Foundations in Machine Learning Project Proposal

Mortality Prediction of Patients in Intensive Care Units

Yasmina Hobeika
Master's in DSBA
B00804284
yasmina.hobeika@student-cs.fr

Victor Hong
Master's in DSBA
B00802726
victor.hong@student-cs.fr

Ali Najem
Master's in DSBA
B00806315
ali.najem@student-cs.fr

Arindam Roy
Master's in DSBA
B00802759
arindam.roy@student-cs.fr

PROBLEM DEFINITION & MOTIVATION

Resources in hospitals, particularly in Intensive Care Units (ICU) can be scarce and in emergency situations where ICU capacity is stretched, decisions need to be made on whether to discharge existing patients who have been warded in lieu of incoming patients who require urgent medical attention. This situation raises the question of which characteristics, and its corresponding thresholds should be used to discharge existing patients and to accept incoming patients. These characteristics are critical to avoid the risk of prematurely discharging already warded patients and admitting incoming patients who require urgent medical attention.

This project aims to predict the survival of a given patient coming into the hospital. We have data collected from each patient with features based on demographics such as age and sex and including medical conditions: Body vitals (blood pressure, temperature, SO2 levels…) and disease history (diabetes, infections, heart failure…). This data is collected at hospitals and monitored throughout patients' stay at the ICU. By applying different machine learning models, we are able to better predict the outcomes and severity of the situation of individual patients

In this project, we will be using data collected by the ICUs from hospitals to predict the survival of a patient. This project has wide scale implementations, a few of which include: Patient prioritization, prognostic healthcare, healthcare infrastructure recommendations, hospital operation recommendations.

METHODOLOGY

We have data collected from each patient with features based on demographics such as age and sex and including medical conditions: Body vitals (blood pressure, temperature, SO2 levels…) and disease history (diabetes, infections, heart failure…). This data is collected at hospitals and monitored throughout patients' stay at the ICU. By applying different machine learning models, we are able to better predict the outcomes and severity of the situation of individual patients.

This project will be divided into three phases for the model implementation. The first being the EDA phase, where the aim will be to find remarkable features of the data. The second phase of this project will be data cleaning, where the aim will be to remove unnecessary or highly correlated features. This phase may also include feature engineering, SVD and PCA steps. The final step of the model implementation will be to test various ML models, such as linear regression, logistic regression, SVM, KNN, decision trees, random forest, et cetera. The model that performs the best in evaluation will be selected.

SOURCE OF DATA

This project will utilize data from the Kaggle repository named Patient Survival Prediction [0]. This dataset has 85 features which include BMI, age, weight, heart rate, temperature, SPO2, blood pressure and many more. It has also a binary column which indicates if the patient died.

EVALUATION

Since the problem in hand deals with binary classification, we will mainly be relying on the confusion matrix and the related statistics like F1, F2 scores. Adding to this, with the null hypothesis being "the patient does not survive", it seems logical to reduce the type 1 error. With further discoveries and analysis, the evaluation method may change.

KEYWORDS

Machine learning, Single Value Decomposition (SVD), Principal Component Analysis (PCA), K-nearest neighbor (KNN), Decision Tree (DT), Random Forest (RF).

LITERATURE REVIEW

Paper 1: **Mortality prediction of patients in intensive care units using machine learning algorithms based on electronic health records [1]**

Purpose of study: Improving predictive models for intensive care unit (ICU) inpatients requires a new strategy that periodically includes the latest clinical data and can be updated to reflect local characteristics. We extracted data from all adult patients admitted to the ICUs of two university hospitals with different characteristics from 2006 to 2020, and a total of 85,146 patients

were included in this study. Machine learning algorithms were trained to predict in-hospital mortality.
Key Features: 70 features including type of admission, sex, age, type of ICU admit, antibiotics taken, and existing comorbidities.
Methodology: Training model with KNN, DT, RF, XG Boost, LGB, SVM & ANN; where 80% of data is used to train model, and 20% of data used for testing
Results Obtained: ML can accurately collect patient data to support clinical decision-making. Study shows ML-based models have accurate predictive power in the ICUs. Conventional models have strengths in standardization and center-to-center comparisons. Using ML models in conjunction with conventional scoring systems can provide more useful information for predicting the prognosis of critically ill patients.

Paper 2: Early hospital mortality prediction using vital signals [2]
Purpose of study: Predict mortality using features extracted from the heart signals of patients within the first hour of ICU admission
Key Features: 12 Features extracted from the heart signals of patients
Methodology:
1- Preprocessing: Remove noise and use FIR filter to interpolate new samples to resample the signals with a lower sampling rate.
2- Feature Extraction: Each signal is described in terms of 12 statistical and signal-based features which were extracted from the patient's ECG signal
3- Classification: They used eight classifiers: decision tree, linear discriminant, logistic regression, support vector machine (SVM), random forest, boosted trees, Gaussian SVM, and K-nearest neighborhood (K-NN).
The results they got: The decision tree may provide the best choice as a tradeoff between transparency and accuracy.

Paper 3: Prediction of Mortality in an Intensive Care Unit using Logistic Regression and a Hidden Markov Model [3]
Purpose of study: This paper presents the results of an improved patient specific ICU mortality assessment algorithm. It makes use of several computationally sophisticated techniques.
Key Features: Age, PF Ratio, Creatinine, Blood urea nitrogen, Bilirubin, Glasgow Coma Score, Glucose, Disease Development.
Methodology: Logistic regression, Hidden Markov Model, trained on 4000 observations, and validated further on two sets of another 4000 observations.
Results Obtained: Algorithm outperforms commonly used critical illness severity assessment score, SAPS-I. In addition, since it uses the instantaneous values of vitals and labs, it could be used to obtain patient specific trajectories of mortality risk assessment to help prioritize resources and determine the appropriate plans.

Paper 4: Prediction of hospital mortality in intensive care unit patients from clinical and laboratory data: A machine learning approach [4]
Purpose of study: Develop a binary classifier for the outcome of death in ICU patients based on clinical and laboratory parameters
Key Features: 50 Features including body vital signs (temperature, SO2 levels, BP, etc.)
Methodology:
1- Logistic regression
2- Hidden Markov Model
3- Trained on 4000 observations, and validated further on two sets of another 4000 observations each.
The results they got: Machine learning has great potential to provide us support to critically question existing methodologies, allowing improvements that reduce mortality.

Paper 5: Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning based, retrospective analysis of the MIMIC-III database [5]
Purpose of study: The predictors of in-hospital mortality for intensive care units (ICUs)-admitted heart failure (HF) patients remain poorly characterized. We aimed to develop and validate a prediction model for all-cause in-hospital mortality among ICU-admitted HF patients.
Key Features: 52 Features including age, hypertension, heart rate, blood pressure, SPO2…
Methodology:
1- Data extraction: using PostgreSQL
2- Missing data handling: For normally distributed continuous variables, the missing values were replaced with the mean for the patient group. For skewed distributions related to continuous variables, missing values were replaced with their median.
3- Algorithms Used: XGBoost and Lasso (from 52 till 20 features)
The results they got: With a high AUC of 0.8416 (95% CI 0.7864 to 0.8967) and a wide net benefit threshold range (>0.1), this nomogram can be widely used to enhance more accurate clinical decision-making.

REFERENCES

[0] M. Agarwal, "Patient survival prediction," Kaggle, 26-Dec-2021. [Online]. Available: https://www.kaggle.com/datasets/mitishaagarwal/patient.

[1] M. H. Choi, D. Kim, E. J. Choi, Y. J. Jung, Y. J. Choi, J. H. Cho, and S. H. Jeong, "Mortality prediction of patients in intensive care units using machine learning algorithms based on Electronic Health Records," Nature News, 03-May-2022. [Online]. Available: https://www.nature.com/articles/s41598-022-11226-4.

[2] R. Sadeghi, T. Banerjee, and W. Romine, "Early hospital mortality prediction using vital signals," Smart Health, 06-Jul-2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S2352648318300357.

[3] "Prediction of mortality in an intensive care unit using logistic regression and a hidden Markov model," IEEE Xplore. [Online]. Available: https://ieeexplore.ieee.org/document/6420413.

[4] E. Caires Silveira, S. Mattos Pretti, B. A. Santos, C. F. Santos Corrêa, L. Madureira Silva, and F. Freire de Melo, "Prediction of hospital mortality in intensive care unit patients from clinical and laboratory data: A machine learning approach," World journal of critical care medicine, 09-Sep-2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9483004/.

[5] F. Li, H. Xin, J. Zhang, M. Fu, J. Zhou, and Z. Lian, "Prediction model of in-hospital mortality in intensive care unit patients with heart failure: Machine Learning-based, retrospective analysis of the mimic-III database," BMJ Open, 01-Jul-2021. [Online]. Available: https://bmjopen.bmj.com/content/11/7/e044779.info.