Predicting In-Hospital Mortality and readmission on eICU Dataset

Ankita Singh Joel Dsouza Purav Zumkhawala

***Abstract*—A patient is placed in an Intensive Care Unit (ICU) when the patient's condition is extremely critical and cannot afford any errors. Incorrect readings of vital signs or mistakes by medical staff can cost a persons life. The aim of the project is to predict patients in-hospital mortality rate upon being discharged from the ICU with the help of supervised machine learning techniques like regression and boosting. The eICU Collaborative Research Database [4] was utilised to explore and analyze various features such as vitals measurements, diagnosis reports, ICD-9 codes, etc and to predict the mortality rate of patients. Two models were developed for prediction - the Admission Baseline model, and the Retrospective Derived Features model. Similar models were built for patient re-admission prediction. The data was processed and analyzed using Logistic Regression and XGBoost and the final f-1 score and AUROC was utilized for the assessment of accuracy and reliability of the results. Finally, leave-one-out hospital cross validation was performed to evaluate individual hospital performance.**

***Index Terms*—Mortality, eICU, Re-admissions**

I. Introduction

It is important to monitor a patients condition regularly upon ICU discharge due to the critical condition they're in. Mortality prediction is useful in this situation as accurate knowledge of a patient's disease state and trajectory enables better care for the patient. In this project the The eICU Collaborative Research Database [4] was utilised to explore various features such as vitals measurements, diagnosis reports, ICD-9 codes, etc and to predict the mortality rate of patients. The eICU dataset consists of 31 tables of medical information such as vitals, diagnosis, and lab results collected from 355 intensive care units, from 208 hospitals, across the United States with more than 200,000 patient records. Features were derived and two models were developed for in-hospital mortality prediction - the Admission Baseline model, and the Retrospective Derived Features model. The features for the Admission Baseline model are age, gender, and features needed to calculate the SAPS II score at admission (3 features). The Retrospective Derived Features Model utilizes the features age, gender, features needed to calculate the SAPS II score, and all Elixhauser comorbidities (36 features). Similar models are built for patient re-admission prediction. The data was processed and analyzed using Logistic Regression and XGBoost and the final f-1 score and AUROC was utilized for the assessment of accuracy and reliability of the results. Finally, leave-one-out hospital cross validation was performed to evaluate individual hospital performance.

II. Related Work

Earlier works [1] [2] utilized the MIMIC-III dataset [3] to perform re-admissions and mortality prediction analysis. The MIMIC dataset is outdated and consists of records of 49,785 hospital admissions of 38,597 patients only. It consists of 26 tables that describe vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, etc. In the paper 'Unfolding physiological state: Mortality modelling in intensive care units' the authors implemented 6 different models in three prediction regimes: baseline prediction, dynamic (time-varying) outcome prediction, and retrospective outcome prediction to predict inhospital, 30 day post-discharge, and 1 year post-discharge patient mortality. Apart from patient information, vital measurements, diagnosis, and lab results this paper also utilizes the doctors notes by using latent variable models to derive features. Features were used to extract SAPS-II and Elixhauser Scores which were used as features and achieves an AUROC of 0.771, Sensitivity of 0.999, Specificity of 0.010 for the Admission Baseline Model. For the Retrospective Derived Features Model the AUROC, sensitivity and specificity values are 0.901, 0.997, and 0.108 respectively.

The paper 'Prediction of ICU Readmissions Using Data at Patient Discharge' predicts the possibility of patient re-admissions. It identifies risks of ICU readmissions at 24 hours, 72 hours, 7 days, 30 days, and bounceback readmissions in the same hospital admission with an AUROC for 72 hours of 0.76 and for bounceback of 0.84.

III. Methods

A. Data

We used the eICU Collaborative Research Database (eICU-CRD), a publicly available multi-center database sourced from the Philips eICU programme.The eICU Collaborative Research Database is populated with data from a combination of many critical care units throughout the continental United States. The data in the collaborative database covers patients who were admitted to critical care units in 2014 and 2015. The eICU-CRD contains 200,859 distinct unit stays, from which we exclude non-ICU stays and ICUc stays which do not have the Apache Variables present. We considered ages of patients between 18 to 89 which are for adult patients. Ages of patients above 89 were scrambled and so would skew our

results and hence not considered. Below 18 years patients were children and do not fit the general conditions of patient predictions. All numerical variables were mean imputed for missing values. Categorical variables that were missing were dropped. and not used in further computation. The response variable was the hospitalDischargeStatus which provided us with the status of the patient while leaving the hospital and was used as the label in our model. For the readmission prediction, we used the same Vital signs and demographic data but changed the response variable to unitStayType.

B. Pre-processing

Pre-processing step is same for both mortality and readmission. In the first step we drop every record where the hospitalDischargeStatus is empty. The data is de-identified such that all the patients having age more than 89 years were assigned a string value "¿ 89" so we gave them a constant value of 91 approximate to the median of all these '¿89' ages given. We can use this to identify these patients behaviour and prediction after some experiment. Our purpose is to predict mortality and readmission in adults and thus for both the models we remove every record with patient having an age less than 18 years. We drop every record that has any data as "(empty string). Next as logistic regression can not work with categorical data such gender we convert these variables into to numerical form using dummy variables. Finally we drop all the foreign keys and primary keys making the input data ready for training and mapping the labels to 0 and 1 for 'Expired' and 'Alive' labels respectively. We separate the samples with different labels so that we can experiment with the ratio of data from different labels.

C. In-hospital Mortality

We predict only In-hospital mortality as the data is not present for a patient after he leaves the hospital so we do not know the date of death of that patient and thus our prediction is limited to in-hospital mortality. We use Logistic Regression and XGBoost methods for training our models. Initial challenge we faced is that the data is imbalanced, number of samples for 'Expired' label is only 10 percent of the number of samples having 'Alive' label. To remedy this in logistic regression for which we use the sklearns library, adding the parameter class_weight="balanced" does the trick and for XGboost we assign weight to the positive label samples which in our case are the samples having 'Expired' label. We calculate the weight using the formula *weight = samples of negative class / samples of positive class* and give this value to the scale_pos_weight parameter in the XGBoost Classifier from the XGBoost library. We did not normalize the data as normalizing reduced the score consistently.

1) Comorbidity Extraction : The diagnosis table from the dataset is used to extract the ICD-9 codes assigned to all the patients based on their diagnosis. Each of the ICD-9 codes corresponds to a certain diagnosis and has a specific weight. These weights are added up and utilized to calculate the Elixhauser comorbidity scores for each of the patients.

2) Logistic Regression: We perform five fold stratified cross validation, balancing the weight for each label and having L1 penalty. We tried using both L1 and L2 penalties but L1 converged faster and thus we train using L1 penalty. We tried tuning the C parameter using grid search but the results did not improve much and it took a long time to run as the data-set is quite large due to which we did not tune the parameter C and used the default value. we use this training both the baseline model and the retrospective model. In the baseline model we have gender,age and all the features for calculating the SAPS II score which makes 29 features. For retrospective model we have all the features of the baseline model plus the co-morbidity features we get from co-morbidity extraction for calculating the Elixhauser score which makes total 55 features.

3) XGBoost: We use the same features as we did in the logistic regression for both the baseline model and retrospective model. We tried different depths and found that max_depth 3 works best. We try using different number of trees but as we have weighted the samples we get accuracy on 'Expired' labels more when the number of trees are less. When we increase the number of trees and as XGBoost learns to predict the miss classified samples from the previous trees and as the number of samples for 'Alive' class are more the accuracy on 'Expired' class decreases on increasing number of trees. We can see the same in the result for both the baseline and retrospective model.

4) Leave-one-out validation: We tried this so that we can get an overview if the model works over all the hospitals or some. This would tell us if the training is influenced by some hospitals or by the actual input data. There are around 208 different hospitals in this data-set and we performed training by leaving samples from one hospital out, training on the rest of the samples and testing the model on the left out hospitals samples. We did not use AUROC score for measuring this as after pre-processing there were some hospitals which had samples of only one label and thus this measurement would not work on this testing data. We measure the performance using F1 score and report the standard deviation across all the hospitals.

5) Readmission: We compiled the demographics and vital signs obtained by patients of various hospitals each segregated by their respective unitStayID. The patient with the same hospitalStayID and a different unitStayID is categorized as a short term readmission whereas the patient with different HospitalStayID and different UnitStayID are categorized as long term readmission. We used the column in the patient table with unitStayType to decipher these categorizations.

We mapped readmission as 1 and all other types as 0 and

model led the data with logistic regression to calculate the F1 score and the AUROC value to test our model performance. The data was split with the baseline model as obtained from the mortality prediction to check the model robustness. The model parameters were regularized with L2 regularization penalty because the predators were uniformly affecting the response variables with all variables having similar slopes. The L1 regularization was reducing all variable values to 0 at around the same time giving skewed results.

Readmission prediction can be extended to classify both bounce back or short term readmission and long term readmission along with more parameters such as hospital wise cross validation to check hospital dependencies and more features such as ethnicity, weight, height etc to see if these predictors have significant effect on the prediction.

IV. Results

TABLE I
IN-HOSPITAL MORTALITY PREDICTION

	eICU	
	F1 score	AUROC
ABM LR	85.4	75.21
RDFM LR	85.88	75.28

TABLE II
RE-ADMISSIONS PREDICTION

AUROC	0.575

TABLE III
LEAVE-ONE-OUT HOSPITAL CROSS VALIDATION

Average F1 score	85.24
Standard Deviation	0.0886

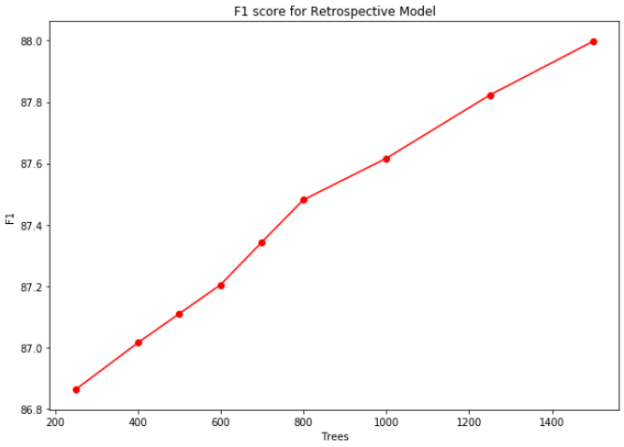

Fig. 1. F1 score of Retrospective model

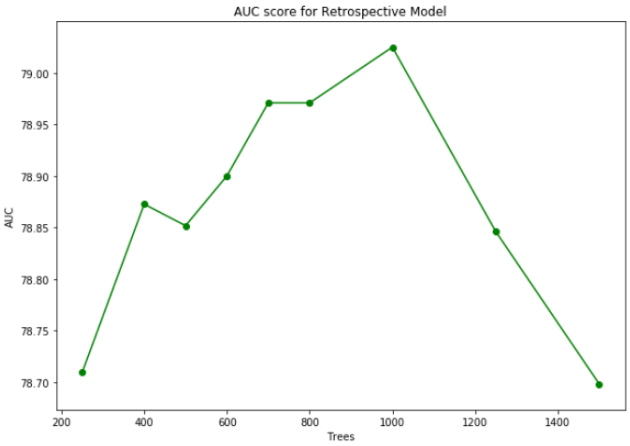

Fig. 2. AUC score for Retrospective model

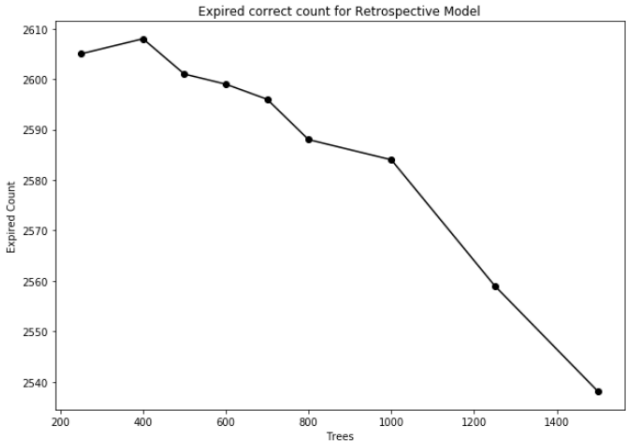

Fig. 3. Expired correct count for Retrospective model

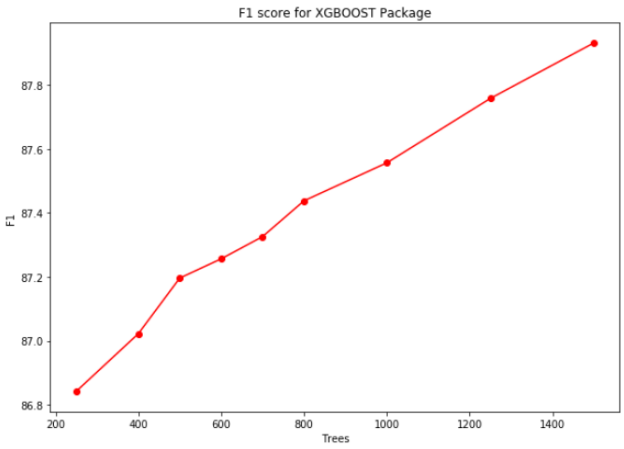

Fig. 4. F1 score for XGBoost package

V. Conclusions

Modern electronic healthcare records contain an increasingly large amount of data including high-frequency signals

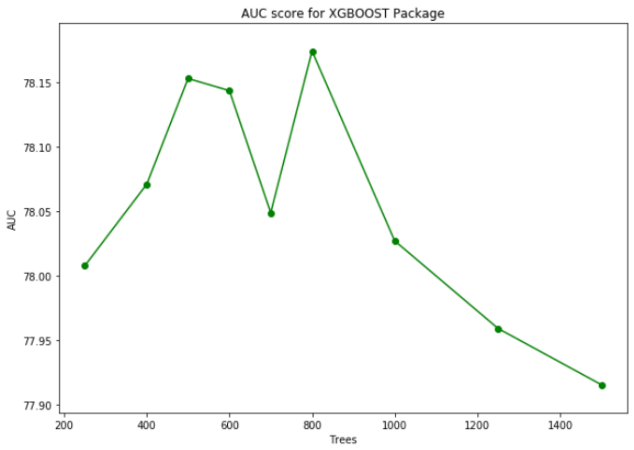

Fig. 5. AUC score for XGBoost package

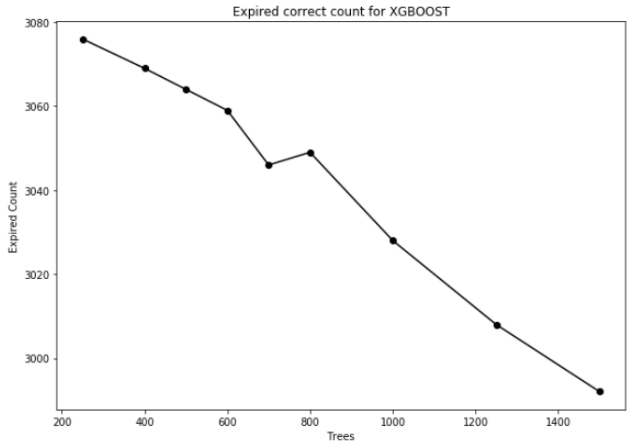

Fig. 6. Expired correct count for XGBoost

from biomedical instrumentation, intermittent results from lab tests, and text from notes. Such voluminous records can make it difficult for care providers to identify the information relevant to diagnose a patient's condition and stratify patients with similar characteristics.This work created a model for predicting in- hospital mortality and readmission using a baseline model and a retrospective model for the eICU database. The work found that the predictions for mortality can be made with an AUROC of 0.75 approximatley for both the models and readmission can calculated with 0.55 which is not good but we aim to improve this in our future works.

The models and results explored in this work could ultimately be useful for interpretable models of disease and mortality.

VI. Future work

In future we plan to include the notes from the caregiver and the textual data in the data-set for more comprehensive analysis of the eICU data-set. We can try region vise and age category vise cross validation which would give us more insights about the mortality in all of these hospitals which are spread across the country. We would try to improve the readmission score as we did not have a lot of time for brain storming and feature extraction for this.

Acknowledgment

'Predicting In-Hospital Mortality using eICU Dataset' was a project for the course CSCE 633 Machine Learning. We would like to thank Dr. Bobak Mortazavi for giving us the opportunity to experiment with and implement this project, and we are very thankful for his guidance and feedback.

References

[1] Pakbin, Arash and Rafi, Parvez and Hurley, Nate and Schulz, Wade and Krumholz, M Harlan and Mortazavi, J Bobak, "Prediction of ICU Readmissions Using Data at Patient Discharge," IEEE 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4932–4935, 2018.

[2] Ghassemi, Marzyeh and Naumann, Tristan and Doshi-Velez, Finale and Brimmer, Nicole and Joshi, Rohit and Rumshisky, Anna and Szolovits, Peter, "Unfolding physiological state: Mortality modelling in intensive care units," ACM Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 75–84, 2014

[3] Johnson, Alistair EW and Pollard, Tom J and Shen, Lu and Li-wei, H Lehman and Feng, Mengling and Ghassemi, Mohammad and Moody, Benjamin and Szolovits, Peter and Celi, Leo Anthony and Mark, Roger G, "MIMIC-III, a freely accessible critical care database," Nature Publishing Group, Scientific data, vol. 3, pp. 160035, 2016

[4] Pollard, Tom J and Johnson, Alistair EW and Raffa, Jesse D and Celi, Leo A and Mark, Roger G and Badawi, Omar, "The eICU Collaborative Research Database, a freely available multi-center database for critical care research," Nature Publishing Group, Scientific data, vol. 5, 2018