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 propose the development of a GNN-based model to accurately predict the mortality of mechanically ventilated patients based on EHR data. First, in this proposal, we present a brief introduction to mechanical ventilation. Then, we provide some insights on the specific dataset that will be used (MIMIC-III), and the AI model that will be proposed and implemented to satisfactorily fulfill the prediction task. Finally, we present some objectives of the project, different motivations, and the expected outcome.

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 [6]. 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. However, despite the use of lung protective mechanical ventilation strategies, it is associated with high morbidity and mortality rates [5]. Therefore, the early assessment of mortality risk for mechanically ventilated patients is crucial in critical care.

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 [2]. Among the data, we find vital signs, demographics, physician notes, and laboratory tests results. The use of this large database along with some novel techniques in the Artificial Intelligence (AI) field could be highly beneficial for the early prediction of mortality in mechanically ventilated patients. This work focuses on a Graph Neural Network (GNN) based architecture. Considering the fact that a graph is a suitable representation for EHR data, and that GNNs have demonstrated a satisfactory performance for multiple biomedical applications in healthcare, we expect to achieve accurate results for the mortality prediction task. The rest of this document includes three sections presenting the *what*, *why*, and *how* of the proposed work.

Data Mining - CS570, Spring 2022, .

ACM ISBN …$

2 WHAT

Patient mortality prediction is an essential task in critical care, which predicts the probability of death. Said task can have a positive impact in the medical decision-making process, resulting in the change of treatments or strategies in an attempt to reduce the mortality risk [4]. Besides, this task can be automated by the use of AI techniques, which not only avoid unnecessary delays in the obtention of the results, but also outperform humans.

As it was mentioned above, graphs are a suitable representation of EHR data, which will be the input of the proposed model. The nodes and edges can represent patients, diagnosis, medications, tests results, and different data stored in the MIMIC-III database, and capture the relationships that exist among them. This was the reasoning behind the selection of a GNN-based model, which will output the desired prediction. Given that mortality prediction is a vast area of research in healthcare, and has many applications in the real world, several research groups focus on the development of new models, or the improvement of existing models for this specific task. A simple search using Google Scholar with the following keywords: *mortality prediction, artificial intelligence*, and *ICU*, shows 18,400 results in the last ten years. However, most of the results use traditional Machine Learning or Deep Learning techniques, and are focused towards the mortality prediction for patients with heart failure and sepsis, not for mechanically ventilation patients.

3 WHY

There are several motivations for the selection and implementation of this work. Among them, we find: 1) It is a real world application that could be introduced into the clinical routines in the future to support the physicians in the medical decision-making process. 2) The number of patients that require mechanical ventilation per year is significantly high, as well as the associated mortality risk. 3) MIMIC-III is publicly available and contains a large number of observations for ICU patients, thus, the data will not be an obstacle for the successful implementation of the project. 4) The scarcity of studies and research focused on mortality prediction for mechanically ventilated patients using AI techniques [3]. 5) The lack of studies and research focused on mortality prediction for mechanically ventilated patients using specifically GNN-based models. To the best of my knowledge, there are no studies yet using GNN-based models. 6) The project involves the implementation of different topics covered in class, such as data cleaning, data reduction, and GNNs.

4 HOW

4.1 Problem Formulation

Propose and implement a GNN-based model using a cohort of mechanically ventilated patients from the MIMIC-III dataset for early mortality prediction.

A graph G can be mathematically defined as a tuple $G = (V, E)$, where V is the set of nodes — in this case, the selected patients, vital signs, medications, laboratory test results, and diagnosis — and E is the set of edges that connect the different pairs of nodes, establishing the existing relationships. This graph G is represented by an adjacency matrix A, which is filled with zeros and ones depending on the connections that exist among the nodes. Said graph will be the input to our implemented model, and the output will be the mortality prediction. Basically, this is a binary classification problem. To achieve this, we proposed the following objectives: i) Preprocess the MIMIC-III database and select the mechanical ventilation cohort of patients, ii) conduct feature selection and construct the required graphs, iii) implement and train the GNN-based model, iv) evaluate and validate the model.

4.2 Proposed Outline

The proposed work will be organized as follows: section 1 will present an introduction about mechanical ventilation in critical care, and GNNs for applications in healthcare in recent years. Section 2 will summarize the previous work in mortality prediction with AI techniques in general, and will also present the state-of-the-art specifically for mortality prediction in mechanically ventilated patients. Additionally, it will present a brief background. Section 3 will provide a description of the methods. It will comprise different subsections: subsection i will include a description of the MIMIC-III database, and the selection for the specific cohort of patients that will be used in this project; subsection ii will present the data preprocessing step and the feature selection process; subsection iii will present the architecture of the proposed model and provide the corresponding details; and subsection iv will present the evaluation of the model, including the selected models for the bench-marking, the chosen evaluation metrics, and the statistical analysis. Section 4 will describe the obtained results, including the training specifications, and using different tables and visualizations to improve the organization of the document. Section 5 will include a discussion of the proposed work and its limitations. Section 6 will discuss the future directions. Finally, section 7 will present a general conclusion.

4.3 Possible Outcomes

At the end of this work, we will achieve the development of a complete pipeline for mortality prediction in mechanically ventilated patients. It will provide detailed information about the data cleaning process, the discretization of some variables, the selection of a suitable patient cohort for the prediction problem, the feature selection, and the contruction of the required graphs. Additionally, we will have an implemented GNN-based model for mortality prediction using EHR data, which will probably lead to a higher accuracy than the one obtained with traditional Machine Learning and Deep Learning models. The corresponding code and files in a GitHub repository.

REFERENCES

[1] Mora Carpio AL and Jorge I Mora. 2017. Ventilator management. (2017).
[2] 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.
[3] 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.
[4] Reza Sadeghi, Tanvi Banerjee, and William Romine. 2018. Early hospital mortality prediction using vital signals. *Smart Health* 9 (2018), 265–274.
[5] 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.
[6] James M Walter, Thomas C Corbridge, and Benjamin D Singer. 2018. Invasive mechanical ventilation. *Southern medical journal* 111, 12 (2018), 746.