Early Mortality Prediction for Mechanically Ventilated Patients Using Graph Neural Networks

Daniela Chanci Arrubla
dchanci@emory.edu
Emory University
Atlanta, Georgia, USA

ABSTRACT

Critically ill patients with respiratory failure, or severe complications to breathe, require mechanical ventilation to survive. However, such life support method is associated with high mortality. Therefore, in this work, we proposed the development of a GNN-based model to accurately predict the mortality of mechanically ventilated patients based on EHR data. More specifically, we extracted and preprocessed data from the publicly available MIMIC-III database, calculated the pairwise patient similarity to obtain the edges connecting similar patients (above 0.9 similarity), constructed a graph network, trained three different models: GraphSAGE, GCN, and GAT, obtaining accuracies of 62%, 65%, and 66%, respectively. These experimental results act as a first step towards the creation of a more robust and complex model for the early mortality prediction in mechanically ventilated patients.

KEYWORDS

Mechanical Ventilation, Graph Neural Networks, Electronic Health Records, Mortality.

1 INTRODUCTION

Mechanical ventilation is an invasive form of life support used in the intensive care unit (ICU) for acutely ill patients with failure of the respiratory function, or for surgical patients [9]. In this regard, the mechanical ventilator moves high concentrations of oxygen into the lungs, and removes the carbon dioxide from the body [1]. Prior studies have shown that, annually, approximately 20 million patients worldwide require mechanical ventilation. Moreover, this number has been increasing in the last years due to the aging population and to the devastating impact of COVID-19. Consequently, several lung protective mechanical ventilation strategies have been proposed and implemented; however, this form of life support is still associated with high morbidity and mortality rates [8]. Therefore, the early assessment of mortality risk for mechanically ventilated patients is crucial in critical care. Such assessment could have a positive impact in the medical decision-making process, resulting in the adjustment of treatments or strategies for patients with high mortality risk, or the prevention of unnecessary prolonged mechanical ventilation events [7].

Nowadays, electronic health records (EHR) store large databases with valuable patient medical data. A widely known publicly available database is the Medical Information Mart for Intensive Care III (MIMIC-III) database, which contains deidentified information of patients admitted to the Beth Israel Deaconess Medical Center [3]. Said database combined with the use of novel Artificial Intelligence (AI) models could address the mortality prediction task for mechanically ventilated patients. Recently, some studies used widely known Machine Learning models for this purpose. However, they disregarded the relationships that exist among patients with a high similarity percentage, or among different clinical events. In this work, we focus towards the implementation of a Graph Neural Network based (GNN) model for the mortality prediction task. To the best of our knowledge, there are no studies yet using GNN-based models. Additionally, a graph is a suitable representation for EHR data in this specific task given that it not only uses the patient features but leverages the possible existing relationships among them, which could lead to the obtention of satisfactory results [5]. The rest of the document presents the related work, problem formulation, methods, experimental results, discussion, future work, and conclusions.

2 RELATED WORK

This section presents an overview of the past and state-of-the-art research on the areas related to the proposed work, namely mortality prediction for mechanically ventilated patients, and application of GNN-based models to the MIMIC-III database.

Considering the continuous growing interest in personalized medicine, there have been several studies and research groups focused towards the development of mortality prediction models. However, most of them study other clinical scenarios, such as sepsis patients, or heart failure patients. Although mechanical ventilation is associated with mortality rates above 50%, there is a lack of studies in this topic. In the literature we can find three studies that focus towards the mortality prediction in mechanically ventilated patients, all of them published during the last year, which indicates that the interest in this topic might grow in the near future.

In this regard, Kim et al. collected data from five hospitals in the Republic of Korea for adult patients who went through mechanical ventilation between the years 2010 and 2019. They used a supervised approach to train five Machine Learning algorithms: balanced random forest (BRF), light gradient boosting machine (LGBM), Extreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), and logistic regression; obtaining an accuracy of 76% for the 30-day mortality prediction with the XGBoost model [4]. Similarly, George et al. trained a feedforward neural network with regularization strategies to predict 3-month and 9-mont mortality on a retrospective cohort study obtained from MIMIC-III for adult patients with acute respiratory failure that needed mechanical ventilation therapy for at least 7 days, obtaining an accuracy of 69%

Data Mining - CS570, Spring 2022, .

ACM ISBN . . . $

for both studied outputs [2]. Lastly, for the mechanical ventilation mortality prediction task, Zhu et al. adressed the in-hospital mortality prediction. They trained 7 different models, namely k-nearest neighbors (KNN), logistic regression, bagging, decision tree, random forest, XGBoost, and a neural network; using a retrospective cohort obtained from the MIMIC-III databaset. As Kim et al., they obtained the best results for the XGBoost model [11].

Additionally, there are some studies using GNN-based models with the MIMIC-III dataset to address different biomedical applications that could be beneficial for the patients or for healthcare personal in general. For instance, a recent study published in 2021 by Liu et al. proposed a multi-relatonal hyperbolic GNN for the creation of a medical triage chatbot that can predict the diagnose with high accuracy based on the information provided by the patient. More specifically, they defined three types of nodes: user, symptom, and diagnosis; two type of edges: user-symptom and user-diagnosis; and used multi-relational aggregation and hyperbolic aggregation, obtaining and accuracy of approximately 63 % for the prediction of the diagnosis with top-10 rankings for the MIMIC-III dataset [6].

3 PROBLEM FORMULATION

In this work, we are interested in implementing a 30-day prediction mortality model for adult patients treated with mechanical ventilation for at least 24 hours. Given a dataset of N patients with M features, represented as $\left\{\left\{F_{ij}\right\}_{j=1}^{M}\right\}_{i=1}^{N}$, the objective is to create a graph (G) with N nodes (V) using the patient similarity measures to analyze the relationships among the patients, define the edges (E) in the graph, and train a GNN model to predict the mortality, which is a binary output $O = \{1, 0\}$.

3.1 Preliminaries

A graph G is a nonlinear data structure composed by a set of nodes V and a set of edges or links E. Mathematically, it is defined as a tuple $G = (V, E)$, and using a GNNs we can address different tasks, such as node classification, edge classification, and graph classification. In this work, we have a homogeneous graph with one type of nodes (patients) and one type of links (patient-patient). The main idea in GNN-based models for the node classification task is to update the embedding of each node in each layer of the model, following the message passing update rule of said layer, which in general consists in gathering the embedding representations of the neighboring nodes, aggregate them, and apply a suitable transformation.

3.2 Motivations and Challenges

There are three main motivations for this work:

(1) The lack of studies and research focused on mortality prediction for mechanically ventilated patients using AI techniques.
(2) Current methods for mortality prediction using traditional Machine Learning and Deep Learning models disregard the possible relationships that exist among patients.
(3) The use of a graph representation allows us to establish relationships among patients using clinical events and data contained in Electronic Health Records (EHR).

Additionally, there was one main challenge that we encountered when defining the methods and obtaining the experimental results: the patient similarity. It was a critical step in this work given that the edges among the different nodes were created based on this similarity measure. In the literature it is possible to find several supervised, unsupervised, and semi-supervised approaches to define the similarity between two patients using healthcare data, often extracted from EHR, that lead to different results. It was also difficult to defined the threshold that divides similar patients from non-similar patients. After a careful revision of the literature I selected the method that will be explained in the following section given that it was the most suitable for the data and the task.

4 METHODS

In this section, the methods used for the mortality prediction of mechanically ventilated patients are discussed.

4.1 Database, Study Design, and Feature Extraction

The critical care database MIMIC-III was used. It contains information of 61,051 ICU admissions for patients ($age \geqslant 16$) admitted to critical care unit in the Beth Israel Deaconess Medical Center in Boston between 2001 and 2012. This database was obtained from Physionet after the obtention of a required certificate. For this work we defined the following inclusion criteria:

- Patients between the age of 18 years old and 90 years old (inclusive).
- Mechanical ventilation events with a duration of at least 24 hours.
- Only the first mechanical ventilation event of an ICU stay in the case that multiple events occur during a single stay.

Additionally, after a careful review of the literature for mortality prediction applications, we selected the variables to compose the feature vector of each patient. Among these variables, there are static data, such as age and gender, and temporal data, such as vitals and laboratory tests. For the temporal data we only used the data extracted during the first 24 hours after the mechanical ventilation starting time. Specifically, for the vitals, we extracted the maximum value, the minimum value, and the mean value. For the laboratory tests, we extracted the maximum value and the minimum value. A list of all the extracted variables divided by category can be observed in Table 1.

4.2 Data Preprocessing

After extracting the data from MIMIC-III, some data preprocessing steps were implemented in the following order:

(1) Discard variables missing for more than 20% of the patients.
(2) Transform categorical variables, i.e., gender.
(3) Apply multiple imputation to replace the missing data using the Multivariate imputation by chained equations (MICE) method.
(4) Normalize the features to be in a range from 0 to 1.

However, for the patient similarity analysis, the data was not normalized.

Table 1: Selected variables from the MIMIC-III database

Category	Variables
Demographics	Age, gender
Hospital events	Hours from hospitalization to ICU, hours from ICU to mechanical ventilation event
Comorbidities	Congestive heart failure, hypertension, diabetes complicated, diabetes uncomplicated, metastatic cancer, solid tumor, peripheral vascular disease, hypothyroidism, liver disease, chronic pulmonary disease, lymphoma, coagulopathy
Elixhauser score	Elixhauser Vanwalraven score
Morbidity severity score	OASIS, LODS, SAPSII, SOFA
Vitals	Heart rate, systolic bloodpressure, diastolic blood pressure, mean arterial pressure, respiratory rate, temperature, oxygen saturation
Laboratory tests	Albumin, bands, bicarbonate, creatinine, chloride, glucose, hematocrit, hemoglobin, lactate, platelet count, potassium, sodium, blood urea nitrogen, and white blood cells count
Output	30-day mortality

4.3 Patient Similarity Analysis

The approach to establish the patient similarity is an unsupervised pairwise similarity approach proposed by Wang et al. in 2019 [10]. For this purpose we use the following data: age, gender, set of laboratory tests, and set of comorbidities. Therefore, for each of these data, a pairwise similarity measure was calculated, and then patient similarity measure was obtained using a weighted sum.

Let i and j represent the data corresponding to two different patients:

Feature similarity for age:

$$FSA\,(i,j) = \frac{min\,(age_i, age_j)}{max\,(age_i, age_j)} \tag{1}$$

Feature similarity for gender:

$$FSG\,(i,j) = \begin{cases} 1, & if\ gender\ is\ the\ same \\ 0, & otherwise \end{cases} \tag{2}$$

Feature similarity for laboratory tests:

$$FSL\,(i,j) = 1 - \frac{d_{lab}\,(i,j) - min\,(d_{lab})}{max\,(d_{lab}) - min\,(d_{lab})} \tag{3}$$

where

$$d_{lab}\,(i,j) = \sqrt{\sum_{k=1}^{k} \left(L_{ik} - L_{jk}\right)^2} \tag{4}$$

and k is the total of laboratory tests.

Feature similarity for comorbidities:

$$FSC\,(i,j) = 1 - normalized\ hamming\ distance\,\left(C_i, C_j\right) \tag{5}$$

where C *is the comorbidities vector.* Originally, in the approach proposed by Wang et al., the cosine distance is used for this measure. However, we implemented this change because the hamming distance was more suitable for our data.

Patient similarity:

$$PS\,(i,j) = 0.4{*}FSC(i,j) + 0.4{*}FSL(i,j) + 0.1{*}FSA(i,j) + 0.1{*}FSG(i,j) \tag{6}$$

with values closer to 1 indicating more similarity between the patients. Finally, the threshold was selected to be 0.9.

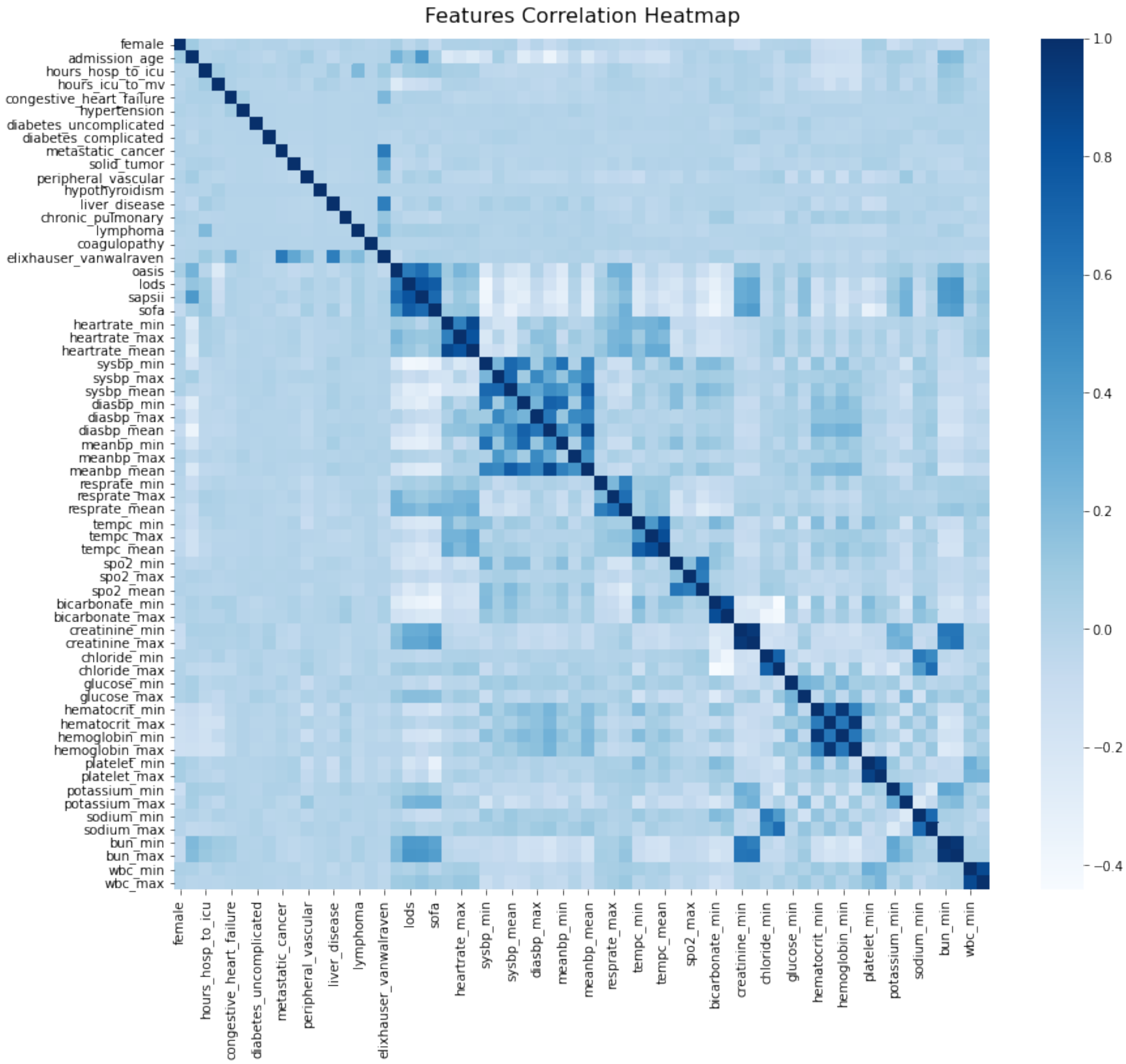

Figure 1: Correlation of the 64 extracted features for the 1000 patients cohort.

4.4 Graph Neural Network models

The dataset was divided into a training and testing set, 80% and 20% respectively, by creating the corresponding masks for the nodes. In addition, for the node classification task we used the following networks: GraphSAGE, Graph Attention Network (GAT), and Graph Convolutional Network (GCN). These models were implemented and trained in PyTorch Geometric, using two layers, and the Adam optimizer with a learning rate of 0.0001 for 1000 epochs. Finally, for the evaluation metric, we selected the accuracy.

5 EXPERIMENTAL RESULTS

This section presents the experiments conducted to evaluate the proposed approach for the mortality prediction problem.

5.1 Selected Cohort of Mechanically Ventilated Patients

The data extraction process was conducted using PostgreSQL. A total of 10183 patients of the MIMIC-III database met the inclusion criteria, and a total of 70 features was extracted for each of them. However, due to time constraints and computational requirements constraints, only 1000 random patients were selected for the subsequent steps. A summary of the most important characteristics of the patients prior to the normalization step can be observed in Table 2.

After inspecting the obtained data, following the inclusion criteria, 3 laboratory tests were excluded given that they were missing for more than 20% of the patients. These are albumin, bands, and lactate, which were missing for 52.25%, 81.91%, and 22.81%, respectively. Considering that for each of this variables we included the

Table 2: Cohort characteristics

Characteristic	Value
Age	Mean: 63.28, Std: 16.48
Gender	F: 37.2%, M: 62.8%
Elixhauser Score	Mean: 0.50, Std: 1.73
OASIS	Mean: 37.43, Std: 8.24
LODS	Mean: 43.57, Std: 14.13
SAPSII	Mean: 5.63, Std: 2.83
SOFA	Mean: 6.04, Std: 3.56

minimum value and the maximum value, as explained above, the features decreased from 70 to 64 per patient.

Finally, after having the final feature set, the correlation was plotted to inspect further the extracted variables (See Figure 1). It is possible to see that most of the features are not highly correlated to each other.

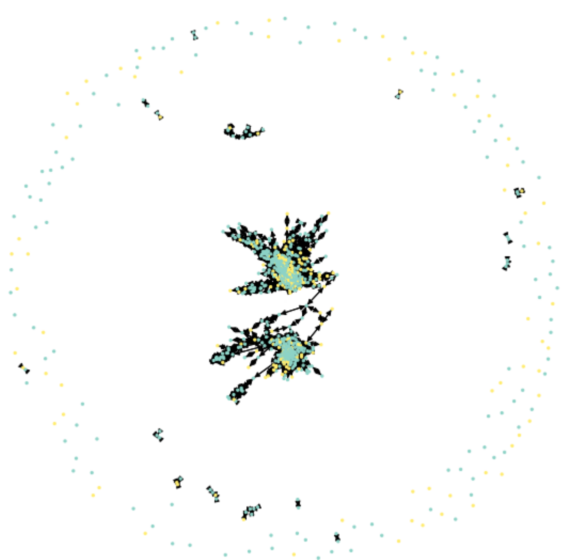

Figure 2: Graph created in PyTorch Geometric

5.2 Graph Network

Using PyTorch Geometric, we created an undirected homogeneous graph with one type of nodes, patients, an one type of links, patient-patient. A summary of the graph is:

- number of nodes: 1000
- dimension of the nodes representation: 64
- number of edges: 9742
- number of labels: 1000
- number of classes: 2

The number of edges represents the number of pair of patients for which the similarity score was above 0.9. In Figure 2, we can observe a visualization of the constructed graph. It is possible to identify some isolated nodes, as well as two main clusters of nodes densely connected. Additionally, the color of the nodes represent the class to which they belong.

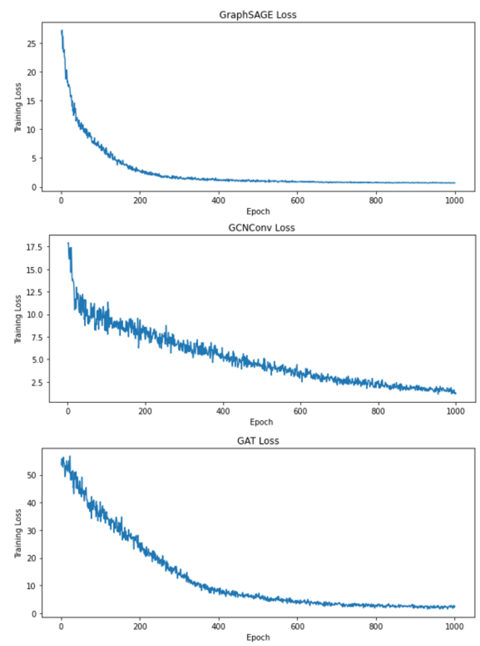

Figure 3: Loss curves for trained models

5.3 Mortality Prediction Models

Using the 1000 patients cohort extracted from the MIMIC-III database, we obtained 800 patients for training, and 200 patients for testing. The models were trained using PyTorch Geometric with the hyperparameters specified in the section above. For the GraphSAGE network the accuracy for the testing set was 0.62, for GCN it was 0.65, and for GAT it was 0.66. The results are summarized in Table 3 and the loss curves for the training process are depicted in Figure 3.

6 DISCUSSION

This project studies the use of GNN-based models for the mortality prediction task in mechanically ventilated patients. Despite the fact that our model did not achieve a higher accuracy than the more traditional Deep Learning and Machine Learning models discussed in the related work section, it is worth mentioning that our model was trained with far less data than theirs. We had access to the data; however, due to mainly time constraints, the dataset was reduced to a more manageable size which was approximately 1/10 of the originally extracted and cleaned dataset. Perhaps, this had a direct impact in the performance of the models. Another key difference between the models in the literature and the model proposed in this work is that we only used data from the first 24 hours after starting the mechanical ventilation event, while they use data for the entire duration of the mechanical ventilation event. The reasoning behind this choice can be found in the title of the project. Basically, we want an early prediction which is accurate using only data of the first 24 hours.

Table 3: Mortality Prediction Results

Model	Accuracy (%)
GraphSAGE	62%
GCN	65%
GAT	66%

Predicting the mortality of patients under any condition or disease is a critical problem in the healthcare field. We focused towards mechanically ventilated patients given the high number of patients needing this form of life support annually, and the high mortality rates associated to it. Having a model with high accuracy could be beneficial for both patients and practitioners, considering that it could help to make better decisions in the medical decision-making process.

Finally, it is worth mentioning that this results are preliminary. It would be worth to invest more time to be able to work with the full dataset, to train the models in a more careful manner, and also try to train more advanced GNN-based models rather than the most widely known used in this work. This would be feasible with the data that we have so far, considering that the data extraction and cleaning process was extremely time consuming, and it was done in a very thoughtful manner, it would be worth to keep working on the last steps of the method proposed in this document.

7 FUTURE DIRECTIONS

There are two main aspects to be considered for the future work. The study proposed in this work is a retrospective study. It would be more challenging, yet more useful for real-worl applications to conduct a prospective study instead. The other aspect to consider is related to the nature of the EHR data, and the ability of different graphs to accurately represent information. In this regard, the future work can focus in creating a representation of the EHR using a heterogeneous graph, where diagnoses, medications, vitals, labs, among other variables, are represented by different types of nodes. This would enable the identification of different types of relationships without the need of similarity measures calculation, which could be beneficial for the mortality prediction task.

8 CONCLUSION

In this work, a graph network representation of mechanically ventilated patients for early mortality prediction was constructed using EHR data extracted from the MIMIC-III database. The approach was constituted by three main steps: (1) The extraction of the features from the database, (2) the patient similarity analysis, and (3)the training of the GNNs. We demonstrated the ability of GNNs to deal with EHR data, obtaining results close to the ones found in the literature. The best results were obtained for the GAT model. However, further evaluation and validation steps (i.e., more evaluation metrics, and a statistical analysis) are needed.

9 DATA AVAILABILITY

The scripts and code used for this project will be uploaded to the GitHub repository: https://github.com/dchancia/MechVent_Mortality_Prediction.git

REFERENCES

[1] Mora Carpio AL and Jorge I Mora. 2017. Ventilator management. (2017).

[2] Naomi George, Edward Moseley, Rene Eber, Jennifer Siu, Mathew Samuel, Jonathan Yam, Kexin Huang, Leo Anthony Celi, and Charlotta Lindvall. 2021. Deep learning to predict long-term mortality in patients requiring 7 days of mechanical ventilation. *PloS one* 16, 6 (2021), e0253443.

[3] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3, 1 (2016), 1–9.

[4] Jong Ho Kim, Young Suk Kwon, and Moon Seong Baek. 2021. Machine Learning Models to Predict 30-Day Mortality in Mechanically Ventilated Patients. *Journal of clinical medicine* 10, 10 (2021), 2172.

[5] Zheng Liu, Xiaohan Li, Hao Peng, Lifang He, and S Yu Philip. 2020. Heterogeneous similarity graph neural network on electronic health records. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 1196–1205.

[6] Zheng Liu, Xiaohan Li, Zeyu You, Tao Yang, Wei Fan, and Philip Yu. 2021. Medical triage chatbot diagnosis improvement via multi-relational hyperbolic graph neural network. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1965–1969.

[7] Reza Sadeghi, Tanvi Banerjee, and William Romine. 2018. Early hospital mortality prediction using vital signals. *Smart Health* 9 (2018), 265–274.

[8] Martin Urner, Peter Jüni, Bettina Hansen, Marian S Wettstein, Niall D Ferguson, and Eddy Fan. 2020. Time-varying intensity of mechanical ventilation and mortality in patients with acute respiratory failure: a registry-based, prospective cohort study. *The Lancet Respiratory Medicine* 8, 9 (2020), 905–913.

[9] James M Walter, Thomas C Corbridge, and Benjamin D Singer. 2018. Invasive mechanical ventilation. *Southern medical journal* 111, 12 (2018), 746.

[10] Ni Wang, Yanqun Huang, Honglei Liu, Xiaolu Fei, Lan Wei, Xiangkun Zhao, and Hui Chen. 2019. Measurement and application of patient similarity in personalized predictive modeling based on electronic medical records. *Biomedical engineering online* 18, 1 (2019), 1–15.

[11] Yibing Zhu, Jin Zhang, Guowei Wang, Renqi Yao, Chao Ren, Ge Chen, Xin Jin, Junyang Guo, Shi Liu, Hua Zheng, et al. 2021. Machine Learning Prediction Models for Mechanically Ventilated Patients: Analyses of the MIMIC-III Database. *Frontiers in Medicine* 8 (2021), 955.