Prediction of Cardiac Death in ICU Patients

Partha Sarathi Bera, Harsh Kohli, Jennifer S. Johnson, and Francisco Duran
Georgia Institute of Technology

Abstract

Globally, Heart Failure claims more lives than colorectal cancer, breast cancer, prostate cancer, influenza, pneumonia, auto accidents, HIV, firearms, and house fires combined [2]. Effective predictors of intensive care unit (ICU) mortality can help identify high-risk patients earlier, improve resource allocation, and create more accurate risk models. When cardiopulmonary resuscitation (CPR) is provided quickly, and there is an effective system of care, the chances of survival with full neurological recovery increases. As recent as 2016, the incidence of cardiac arrest in hospitalized patients (IHCA) was 209,000 with a survival rate for adults of 24.8% [3]. Leveraging patient medical records, biometric data, and machine learning algorithms, this research project aims to identify mortality predictors of cardiac arrest in ICU patients using data derived from the publicly available Medical Information Mart for Intensive Care (MIMIC-III) database [4]. A video presentation of this work and the project code can be found at the following link: https://github.gatech.edu/hkohli3/ICU-Mortality-Prediction-Team36

Introduction and background

Machine-learning (ML) models aimed to predict the outcome of patients in ICU are valuable for many reasons, some of which are particularly relevant. For example, stratifying patients by their risk to adverse events provides a way of more accurately evaluating desired metrics, e.g., if one hospital has higher mortality rate than another, it may just reflect a difference in the average health of the two different populations. The ability to empirically separate patients by risk essentially allows the calibration of these evaluations for more accurate comparisons [8]. Optimizing resource utilization is another benefit. ICUs are congested, and many patients are not able to receive critical care that would be beneficial to them [6]. Both total expenditure and expenditure per day in the ICU were highest for patients whose outcomes were the most unexpected when compared to a physician predicted prognosis [4]. Being able to predict how at risk various patients are throughout their stay provides an empirical basis for scheduling therapies, allocating resources, and optimizing discharge time [8].

Physicians have to make clinical decisions everyday based on a small set of data known to them by their medical training and experience, and data based on the current patient. However, a big-data derived prognostic model provides the advantage of being supported by more data than any physician's experience, with the benefit of a traceable and reproducible bias.

Studies have shown that prognostic models can improve patient care [9]. Heart Failure (HF) is a complex syndrome characterized by the inability of the heart to supply sufficient blood flow to the body. HF is clinically diagnosed by left ventricular ejection fraction [1]. Symptoms of HF include shortness of breath, fatigue, edema, rapid or irregular heartbeat, persistent cough or wheezing with white or pink blood-tinged phlegm; swelling of the abdomen, fluid retention, nausea, and decreased alertness. Real-time identification of the diseases or conditions of a hospitalized patient is important for HF identification. A number of algorithms and interventions have been developed to improve disease cohort identification for HF [5], [7], [8], [9], [10]. The studies surveyed studying real-time identification of hospitalized patients with HF have relied on a small number of clinical factors and have had mixed success [10]. Five different algorithms were identified from the surveyed publications:

1. The presence of HF on the patient's admissions history (acute myocardial infarction and atherosclerosis).
2. The presence of at least 1 of the following characteristics: HF, inpatient oral or intravenous loop diuretic use, brain natriuretic peptide level of 500 pg/mL or greater.
3. Logistic regression of clinically relevant structured data elements.
4. ML model using unstructured data.
5. ML model using unstructured and structured data.

Algorithms based on 1, 2, and 3, were insufficient for real-time identification of hospitalized patients with heart failure [10]. Algorithms based on 4 and 5 had high predictive accuracy of ML, which demonstrates that support of such analytics in future electronic health records (EHR) systems can improve cohort identification. Algorithms 3, 4, and 5, preliminarily, had areas under the receiver operating characteristic curves of 0.953, 0.969, and 0.974, respectively.

This study implements an ML approach with structured data to identify hospitalized patients with HF that can help reduce the mortality rate associated with cardiac death.

Approach and implementation

The methodology for extracting pertinent features from the MIMIC-III dataset, converting the raw data to a standardized format, and predicting patient outcomes is described below.

Dataset description

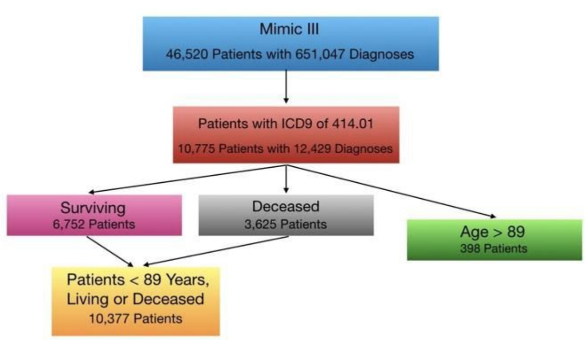

Figure 1. MIMIC-III Database

This study leverages the MIMIC-III dataset, which is a large and public database developed by the MIT Lab for Computational Physiology, comprised of de-identified health data associated with nearly 50,000 critically ill patients admitted to ICUs of the Beth Israel Deaconess Medical Center from 2001 to 2012. The database contains demographics, vital sign measurements made at the bedside, laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (both inside and out of the hospital [4]. Hospitalizations for patients between 18 and 89 years old were included.

From a random 75% sample of hospitalizations, an algorithm for HF identification using EHR data was developed using an ML approach with structured data. From the MIMIC-III data, the following unique diagnoses are relevant to the HF algorithm:

Table 1. Unique Diagnoses

ICD9	Description	Diagnoses count
414.01	Coronary atherosclerosis	12,429
401.9	Hypertension	20,703
428	Heart Failure	18,611
427	Cardiac Dysrhythmia	20,301

Cohort identification

From the MIMIC-III database, the cohort was selected based patients with a diagnosis of Coronary Atherosclerosis (ICD9 code of 414.01), Hypertension (401.9), Heart Failure (428), and Cardiac Dysrhythmia (427) as shown above in Table 1. Using Apache Pig, a number of features including flags were derived from the MIMIC-III structured data to aid in determining the case and control cohorts. Table 2 below summarizes each cohort definition and the methods used to manipulate data in the demographics.csv file.

Table 2. Cohort Definitions

Cohort	Definition	Flags	Method used
Heart Failure (HF) Patient	ICU patients with a diagnosis of coronary atherosclerosis, hypertension, heart failure, and cardiac dysrhythmia	HEART_ATTACK_FLAG == 1 ATHERO_DIAGNOSIS_FLAG == 1	String matching of relevant words and ICD-9 codes
Case	HF patient who expires due to cardiac complications while a patient in the hospital	DEATH_FLAG == 1 HEART_DEATH_FLAG == 1 OUTSIDE_DEATH_FLAG == 0 (HOSPITAL_EXPIRE_FLAG == 1)	Non-null/non-empty date of death and heart death description fields

Table 2. Cohort Definitions (continued)

Control	HF patient who expires due to non-cardiac complications while a patient in the hospital	DEATH_FLAG == 1 HEART_DEATH_FLAG == 0 OUTSIDE_DEATH_FLAG == 0 (HOSPITAL_EXPIRE_FLAG == 1)	Non-null/non-empty date of death and null/empty heart death description fields
Survivor	HF patient who survives the ICU visit beyond a threshold number of days after discharge	DEATH_FLAG == 0	Empty date of death field in demographics.csv file

Feature selection

Preliminary data exploration was conducted on the demographic characteristics found in the population separating the deceased from those patients who survived. The sample was further divided by race, admission status, and insurance type to try to find interesting correlations that could be further explored.

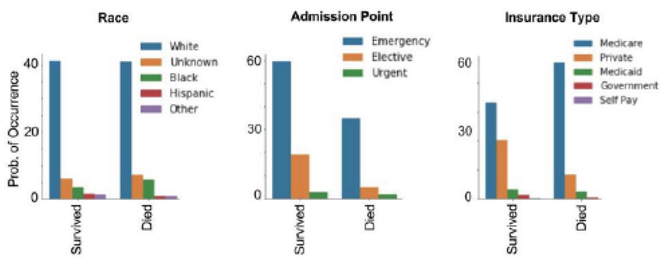

Figure 2. Exploratory Analysis of Patient Data

Looking at just the three examples provided above in Figure 2, the feature 'race' seems to have some effect on outcome, as the average patient who dies is more likely to not self-classify as white. Similarly, these patients who die are far more likely to be admitted to the ICU via the ER and are more likely to be on Medicare insurance. Of course, there are confounders for this data (for example, Medicare patients are also more likely to be older). Applying an ML approach on these correlated demographic, admission, and other data enable this algorithm to predict patients most likely to pass away.

A feature set involving all demographic data available (race, religion, language, insurance type, marriage status, age, sex among others) was created as well as a naive interpretation of medical history including number of appointments. Combined after one-hot encoding, the feature set is approximately 120 features.

Model architecture

Several different architectures were examined. Given that there were only about 8200 training examples (after removing the validation and test tests), all necessary steps were taken to ensure the model does not overfit to the training data. For all experiments, the batch size was fixed at 32 since variants of stochastic gradient descent optimizers operate within a small batch regime and using a larger batch size has been shown to lead to significant degradation in the quality of the model due to possible overfitting. The Adam optimizer was used as it combines the regularization effect of RMSProp with Momentum and normalizes the weights by sum of their root mean squared values so that they do not become too big leading to high variance in the network. Momentum ensures that the weight do not change too much based on a small, noisy batch of inputs as the gradients preserve information from previous batches. The learning rate for Adam was fixed at 0.0005 (down from 0.001 which is default) to facilitate a slower, smoother convergence. Dropout with keep probability of 0.5 was applied on all layers for added regularization and to prevent co-adaption between features. Softmax activation was applied on a 2-neuron output layer for all architectures and the model was optimized on the cross-entropy loss between the labels and the output logits.

Multilayer perceptron

A simple multi-layer perceptron (MLP) was trained with three hidden layers. The inputs used were one-hot representations of the 6 features – age, ethnicity, gender, language, marital status and religion for a total of 121 sparse features. The hidden sizes of the three layers were set to 200, 100 and 50 respectively, which was subsequently followed by the output layer. ReLU activation was used for the hidden neurons.

Highway maxout with sparse mixture of experts layer

Highway Maxout Networks (HMN) [11] are inspired from Long-Short Term Memory networks or LSTM networks which use a gated mechanism to allow gradients to better flow over many time-steps and are more robust to vanishing and exploding gradients than traditional Recurrent Networks. Highway Networks use a similar gated mechanism to create an "information highway" that allows the network to adaptively allow the input to either directly pass onto the next layer or be acted upon by the activation function of the present layer H(x). Maxout Networks compute the maximum of activations of k neurons (k being the window size) and learn a piecewise linear function which this leads to a better and more efficient model averaging over dropout (which approximates this by training a random sub-network at each iteration). Combining highway networks with maxout over activations have yielded state of the art results on many tasks involving deep networks [12].

Maxout activation with a pool size of 16 is applied to a sparse mixture of experts layer. The mixture of experts layer also uses a gating mechanism to combine information from several, smaller sub-networks. A sparse combination of individual experts is determined through gating in a process akin to condition computation where only parts of the network are active on a per-example basis. In our case, 16 experts were used each with an output size of 200. An information highway is created by applying a second maxout to the combined output of the mixture of experts layer directly and the output from the first maxout layer. Like in the case of the MLP, inputs to the network are the 121-dimension sparse representations of the 6 features.

Feature embedding layer

Several unsupervised approaches exist to convert large one-hot representations to a more compact and useful embedding which can be better absorbed by any learning algorithm. These approaches have been widely successful in Natural Language Processing (NLP) tasks, specially when learning embeddings for individual words – Word2Vec and GloVe. Such embeddings are typically pretrained on a large, unsupervised corpus and intrinsic properties of the various categories (words, in this case) and transfer-learned by the subsequent learning algorithm and fine-tuned based on the learning objective.

While due to limitations in the data, we could not pre-train embeddings for the various categories in our features, we added an embedding layer for these within our model architecture to learn these embeddings on the fly for our task. Thus the one-hot representations of our categorial features - ethnicity, gender, language, marital status and religion were first passed through a single dense layer (effectively a linear transformation with an added bias) with hidden size 25. We hypothesize that while the vanilla models should typically learn the inter-relatedness between features, explicitly grouping the related features (like all the one-hot vectors of, say, ethnicity) manually would make the learning objective easier. Such feature engineering is necessary given the small amount of training data as the model, now, does not have to explicitly learn the individual indices of the various categories but only train on their embedded representations.

The 25-dimension embeddings from the six individual features are then concatenated and fed to subsequent layers. The final, 150-dimensional representation was input to both the MLP and Highway Maxout networks discussed above to see which performs better.

Experimental evaluation

The results from the various experiments are described below. Training was carried out for a maximum of 10000 iterations in each run. The best model was saved based on the lowest cross entropy loss on a validation set of 1000 examples. A held-out test set of another 1000 examples was used to determine the model accuracy at the end of training.

Table 3. Test Accuracies of Four Models

Model	**Accuracy**
MLP	0.721
HMN	0.715
MLP + Embedding	0.705
HMN + Embedding	0.75

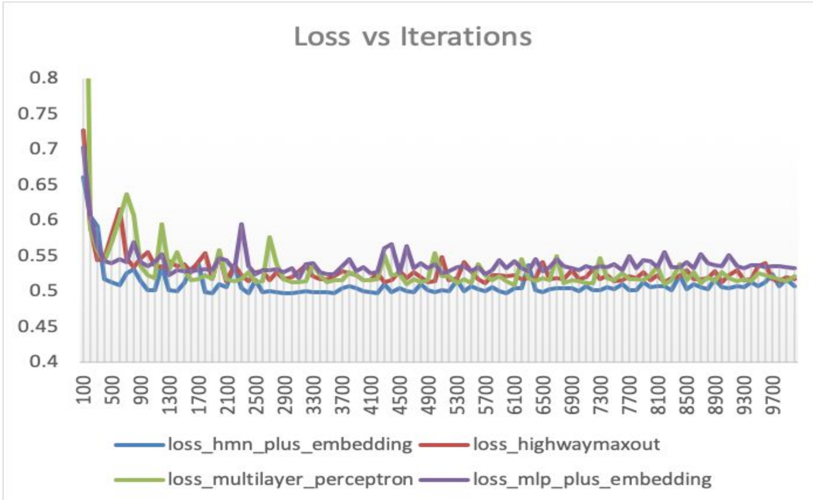

Figure 3. Validation Losses (per model)

Table 3 above shows the test accuracies of the four models and the Figure 3 describes how the validation loss varies with the number of iterations for each model. While there is little to choose from between MLP, HMN and MLP with embedding, consistent with our expectations the Highway Maxout Network with the added embedding layer performs the best with a final test accuracy of 75%. Figure 3 illustrates the validation loss for HMN plus embedding model is consistently better than the other models, and it also converges faster, likely due to a more expressive architecture with more trainable parameters.

Conclusion

One challenge faced in completing this project was in understanding the unstructured data, especially the clinical notes. Given more time, this unstructured data could have been properly incorporated into the ML model, and perhaps have resulted in equal or better accuracy.

The members of Team 36 worked synergistically during this project despite working across two times zones. Summarized in Table 4 below are the contributions of each team member.

Table 4. Team Member Contributions

Team 36 Members	**Contribution**
Partha Sarathi Bera	Feature/cohort data extraction
Harsh Kohli	Machine Learning/Highway Maxout Network modeling
Jennifer S. Johnson	Project integration and final deliverables
Francisco Duran	Feature/cohort data extraction using big data tools

References

1. T.R. Dawber, G.F. Meadors, F.E. Moore. Epidemiologic approaches to heart disease: the Framingham Study. Am J Public Health, 41 (1951), pp. 279-286.
2. Cpr.heart.org. (2019). CPR Facts and Stats. [online] Available at: https://cpr.heart.org/AHAECC/CPRAndECC/AboutCPRECC/CPRFactsAndStats/UCM_75748_CPR-Facts-and-Stats.jsp [Accessed 3 Mar. 2019].
3. Merchant, R. M., Yang, L., Becker, L. B., Berg, R. A., Nadkarni, V., Nichol, G., Carr, B. G., Mitra, N., Bradley, S. M., Abella, B. S., Groeneveld, P. W., American Heart Association Get With The Guidelines-Resuscitation Investigators (2011). Incidence of treated cardiac arrest in hospitalized patients in the United States. Critical care medicine, 39(11), 2401-6.
4. MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available from: http://www.nature.com/articles/sdata201635.
5. Le Gall J, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. JAMA. 1993;270(24):2957–2963.doi:10.1001/jama.1993.03510240069035.
6. Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A., & Szolovits, P. (2014). Unfolding Physiological State: Mortality Modelling in Intensive Care Units. KDD : proceedings. International Conference on Knowledge Discovery & Data Mining, 2014, 75-84.
7. Suzette J. Bielinski. Mayo Clinic. Heart Failure (HF) with Differentiation between Preserved and Reduced Ejection Fraction. PheKB; 2013 Available: https://phekb.org/phenotype/147.
8. Seneff, Michael, and William A. Knaus. “Predicting patient outcome from intensive care: a guide to APACHE, MPM, SAPS, PRISM, and other prognostic scoring systems.” Journal of Intensive Care Medicine 5.1 (1990): 33-52.
9. Selker, Harry P., Joni R. Beshansky, and John L. Griffith. “Use of the electrocardiograph-based thrombolytic predictive instrument to assist thrombolytic and reperfusion therapy for acute myocardial infarction: a multicenter, randomized, controlled, clinical effectiveness trial.” Annals of internal medicine 137.2 (2002): 87-95.
10. Blecker S, Katz SD, Horwitz LI, et al. Comparison of Approaches for Heart Failure Case Identification From Electronic Health Record Data. JAMA Cardiol. 2016;1(9):1014–1020. doi:10.1001/jamacardio.2016.3236.
11. Rupesh K Srivastava, Klaus Greff, and Juergen Schmidhuber. Training very deep networks. In Advances in Neural Information Processing Systems 28, pp. 2377–2385, 2015.
12. Xiong, C., Zhong, V., and Socher, R. 2016. Dynamic coattention networks for question answering. CoRR abs/1611.0160.