Depending on the comorbidities $C_I \subseteq C$ he/she had, measurements of the corresponding predictive biomarkers $l_j \in L_i$, where $i \in I$, were added to the FG as events e_{l_j,t_k}. A corresponding hidden state s_{t_k} was also added. The Bio FFs f_{c_i,l_j} were connected to events e_{l_j,t_k}, e_{c_i} and to the hidden state s_{t_k}. The Transition FF f_{trans} was connected to the previous hidden state $s_{t_{k-1}}$ and the new hidden state s_{t_k}. The probabilities of the hidden states S, denoted by $P_S = \{p_t | t \in T_n\}$, were computed using the BP algorithm, for which p_t is the probability that hidden state s_t is ICU. When p_t is greater than $\hat{p}$, the hidden state is predicted to be ICU. Otherwise, it is predicted to be No ICU.

3 Experimental Setup

In this section, we describe our experimental setup and how we evaluated the proposed model, including the cohort characteristics and the evaluation methods.

Table 2 Demographics and comorbidities of hospitalized COVID-19 patients

		Pre-delta		Post-delta	
		$d = 3$	$d = 7$	$d = 3$	$d = 7$
Total*, N		364	175	24	11
Age, $median(IQR)$		57 (45–68)	60 (49–71)	58 (44–66)	60 (46–66)
Sex, $N(\%)$	Female	193 (53.0%)	84 (48.0%)	13 (54.1%)	6 (54.6%)
	Male	171 (47.0%)	91 (52.0%)	11 (45.8%)	5 (45.5%)
Comorbidity, $N(\%)$	N/A	229 (62.9%)	107 (61.1%)	11 (45.8%)	6 (54.5%)
	Hypertension	80 (22.0%)	40 (22.9%)	8 (33.3%)	3 (27.3%)
	COPD	12 (3.3%)	4 (2.3%)	0 (0.0%)	0 (0.0%)
	Type 2 diabetes	66 (18.1%)	32 (18.3%)	6 (25.0%)	4 (36.4%)
	Renal disease	35 (9.6%)	19 (10.9%)	5 (20.8%)	2 (18.2%)
	Heart failure	12 (3.3%)	5 (2.9%)	1 (4.2%)	0 (0.0%)
	Obesity	13 (3.6%)	6 (3.4%)	0 (0.0%)	0 (0.0%)

*Only patients with at least one lab/vital measurement were included in our analysis. Hence, our dataset was reduced to 388 patients in total. Depending on the prediction tasks ($d = 3$ or $d = 7$), the number of valid patients in each dataset varies. Abbreviations: N/A, no comorbidity; COPD, chronic obstructive pulmonary disease

3.1 Dataset

We validated the proposed model on a COVID-19 dataset collected from the University of Illinois Hospital in Chicago, Illinois. This dataset contains EHRs of 589 patients who were hospitalized with COVID-19 between March 2020 and August 2021, with 533 patients in the pre-Delta data and 56 in the post-Delta data. Demopgraphics and comorbidity information of the patients are listed in Table 2. When the data contain multiple measurements of the same lab/vital within the same day, we used only the latest lab/vital value for that day. Since we want to predict the patients' ICU admissions d days in advance ($d = 3$ and $d = 7$), we treated future ICU admission status as a hidden state to be inferred using current clinical measurements.

3.2 Model Evaluation

We evaluated the model performance with an 80:20 split into training and testing sets on the pre-Delta data. Since the events and states are time-dependent, it would not have made sense to randomly split the data; doing so might result in use of future data points to train the model and predict data points in the past. Thus, the training set consists of the records of the first 80% of the patients admitted during the pre-Delta

period, and the testing set consists of the records of the remaining 20% of the pre-Delta-period patients. We performed variable selection and trained the FFs as well as the threshold $\hat{p}$ on the training set. The accuracy of ICU admission predictions was evaluated on both the testing set and the post-Delta data, using AUC and accuracy as the metrics. We compared the prediction performance of the proposed framework with that of other state-of-the-art methods, which used logistic regression, support vector machine, decision tree, and random forest approaches [4–6]. Furthermore, to evaluate the value of using previous ICU admission status to predict future ICU admission status, we assessed the proposed model without the Transition FFs, i.e., the case when the previous hidden state does not share an edge with the current hidden state (w/o Tr.).

4 Results

In this section, we discuss the experimental results for the FG model in making 3-day and 7-day predictions on ICU admissions of hospitalized COVID-19 patients. We provide details on the predictive biomarkers of each comorbidity and their thresholds, and we explain the model performance on different prediction tasks.

4.1 Predictive Biomarkers

We selected 3 predictive biomarkers for each group, including the group without comorbidities, and the groups with each comorbidity (see Table 3). The dataset for 7-day prediction does not contain enough patients ($N < 10$) with the comorbidities of COPD, heart failure, and obesity, so we merged those patients into the group without comorbidities. From the results, we found that some of the predictive biomarkers are indicators of the severity of the comorbidities. Moreover, we found that the predictive biomarkers differ between the 3-day and 7-day predictions. These findings can help us better understand the critical factors for predicting ICU admission of hospitalized COVID-19 patients and design the factor functions using domain knowledge.

4.2 Model Validation

We validated the ability of our model to infer ICU admissions for hospitalized COVID-19 patients after d days ($d = 3$ and $d = 7$) given their comorbidities, laboratory tests, and vital measurements. The model's performance was compared with that of several state-of-the-art models (see Table 4). Compared to the existing methods, our model achieved the best performance for 7-day predictions and comparable performance on 3-day predictions. Moreover, the proposed model outperformed the FG

Table 3 Predictive biomarkers for ICU admission given the comorbidities

Comorbidity (c_i)	$d = 3$			$d = 7$		
	N	Labs/Vitals (L_i)	ICU admission	N	Labs/Vitals (L_i)	ICU admission
N/A	229	ALB %LYMPH RBC	≤3.4 g/dL ≤15.0% ≤4.2 M/uL	107	CRP ALB AST	> 32.4 mg/L ≤ 3.5 g/dL > 23.0 U/L
Hypertension	80	ALB WBC RBC	≤3.4 g/dL > 5.8 K/uL ≤4.6 M/uL	40	%LYMPH WBC RBC	≤34.0% > 3.3 K/uL ≤ 4.6 M/uL
COPD	12	CRP WBC %MONO	> 19.8 mg/L > 6.0 K/uL ≤11.1%	4	– – –	– – –
Type 2 diabetes	66	RBC CRP HGB	≤4.6 M/uL > 129.6 mg/L ≤11.6 g/d	32	CRP WBC ALB	> 55.8 mg/L > 5.0 K/uL ≤3.5 g/dL
Renal disease	35	WBC ALB %MONO	> 6.6 K/uL ≤ 3.4 g/dL ≤8.3%	19	WBC AST ALB	> 4.1 K/uL > 16.0 U/L ≤3.5 g/dL
Heart failure	12	HGB RBC WBC	≤11.6 g/dL ≤3.9 M/uL > 7.1	5	– – –	– – –
Obesity	13	ALB %LYMPH BUN	≤3.8 g/dL ≤17.5% > 16.1 mg/dL	6	– – –	– – –

∗Abbreviations: N/A, no comorbidity; COPD, chronic obstructive pulmonary disease; ALB, albumin; AST, aspartate aminotransferase; BUN, blood urea nitrogen; CRP, C-reactive protein; HGB, hemoglobin; %LYMPH, lymphocyte percentage; %MONO, monocytes percentage; RBC, red blood cell; WBC, white blood cell

model without Transition FF, demonstrating that a combination of past and current observations improves prediction performance. Our model also demonstrated better robustness in prediction for data affected by the Delta variant; it had a relatively small decrease (9.9% to 13.5%) in performance on the post-Delta data relative to pre-Delta data, while most of the other methods (listed in Table 4) showed a large decline. Nevertheless, the contribution of the post-Delta data versus the pre-Delta data requires a deeper level of analysis than this paper provides. Specifically, we believe a molecular-level analysis is necessary [15, 16].

Summary. The success of our model is due to its ability to integrate past state/observations and current observations to make predictions. The decrease in prediction performance of the proposed model without the Transition FF demonstrates the importance of temporal information in prediction. The existing state-of-the-art methods do not take into account temporal variables and have lower prediction accuracy.

Table 4 Prediction performance. Changes in performance w.r.t. pre-Delta period are provided in parentheses

Method	Pre-delta				Post-delta			
	$d = 3$		$d = 7$		$d = 3$		$d = 7$	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
Proposed	0.81	0.74	0.82	**0.75**	**0.73 (-9.9%)**	**0.64 (-13.5%)**	**0.87 (+6.1%)**	0.66 (−12.0%)
Proposed w/o Tr.	0.76	0.68	0.69	**0.75**	0.67 (−11.8%)	0.62 (−8.8%)	0.67 (−2.9%)	0.62 (−17.3%)
LR	0.85	0.77	**0.83**	0.74	0.62 (−27.1%)	0.57 (−26.0%)	0.57 (−31.3%)	0.57 (−23.0%)
SVM	0.86	**0.80**	0.82	0.72	0.66 (−23.3%)	0.58 (−27.5%)	0.57 (−30.5%)	0.50 (−30.6%)
Random Forest	**0.88**	0.79	0.69	0.70	0.70 (−20.5%)	0.50 (−36.7%)	0.73 (+5.8%)	0.57 (−18.6%)
Decision Tree	0.63	0.63	0.54	0.65	0.60 (−4.8%)	0.53 (−15.9%)	0.71 (+9.2%)	**0.67 (+3.1%)**

∗The best AUC/ACC in each column is in bold. Abbreviations: Tr., Transition factor function; AUC, area under the ROC curve; ACC, accuracy; LR, logistic regression; SVM, support vector machine

5 Limitations and Future Work

One limitation of our work is that we only constructed the factor functions with joint probabilities supported by statistical analysis and have not experimented with more sophisticated factor functions. For example, we could include domain knowledge from clinical experts to construct multivariate factor functions that better explain the relationships between the variables. Second, the proposed model only considers current lab/vital measurements, not patterns/trends in past measurements, when making predictions.

Future work will extend the model to capture the temporal trend, i.e., the rate of change in the lab/vital measurements, of the events. In addition, performing hyperparameter tuning to emphasize different weights among the factor functions may improve the prediction performance and demonstrate the contributions of different factors. The model will be implemented as a toolset to provide advice to triaging physicians. Subsequent work will incorporate the impact of vaccines and emerging mutations automatically as a learning paradigm.

6 Conclusion

We proposed a factor graph-based framework that predicts ICU admissions of hospitalized COVID-19 patients d days in advance. Our model demonstrates comparable and better performance than the state-of-the-art machine learning methods on 3-day

and 7-day predictions, respectively. The relationships between comorbidities and labs/vitals captured by the model shed light on understanding ICU admissions for COVID-19, for which greater severity of comorbidities introduces a higher risk of ICU admissions. Most importantly, the model's prediction performance is robust for the post-Delta data.

Acknowledgements This work was partly supported by the Center for Computational Biotechnology and Genomic Medicine, Carle Foundation Hospital, and the Jump ARCHES endowment fund. We thank Yufu Zhang, Jorge Rodriguez Fernandez, and Jai Nebhrajani for providing and interpreting the data. We also thank our colleagues in the DEPEND group, particularly Krishnakant Saboo, Chang Hu, Anirudh Choudhary, Mosbah Aouad, Yixin Chen, Kathleen Atchley, and Jenny Applequist for their valuable feedback. The project was also supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number UL1TR002003. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Vitiello, A., Ferrara, F., Troiano, V., & La Porta, R. (2021). COVID-19 vaccines and decreased transmission of SARS-CoV-2. *Inflammopharmacology*. https://doi.org/10.1007/s10787-021-00847-2
2. Dougherty, K., Mannell, M., Naqvi, O., Matson, D., & Stone, J. (2021). SARS-CoV-2 B.1.617.2 (Delta) variant COVID-19 outbreak associated with a gymnastics facility. MMWR Morb Mortal Wkly Rep. https://doi.org/10.15585/mmwr.mm7028e2external.
3. Mlcochova, P., Kemp, S., Dhar, M., et al. (2020). SARS-CoV-2 B. 1.617.2 delta variant replication and immune evasion. *Nature, 599,* 114–119.
4. Galanter, W., Rodríguez-Fernández, J., Chow, K., et al. (2021). Predicting clinical outcomes among hospitalized COVID-19 patients using both local and published models. *BMC Medical Informatics and Decision Making*. https://doi.org/10.1186/s12911-021-01576-w
5. Zhang, J., Jun, T., Frank, J., et al. (2021). Prediction of individual COVID-19 diagnosis using baseline demographics and lab data. *Scientific Reports*. https://doi.org/10.1038/s41598-021-93126-7
6. Jiang, X., Coffee, M., Bari, A., et al. (2020). Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials and Continua*. https://doi.org/10.32604/cmc.2020.010691
7. Iwendi, C., Bashir, A., Peshkar, A., et al. (2020). COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in Public Health*. https://doi.org/10.3389/fpubh.2020.00357
8. Loeliger, H. (2004). An introduction to factor graphs. *IEEE Signal Processing Magazine*. https://doi.org/10.1109/MSP.2004.1267047
9. Varatharajah, Y., Chong, M., Saboo, K., et al. (2017). EEG-GRAPH: A factor-graph-based model for capturing spatial, temporal, and observational relationships in electroencephalograms. In: NeurIPS (pp. 5372–5381).
10. Yang, Y., Walter, L., Lu, L., et al. (2014). Forecasting potential diabetes complications. In: *Proceedings of 28th AAAI Conference on Artificial Intelligence* (pp. 313–319).
11. Cao, P., Badger, E., Kalbarczyk, Z., Iyer, R., & Slagell, A. (2015). Preemptive intrusion detection: Theoretical framework and real-world measurements. In: *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*. https://doi.org/10.1145/2746194.2746199.
12. Cao, P. (2019). On preempting advanced persistent threats Using probabilistic graphical models. arXiv:1903.08826 [cs.CR].

13. Yedidia, J., Freeman, W., & Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*. https://doi.org/10.1109/TIT.2005.850085
14. Harrison, S., Fazio-Eynullayeva, E., Lane, D., et al. (2020). Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis. *PLOS Medicine*. https://doi.org/10.1371/journal.pmed.1003321
15. McCallum, M., Walls, A., Sprouse, K., et al. (2021). Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. *Science*. https://doi.org/10.1126/science.abl8506
16. Ghosh, A., Kaiser, M., Molla, M., et al. (2021). Molecular and serological characterization of the SARS-CoV-2 Delta variant in Bangladesh in 2021. *Viruses*. https://doi.org/10.3390/v13112310