ICU Delirium Derived Dataset Manuscript

Mariam Safieldin (mas177@mail.aub.edu)

Manuscript submitted to PhysioNet, still under review

Abstract

Delirium is an acute state of brain failure characterized by unexpected disorientation, a fluctuating course, inattention, and frequently an abnormal degree of awareness. Although delirium is a common illness, it is often overlooked in hospitals. Due to the diverse and fluctuating nature of delirium, as well as its overlap with other psychiatric diseases, it is hard for healthcare professionals to identify delirium patients. This study aims to allow prompt delirium prevention and intervention in hospitalized patients by developing an EHR-based machine learning model capable of accurately predicting incident delirium. Prior studies on delirium used private datasets restricted to certain hospitals or departments. To achieve the goal of developing a model for predicting delirium risk score, in this study, we derived a dataset for ICU delirium and delirium from the MIMIC-III database in order to determine the prevalence of delirium in the ICU clinical context independently and to identify the clinical features that may serve as delirium indicators. We train different machine learning models and propose the best model for predicting the delirium risk score based on our evaluation of the models' performance.

1 Background

Approximately 55% to 80% of delirium cases in the United States are undetected by healthcare teams [1]. Research has found that in intensive care units (ICUs), prevalence rates range from 47.1% to 84.2%. In a study of delirium prevalence across 35 hospital departments, 84.2% of delirium cases were misdiagnosed. The difficulty of detecting delirium is due, in part, to the variability of factors that offset delirium in a patient, including surgery, physical restraints, severe illness such as sepsis and stroke, and many others. While there are several screening methods for delirium, health practitioners must receive specialized training in these techniques to correctly and quickly identify the disease [1].

To achieve our objective of building machine learning models for delirium prognosis in the ICU clinical setting and identifying the features contributing to the models' predictions, we plan to derive a dataset dedicated to patients in ICU clinical setting and identify delirium cases and controls. The source of the ICU delirium extracted dataset would be the Medical Information Mart for Intensive Care (MIMIC-III) database [2]. The feature extraction process for the dataset produced from MIMIC-III for ICU delirium is detailed in the methods section.

2 Methods

This section discusses the extraction process and preprocessing steps for ICU delirium MIMIC-III derived dataset.

MIMIC-III is accessible through **PhysioNetWorks**, a restricted component of PhysioNet [3], after submitting a request that requires completing a recognized course in protecting human research participants that includes Health Insurance Portability and Accountability Act (HIPAA) requirements and signing a data use agreement that outlines appropriate data usage and security standards and prohibits attempts to identify individual patients.

MIMIC-III provides information on 53,423 unique hospital admissions for adult patients (aged 16 or older) admitted to intensive care units between 2001 and 2012. In addition, it includes information for 7870 neonates admitted between 2001 and 2008.The database includes 26 data tables which contains the following information about the patients: demographics, vital sign measurements, laboratory test results, ICU stays, diagnosis, and admission. The following 10 tables were used to extract demographic and clinical information regarding the patients. All tables are provided as CSV files.

1. PATIENTS
2. ADMISSIONS
3. ICUSTAYS
4. CHARTEVENTS
5. D_ITEMS
6. OUTPUTEVENTS
7. INPUTEVENTS_CV
8. INPUTEVENTS_MV
9. DIAGNOSIS_ICD
10. D_ICD_DIAGNOSES

ICU Delirium Features Extraction and Data Tables Merging Steps

The inclusion criteria enforced is the availability of chart records and a minimum age of 18 years old.

1. **Extracting Patients' Demographics and Admission Information**

   Using the subject identifiers (SUBJECT_ID), the **PATIENTS** and **ADMISSIONS** data tables were merged to link each patient to their admission. A single patient could have many admissions. Every admission for the same patient is considered a different record; hence, the admission identifier (HADM_ID) is the primary identifier for merging the tables. These two tables contain information regarding the admission time, admission type, admission location of patients, insurance, language, religion, marital status, ethnicity, gender, and age of the patient. It also gives an indicator showing whether or not a patient's chart information is stored in the database.

   The **ICUSTAYS** data table was integrated with the aforementioned output table to link the admission of each patient to their ICU stay. Consequently, each record in our output table is distinguished by the patient's identifier, the admission identifier, and the ICU stay identifier (ICUSTAY_ID).

2. **Extracting Clinical Charted Observations**

The **CHARTEVENTS** table provides information regarding each patient's charted clinical observations. Each clinical variable is represented by a single record or row within the table. Each row maintains the value, unit of measurement, and item identifier (ITEM_ID) for a specific clinical variable for a particular patient's ICU stay. This table was initially merged with the **D_ITEMS** table, which contains item identifiers (ITEM_ID) and labels, in order to link each item to its label via the ITEM_ID. The **CHARTEVENTS** table comprises 1,459,112,59 rows, each of which contains a value for a specific clinical feature regarding the admission and ICU stay of a specific patient. Due to computational power constraints and memory limits only 1,000,000,00 rows out of the **CHARTEVENTS** table were used, which represents 68.5% of the table. Each record in the table is identified by SUBJECT_ID, HADM_ID, and ICUSTAY_ID.

After analyzing this data table, it was noticed that the table suffered from data entry inconsistencies. The table included 11,847 clinical labels that are not entirely distinct due to inconsistency in the spelling used while entering the clinical variable labels. An example of a data entry inconsistency is shown in Figure 1. This mismatch in the spelling of clinical variables would lead to a dataset with a very high proportion of missing values containing repeated columns with similar names. To the aim of grouping the same clinical variables, yet stored under different names, under the same label; Levenshtein distance was used to identify text similarity between instances of the variable LABEL in the **CHARTEVENTS** table. Two labels were considered very similar if the Levenshtein distance between them is less than the minimum of their lengths divided by a threshold. The threshold is set to 2 if the lengths of the words exceed 5 characters and to 4 if it do not. The threshold value was selected based on experimenting with multiple values to ensure that no irrelevant words are grouped together as similar.

The analysis of this table also revealed that for some patients, the temperature was recorded in Celsius, whilst for others, it was recorded in Fahrenheit. Similarly, some patients' weights were recorded in kilograms (kg), while others were reported in pounds (lbs). Accordingly, all the values for instances of LABEL under temperature that are measured in Fahrenheit were multiplied by 5./9 after subtracting 32 from them to convert them to Celsius. Similarly, all values for instances under weight that are measured in pounds were converted to kilograms by multiplying them by 0.45359237.

After grouping occurrences of the same label in a dictionary under a single title, these instances were all masked to that title thus successfully grouping all their values under a single consistent label. Then, the column LABEL in the **CHARTEVENTS** table was used as a column to pivot the table so that the clinical features represented columns and their values, VALUENUM, represented rows. The resulting table additionally included the following variables as columns: SUBJECT_ID, HADM_ID,

Figure 1: Example of identified data entry inconsistencies in **CHARTEVENTS** data table

ICUSTAY_ID, ITEMID, and CHARTTIME. A total of 159 clinical features were extracted from the **CHARTEVENTS** table.

3. **Extracting Patients' Output Events**

   OUTPUTEVENTS data table includes information related to output from a specific patient during their ICU stay. Similar to **CHARTSEVENTS** data table, this table has the same structure, yet it is smaller in size. **OUTPUTEVENTS** table also suffered from data entry inconsistencies. Similar labels were grouped together and extracted as previously mentioned for the **CHARTEVENTS** data table. A total of 16 output events were extracted from the **OUTPUTEVENTS** table.

4. **Extracting Patients' Input Events**

   INPUTEVENTS_MV and **INPUTEVENTS_CV** data tables include information about patient intake during their ICU stay. The process of extracting clinical features from these two tables follows the same approach for described earlier. A total of 128 input events were extracted from the **INPUTEVENTS_CV** data table and 21 input events from **INPUTEVENTS_MV** data table with 8 input events identified as common between the two tables from the set of extracted events which were combined together under the same labels.

5. The aforementioned transformed tables were merged all together using the following 3 identifiers: SUBJECT_ID, HADM_ID, ICUSTAY_ID. The merged table included included 307 columns representing identifiers, clinical features, demographics, and admission information.

6. **Identifying Delirium Patients**

   Six ICD 9 codes of delirium were used to identify participants who suffered from delirium during their ICU stay in the extracted dataset. The table **DIAGNOSIS_ICD** contains information about the ICD 9 code assigned to every patient's admission. While the table **D_ICD_DIAGNOSES** contains short and a long tiles for every ICD 9 code. These two tables were merged together using the ICD9_CODE identifier, then the output table was merged with the table from the aforementioned extraction steps to obtain an ICD 9 code for every record in our data table. A new binary feature "Delirium" was introduced to indicate whether a person experienced delirium during their ICU stay. An ICU stay was considered a delirium case if the patient was assigned at least one ICD 9 code of those related to delirium in their corresponding ICU stay.

Used Delirium ICD 9 codes and their descriptions:

1. 293.0: Delirium due to conditions classified elsewhere
2. 290.3: Senile dementia with delirium
3. 292.81: Drug-induced delirium
4. 293.1: Subacute delirium
5. 290.41: Vascular dementia, with delirium
6. 291.0: Alcohol withdrawal delirium

The extracted ICU delirium dataset included 38,526 unique records for patients admissions to the ICU. 1714 delirium cases for 1648 unique patients were found out of the extracted ICU stays which represents 4.4% of the extracted records. The number of negative records (Non-Delirium Cases) were downsampled to be double the number of positive records (Delirium Cases) in order to maintain a balance between the number of delirium and non-delirium cases. Statistics of the demographic characteristics of the included patients in the ICU delirium dataset are displayed in Table 1.

Handling Missing Values

A significant proportion of EHR data contains missing values. Handling missing values is an essential step for correctly training a machine learning model. In the extracted MIMIC-III dataset, the percentage of missing values ranged from 99.9% to 0.1%. In prior research using the MIMIC-III data collection [4, 5], the percentage of missing values was restricted to 30%. Other studies allowed the imputation of features with a high proportion of missing data of 99% or less [6, 7]. Since there is no theoretical basis for the threshold to be used to limit the proportion of missing values, we experimented with two thresholds, 35% and 75%, in an effort to determine which threshold would result in better performance.

Characteristic	Train Set	Test Set
Delirium: n (%)		
Yes	1371 (33.3)	343 (33.3)
No	2742 (66.6)	686 (66.6)
Age: median (IQR)	76 (26)	80 (26)
Gender: (%)		
Male	57.5	57.6
Female	42.5	42.4
Marital Status: (%)		
Married	47.5	46.7
Single	24.6	22.1
Widowed	15.3	16.7
Divorced	5.8	7.5
Separated	1.2	1.2
Unknown	0.5	0.7
Missing	5.1	5.1

Table 1: Demographic characteristics of the included patients in ICU delirium MIMIC-III dataset

All missing values in both datasets were imputed using DataWig [8], a Deep Learning toolkit for missing value imputation. To predict the missing values in each of the features of the extracted dataset, DataWig's imputer was trained on the dataset and used to fill in the missing values in both train and test sets.

The datasets included features with missing values ranging from 0.1% to 99%. The missing values percentage was limited to a maximum of 75%. Two versions from every dataset were generated based on the thresholds 35% and 75%. Details about the missing values percentages for every feature in both datasets are listed in Table 4. We will experiment using the datasets with features included based on the missing values percentage threshold.

3 Experimental Results

The results of training the CatBoost classifier using the two versions of the ICU delirium MIMIC-III dataset based on the threshold percentage of missing values are presented in Table 2. The results indicate that a threshold of 75% allows the model to train on more features that contribute to its performance improvement. Except for the specificity, which showed only 0.3% improvement when trained on the dataset with a 35% missingness threshold, the results of the model trained on the dataset with features included based on a 75% missingness threshold were significantly better.

Threshold	**Accuracy**	**AUC**	**F1 Score**	**Recall**	**Precision**	**Specificity**	**NPV**
75%	**92.2**	**97.4**	**88.3**	**87.8**	**88.8**	**94.5**	**93.9**
35%	89.2	93.3	83.0	79.0	87.4	94.3	90.0

Table 2: CatBoost performance on ICU Delirium MIMIC-III dataset with two thresholds for the percentage of missing values

On the basis of the obtained results, a missing value percentage threshold of 75% was selected for the ICU delirium MIMIC-III dataset. More information about the percentage of missing values for every feature is included in Table 4.

Table 3 displays statistical information about the nature of variables in our dataset and their total number of records.

	Categorical	**Numerical**	**Binary**	**Rows**	**Predictors**
ICU Delirium Dataset	9	72	0	(4113, 1029)	81

Table 3: Statistical information on the nature of variables in the ICU delirium MIMIC-III derived dataset

3.1 Missing Values Percentages Per Feature

Table 4 shows the percentage of missing values per every feature in the ICU delirium MIMIC-III derived dataset.

Feature	%	**Feature**	%
Fingerstick Glucose	0.95	Potassium (whole blood)	0.97
Blood Urea Nitrogen (BUN)	1.3	Sodium (serum)	1.3
Religion	1.3	Heart Rate	1.9
Respiratory Rate	1.9	O2 saturation pulseoxymetry	2.0
Temperature (C)	2.0	White Blood Count (WBC)	2.1
Hemoglobin	2.0	Marital Status	5.0
O2 Flow	15	Urine Out Foley	17
Creatinine	20	Magnesium	21
Phosphorous	24	International Normalized Ratio (INR)	28
Partial Prothrombin Time (PTT)	29	GCS Eye Opening	32
GCS Verbal Response	32	GCS Total	32
GCS Motor Response	32	Braden Score	33
Inspired O2 Fraction	33	NBP Mean	34
Weight	35	Dextrose 5 % In Water (D5W)	36
Arterial Base Excess	37	.9% Normal Saline	44
Arterial Blood Pressure mean	44	Ionized Calcium	48
Positive End-Expiratory Pressure (PEEP)	48	Minute Volume	48
Peak Insp. Pressure	48	Po Intake	48
Mean Airway Pressure	48	Language	49
Alanine Transaminase (ALT)	50	Aspartate Transferase (AST)	51
Plateau Pressure	54	Lactic Acid	54
Potassium Chloride (KCL)	55	Arterial pH	55
Arterial CO2	55	Arterial PaCO2	55
Differential-Basos	58	Differential-Eos	58
Differential-Lymphs	58	Differential-Monos	58
FiO2 Set	59	Central Venous Pressure	60
Propofol	61	Albumin	62
Lactate Dehydrogenase (LDH)	63	Low Exhaled Min Vol	63
Sensitivity-Vent	63	Tidal Volume	65
Respiratory Rate Set	65	Resp Rate (Spont)	65
Pressure Support	65	Tidal Volume (Set)	65
High Resp. Rate	66	Gastric Meds	66
Anion gap	67	Chloride (serum)	67
Bicarbonate (HCO3) (serum)	67	Hematocrit (whole blood)	67
Platelet Count	67	Calcium non-ionized	68
Insulin	69	Prothrombin time	71
Saturation of arterial blood with oxygen (SaO2)	71	Oral Gastric	72

Table 4: Missing values percentages in ICU delirium MIMIC-III dataset

References

[1] Melissa LP Mattison. Delirium. *Annals of internal medicine*, 173(7):ITC49–ITC64, 2020.

[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. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[3] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

[4] Yue Yu, Jun Wang, Qing Wang, Junnan Wang, Jie Min, Suyu Wang, Pei Wang, Renhong Huang, Jian Xiao, Yufeng Zhang, et al. Admission oxygen saturation and all-cause in-hospital mortality in acute myocardial infarction patients: data from the mimic-iii database. *Annals of Translational Medicine*, 8(21), 2020.

[5] Yibing Zhu, Jin Zhang, Guowei Wang, Renqi Yao, Chao Ren, Ge Chen, Xin Jin, Junyang Guo, Shi Liu, Hua Zheng, et al. Machine learning prediction models for mechanically ventilated patients: analyses of the mimic-iii database. *Frontiers in Medicine*, 8:955, 2021.

[6] Evan J Tsiklidis, Talid Sinno, and Scott L Diamond. Predicting risk for trauma patients using static and dynamic information from the mimic iii database. *Plos one*, 17(1):e0262523, 2022.

[7] Mengling Feng, Jakob I McSparron, Dang Trung Kien, David J Stone, David H Roberts, Richard M Schwartzstein, Antoine Vieillard-Baron, and Leo Anthony Celi. Transthoracic echocardiography and mortality in sepsis: analysis of the mimic-iii database. *Intensive care medicine*, 44(6):884–892, 2018.

[8] Felix Biessmann, Tammo Rukat, Philipp Schmidt, Prathik Naidu, Sebastian Schelter, Andrey Taptunov, Dustin Lange, and David Salinas. Datawig: Missing value imputation for tables. *J. Mach. Learn. Res.*, 20(175):1–6, 2019.

Acronyms

ALT Alanine Transaminase. 8

AST Aspartate Transferase. 8

BUN Blood Urea Nitrogen. 8

D5W Dextrose 5 % In Water. 8

GCS Glasgow Coma Scale. 8

HCO3 Bicarbonate. 8

INR International Normalized Ratio. 8

KCL Potassium Chloride. 8

LDH Lactate Dehydrogenase. 8

MIMIC-III Medical Information Mart for Intensive Care. 2, 5–8

PEEP Positive End-Expiratory Pressure. 8

PTT Partial Prothrombin Time. 8

SaO2 Saturation of arterial blood with oxygen. 8

WBC White Blood Count. 8