Predicting ICU Readmission using Deep Learning

Shagun Gupta
SHAGUN.GUPTA@MAIL.UTORONTO.CA
University of Toronto
27 King's College Circle,
Toronto, ON M5S, Canada

Abstract

Unplanned readmissions to intensive care units (ICU) are costly and result in poor patient outcomes. They are also recognized as a quality indicator of hospital and provider performance. As such, identifying patients at risk of ICU readmission can help make ICU discharges safer while avoiding unnecessary expenses for the hospital. The aim of this study was to develop a tool for predicting patients at high risk for ICU readmission using clinical data from the openly available Medical Information Mart for Intensive Care III (MIMIC-III) database. In particular, this study compares deep neural networks (DNNs) against the baseline measure, Stability and Workload Index for Transfers (SWIFT), on three different readmission time frames: readmission within 48 hours after discharge, 72 hours after discharge and anytime during the same hospitalization. The deep neural networks outperforms the SWIFT score for all readmission timeframes. The results from this study also further provide motivation for the most promising directions for machine learning predictors on this dataset.

1. Introduction

The intensive care unit (ICU) manages the most critically ill patients in the hospital, providing acute, around the clock care. As a result, ICU costs are about three times that of a general medical or surgical ward, reflecting the higher resource consumption and staffing ratios associated with ICUs (Halpern and Pastores (2010)). As such, decisions regarding discharge to ward-level care are made with an effort to maximize these limited ICU resources (Desautels et al. (2017)). However, premature ICU discharge can expose patients to inadequate levels of monitoring and lack of timely interventions resulting in unplanned ICU readmissions. These ICU readmissions are associated with adverse outcomes including preventable clinical deterioration, increased mortality and increased health care costs (Renton et al. (2011), Kastrup et al. (2013)). Various studies indicate that up to 50% or more of such readmissions may be preventable. And while the rates of ICU readmission are lower than rates of overall hospital readmission, the frequency of ICU readmission in the US has increased over the past 20 years (Kastrup et al. (2013), Kramer et al. (2013)). This has lead researchers and administrators to examine the use of ICU readmission as a metric for hospital and provider performance and employ it a means to incentivize efficient and quality, coordinated patient care (Rosa et al. (2015),Desautels et al. (2017)). These factors suggest that identifying which patients are more likely to be readmitted could help make ICU discharges safer with potential to benefit the patient and promote cost savings for the hospital (Chalmers and Black (2014)).

Risk stratification of patients discharged from the ICU is a complex process with many potential challenges. Previously identified predictors of death or ICU readmission include:

length of ICU stay, Glasgow Coma Scale (GCS) score at the time of ICU discharge, mean arterial blood pressure, and ICU admission source (Chen et al. (1998)). Currently, there are limited assessment tools available to predict risk of ICU readmission. The best known tool is the Stability and Workload Index for Transfer (SWIFT) score (Gajic et al. (2008)). Developed specifically to predict readmission to the ICU during the same hospitalization, it relies on five variables: source of ICU admission, length of ICU stay, two respiratory parameters and neurological status as determined by the Glasgow Coma Scale. However, performance of these tools varies widely in the literature (Kastrup et al. (2013)), with only moderate performance (as measured by AUC) in the best case scenario (Rosa et al. (2015)).

The large-scale adoption of electronic medical records and continuous monitoring of ICU patients have generated a wealth of clinical data presenting opportunities to build predictive models for decision support. As such, this study focuses developing a tool for predicting which patients are at risk for readmission to the ICU using deep learning techniques. A typical hospital contains a variety of ICU units including but not limited to the Coronary Care Unit, Medical Intensive Care Unit and Surgical Intensive Care Unit and therefore contains patients with a wide variety of medical conditions. This makes it important to capture complex relationships across many disparate data types that are easily generalizable to the wide variety of patients in the ICU. In this regard, deep learning techniques have recently demonstrated success in capturing rich data representations with high predictive ability without the need to hand engineer a variety of features. Previously, Chen et al. (2015) demonstrated the use deep neural networks (DNNs) for the of multiple disease classification. More recently, deep neural networks were used to improve palliative care planning by predicting all-cause mortality in the ICU in order to identify patients that could benefit from palliative care (Avati et al. (2017)). These studies demonstrate clear motivation for learning readmission prediction models from data using DNNs.

Therefore, for the purpose of this study ICU readmission is defined as one where a patient is discharged from the ICU to general ward-level care only to return to the ICU shortly thereafter. Three timeframes previously reported in the literature (Reini et al. (2012), Brown et al. (2013), Woldhek et al. (2017)) are considered for the prediction task: (1) readmission within 48 hours after discharge, (2) readmission within 72 hours after discharge, and lastly (3) readmission anytime during the same hospitalization. Deep neural networks are compared against other baseline machine learning algorithms and as well as the SWIFT score. For each readmission timeframe this study demonstrates that machine learning techniques outperform the SWIFT index.

2. Related Work

This section covers previous work done in predicting ICU readmission. Overall, there are few studies examining ICU readmission prediction. This may be due to two reasons:

1. ICU populations are diverse and it is not easy to construct a model with good predictive value without a large dataset.
2. The Centers for Medicare & Medicaid Services (CMS) has currently imposed early readmission penalties for 5 patient cohorts, forcing hospitals to focus their time and dollars on those cohorts.

Other studies in the literature have focused on developing ICU readmission prediction models focus on mathematical fuzzy modeling approach. One such study looked at predicting ICU readmission between 24 to 72 hours after ICU discharge. A data mining approach that combined fuzzy modeling with tree search feature selection was utilized to data extracted from the MIMIC-II ICU database. Fuzzy modeling tools allow for approximation of nonlinear systems when there is limited knowledge of the problem being modeled. This approach identified five key variables as predictors of ICU readmission: mean heart rate, mean urine output, mean platelets, and mean non-invasive arterial blood pressure preceding the 24 hours before discharge. It yielded an AUC score of 0.72 0.04, a sensitivity of 0.68 0.02 and a specificity of 0.73 0.03 (Fialho et al. (2012)). However, the authors used a very stringent inclusion criteria for this study leading to an unusually high ICU readmission rate of 13% - almost double that of anywhere reported in the literature.

The same group later published another study where they hypothesized that patient clinical notes can be used augment clinical information and help improve readmission prediction. They compared fuzzy fingerprint text classifiers to traditional text classifiers. A dual fingerprint system based on word occurrence was employed, i.e. one fingerprint was used to represent patients who were readmitted while another to represent patients who were not readmitted. The fuzzy fingerprint model was able to improve (AUC 0.8) on traditional prediction tools that relied solely on physiological data (AUC 0.69) (Curto et al. (2016)). This new method was not compared to the fuzzy model developed in their original paper above, and again the highly selective cohort resulted in an abnormally high readmission rate.

Only one paper focused on ICU readmission using machine learning methods. A study from a single academic, tertiary care hospital in the UK sought to predict readmission to the ICU within 48-hours of discharge AdaBoost, an ensemble machine learning technique which combines the results from multiple weak decision trees in an iterative manner. However, they used both mortality and readmission as the positive class, i.e. any patient deaths within 48 hours were also considered positive cases for readmission. The predictor achieved an AUC score of 0.7095 compared to the baseline SWIFT score with an AUC of 0.6082 on the test data)Desautels et al. (2017)).

3. Cohort

The publicly available MIMIC-III v1.4 (Johnson et al. (2016)) database was used for this study. The database contains electronic medical records from over 50,000 distinct hospital admissions for approximately 38,600 adult patients admitted to the critical care units in the critical care units of the Beth-Israel Deaconess Medical Center in Boston between 2001-2012.

3.1 Cohort Selection

All patient admissions were initially considered for the study. The following exclusion criteria was then applied to each admission to obtain the final cohort:

1. Remove any non-adults, i.e. patients under 18 years of age.
2. Remove invalid admissions, defined as patients with no charted observations or an incomplete administrative record of admission and discharge.

3. Remove any organ donor accounts which are often recorded as readmissions.
4. Remove any ICU stays less than 4 hours. These stays correspond usually correspond to situations such as surgical preparation.
5. Remove any admissions where the patient died during the hospital stay since the goal of this this study is to predict readmission and not mortality.

This resulted in a cohort of 38,225 admissions for the purpose of predicting in-hospital readmission.

3.2 Data Extraction and Preprocessing

For each ICU admission, demographic variables such as patient age, sex, ethnicity and admission features such as type of admission, admission location were extracted. Clinical features extracted included labs, vitals, Glasgow coma score (GCS), SAPS-II scores, OASIS scores, qSOFA scores. Medical comorbidities were represented by the Elixhauser scores (EH) for 30 co-morbidities. A complete list of variables can be found in Table 1 (Appendix). Next, any features with more than 20% missing values were dropped from each cohort. The dataset was then standardized by subtracting the mean (centering) and dividing by the standard deviation (scaling) of each variable. Lastly, dummy variables were created for all categorical variables. Feature selection was performed by selecting the top 75 features in each cohort by mutual information. Selecting more features did not improve classifier results.

3.3 Prediction Task

A positive case for readmission is defined as one in which a patient is discharged from an ICU and then readmitted to an ICU within the same hospitalization. Three time frames were considered for readmission:

1. Readmission to the ICU within 48 hrs of discharge. There were a total of 656 positive readmissions for this time frame, resulting in a readmission rate of 1.72%
2. Readmission to the ICU within 72 hrs of discharge. There were a total of 1079 positive readmissions for this time frame, resulting in a readmission rate of 2.82%
3. Readmission to the ICU anytime within the same hospitalization. There were a total of 2374 positive readmissions for this time frame, resulting in a readmission rate of 6.21%

These readmission rates are consistent with previous studies (Kastrup et al. (2013)).

4. Methods

This section covers the details of the evaluation methodology, algorithms analyzed, and the metrics used for evaluation.

4.1 Methodology

Given an input set of features describing a ICU admission, the goal is to produce an output value indicating whether the patient is likely to be readmitted to ICU or not.

Each record m is represented by its vector of 75 features $x_m = [x_{m_1}, x_{m_2}, \ldots, x_{m_{75}}]$ and its associated label $y_m \in \{\mathit{0}, \mathit{1}\}$ (with 1 indicating that the patient has been readmitted and 0 otherwise). Each algorithm was assessed using *nested cross validation* approach, with a set of 5 train/test set splits (with non overlapping test sets). In the inner loop, the hyperparameters for each model are selected using ten-fold cross-validation on the training data to determine the optimal values with accuracy as an objective. In the outer loop, scores are estimated by averaging test set scores over the 5 dataset splits.

4.2 Algorithms

A total of five different classification algorithms were analyzed for the task of readmission prediction:

1. Fully connected feedforward neural network (FFNN) (Rumelhart et al. (1986)): Deep neural networks, also often called fully connected feedforward neural networks are type a neural networks with multiple nonlinear layers with one prediction layer on the top to solve classification task. The FFNN was implemented using Keras with the following configurations: sigmoid for the output layer activation function, learning rate = 0.001 with the Adam optimizer (Kingma and Ba (2014)), 10 training epochs, batch size = 128, L2 regularization with $\lambda = 0.01$, and the number of layers and units per layer selected from a search within the discrete sets $\{3, 4, 6, 8\}$ and $\{6, 12, 18\}$ respectively. The hidden layer activations tested included ReLU, eLU, SeLU and tanh.

2. SWIFT score (Gajic et al. (2008)): Calculated as defined in the study by Gajic et al., (2008). The maximum possible score is 64, while a score of 15 or more is associated with higher risk of readmission (Kareliusson et al. 2015).

3. Logistic regression (LR) (Fan et al. (2008)): uses logistic function to predict each person's readmission score/probability. The parameter C (the inverse of regularization strength) was selected from a search within the values $\{10^{-15}, 10^{-13}, \ldots, 10^{-1}, 1, 2, 10, 100\}$.

4. Naive Bayes (NB) (Zhang (2004)): implemented using sci-kit learn, version 0.18.1. NB makes an independence assumption between features (which is often inaccurate) to model the conditional probability.

5. Majority Class (MC): a trivial baseline classifier that always predicts the most frequent label in the training set. Note: when a probability of a label is required for a test example, it ignores the test example and reports the overall probability of the label in the training dataset.

4.3 Evaluation Metrics

Each algorithm was evaluated using the standard metrics including: Accuracy, Precision, Recall, F1-Score, defined as follows:

$$
\begin{aligned}
Precision =& \frac{tp}{tp + fp}, && (1)\\
Recall =& \frac{tp}{tp + fn}, && (2)\\
F1{-}Score =& \frac{2 \times Precision \times Recall}{Precision + Recall} && (3)
\end{aligned}
$$

where tp, fp, and fn indicate respectively the amount of true positives, false positives, and false negatives.

Average Precision (AP), a commonly ranking metric, was also used for evaluation. AP can be defined as:

$$AveP = \sum_{k=1}^{n} P(k)\Delta r(k) \tag{4}$$

where k is the rank in the sequence of records ranked using readmission probabilities, n is the number of records, $P(k)$ is the precision at cut-off k in the list, and $\Delta r(k)$ is the change in recall from records $k-1$ to k.

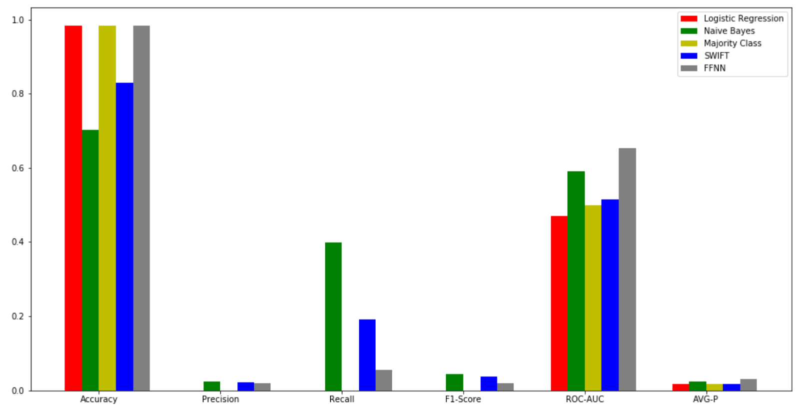
Lastly, the area under the ROC curve (AUROC) was also examined.

5. Results

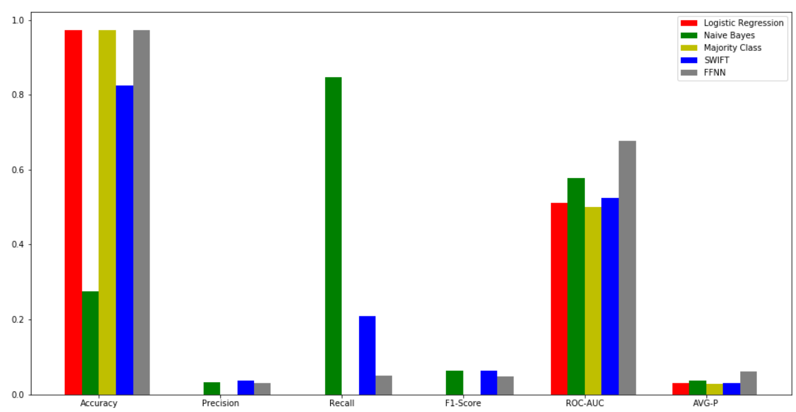
The FFNN (deep neural network) outperformed the SWIFT baseline for each readmission timeframe. Performance characteristics for the FFNN model, the SWIFT score and other ML algorithms are summarized in Figure 1. Average ROC curves are presented in Figure 2. ROC curves show the true positive rate (the fraction of positive cases which receive a positive label) as a function of the false positive rate (the fraction of negative cases receiving a positive label). From these results, the following observations can be made for each of the datasets:

Readmission within 48-hours:

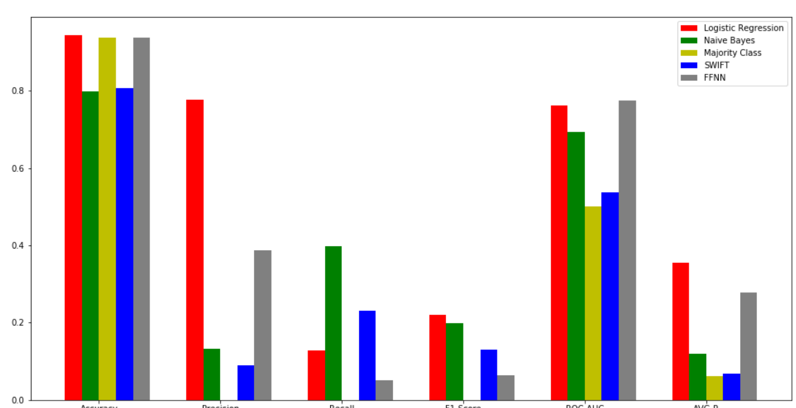
1. The SWIFT score (AUC 0.52) score performs no better than the majority classifier (AUC 0.5). The FFNN (AUROC 0.65), however, performs significantly better than both the SWIFT score and the majority classifier. Naïve Bayes (AUROC 0.59) also outperforms the majority classifier, unlike LR (AUROC 0.47) which fails to improve over the baseline Despite the improvement in AUC for the FFNN over the SWIFT score, the precision and recall remain low in addition to a low average precision score, i.e. poor ranking.
2. When comparing accuracy, the FFNN is comparable to the majority class algorithm. While in most cases this may be a cause for concern regarding whether any of the classifiers have actually learned to predict readmission, the FFNN outperforms MC on AUC indicating strong evidence of learning. If the predictions were random, the AUC for the FFNN would be 0.5, which is the case for the majority classifier. This is true for all datasets including the 72-hour readmission and anytime readmission.



(a) Readmission within 48 hrs

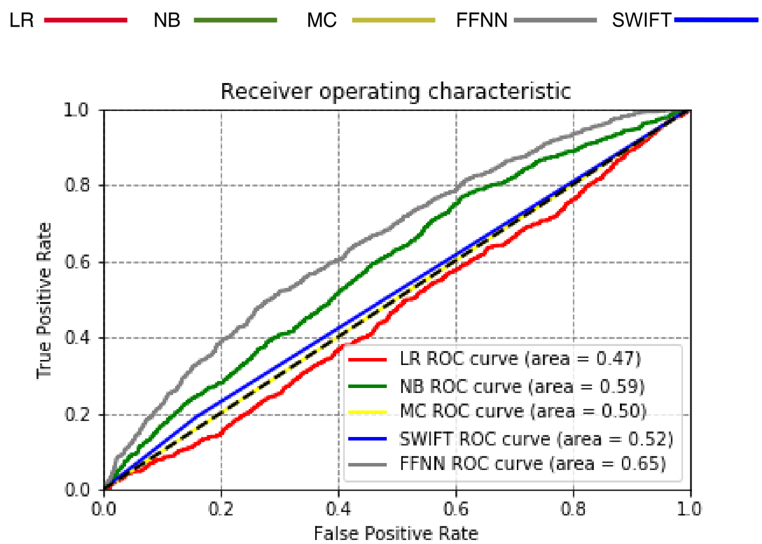


(b) Readmission within 72 hrs

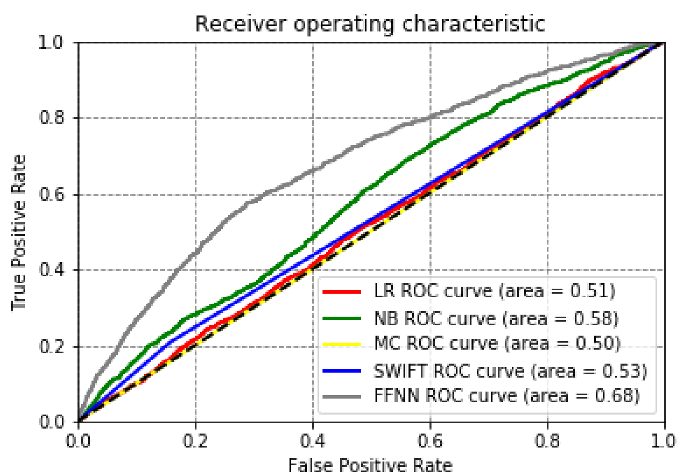


(c) Readmission anytime within same hospitalization

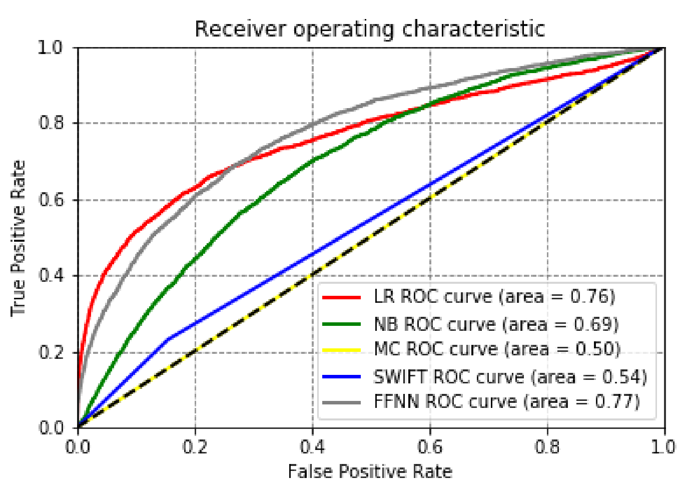
Figure 1: Comparison of accuracy, precision, recall, F1-score, ROC-AUC, and Average Precision for all three readmission datasets.



(a) Readmission within 48 hrs



(b) Readmission within 72 hrs



(c) Readmission anytime within same hospitalization

Figure 2: Receiver operating characteristic (ROC) curve with AUCs shown in the legend.

Readmission within 72-hours:

1. Similar to the results for readmission within 48 hours, the SWIFT score (AUC 0.54) does not do much better than the majority classifier (AUC 0.5), while the FFNN (AUROC 0.68) has the best performance. However, again, despite this improvement, the F1 score and average precision scores both remain very low. This is likely a result of the highly imbalanced classes. Naïve Bayes performs well on recall (over 0.8), however precision remains poor, likely because the NB algorithm overcompensates for the positive class.

Readmission anytime within the same hospitalization:

1. The SWIFT score (AUCROC 0.56) exhibits poor performance, consistent with previous studies (Kastrup et al. (2013)). Both the FFNN and LR show considerable improvement over the baseline (AUROC 0.77 and 0.76 respectively). NB performance is consistent on all datasets (AUCROC 0.69). However, when examining precision, LR outperforms all other algorithms by a large margin. This is reflected in Figure 2, where the ROC curve for the LR is more towards the upper left region than that of the FFNN. It also performance well for ranking with an AP score of 0.36. LR also performance well for ranking with an average precision score of 0.36, followed by the FFNN.

Finally, to explore the relationship between features and the readmission outcome, feature analysis was performed. A general method for measuring the amount of information a feature xk provides w.r.t. predicting a class label y is to calculate its mutual information (MI) I(xk,y). The results for each readmission time frame are in Tables 1, 2 and 3. The top ranking features for readmission within 48 hours and readmission within 72 hours are almost identical with admission type (emergency), gender, endotracheal intubation (Y/N) and the first GCS (Glasgow Coma Score) score components for eyes and verbal as some of the highly ranked features. The top ranking features for readmission anytime during the same hospitalization were dominated largely by lab values. Interestingly, only three of the five predictors for the SWIFT score appeared in the top 10 ranked features for any dataset: admission type (emergency) feature (all readmission timeframes), ICU length of stay (any type readmission) and lastly some component of the GCS score (all readmission timeframes). The respiratory features, last measured Pao2/Fio2 ratio and last arterial blood gas Paco2, did not appear in the top 10 ranked features for any readmission dataset.

6. Discussion

This study examined the task of predicting which patients are likely to be readmitted to the ICU within 48 hours of discharge, 72 hours of discharge or anytime during the same hospitalisation. To current knowledge, this is the first study to use deep neural networks to predict ICU readmission. Deep learning methods were able to outperform the standard ICU readmission index, the SWIFT score, for all readmission timeframes considered in this study. However, looking at individual readmission timeframes, the FFNN still only had average performance at best for 48-hour readmission and modest performance for 72-hour readmission. The FFNN performed best for readmission anytime during the same

Table 1: Ranking of the most important features using Mutual Information for 48 hr readmission

Rank	**Feature**	**Type**
1	admission_type_EMERGENCY	Type of admission: emergency
2	race	Race
3	gender_M	Gender
4	endotrachflag	Endotracheal intubation (Y/N)
5	insurance_Medicare	Insurance type: Medicare
6	first_gcseyes	First GCS score for eyes
7	first_gcsverbal	FIrst verbal GCS score
8	spo2_max	Lab Value (max blood oxygen saturation)
9	last_gcsverbal	Last verbal GCS score
10	first_bicarbonate	Lab Value

Table 2: Ranking of the most important features using Mutual Information for 72 hr readmission

Rank	**Feature**	**Type**
1	admission_type_EMERGENCY	Type of admission: emergency
2	race	Race
3	first_hematocrit	Lab Value (first hematocrit measurement)
4	first_gcseyes	First GCS score for eyes
5	endotrachflag	Endotracheal intubation (Y/N)
6	first_gcsmotor	First motor GCS score
7	gender_M	Gender
8	first_hemoglobin	Lab Value
9	first_bicarbonate	Lab Value
10	insurance_Medicare	Insurance Type

Table 3: Ranking of the most important features using Mutual Information for readmission anytime within same hospitalization

Rank	Feature	Type
1	first_hematocrit	Lab Value
2	first_hemoglobin	Lab Value
3	admission_type_EMERGENCY	Type of admission: emergency
4	last_sodium	Lab Value
5	last_inr	Lab Value (international normalized ratio)
6	iculos	ICU length of stay
7	last_bun	Lab Value (blood urea nitrogen)
8	last_creatinine	Lab Value
9	first_inr	Lab Value (international normalized ratio)
10	first_gcsverbal	First verbal GCS score

hospitalization, showing good predictive power, with an AUROC of 0.77. However, it is important to note that logistic regression performed just as well in this case while also showing much better precision. This suggests that a more complex models do not necessarily guarantee better results and that more traditional machine learning algorithms may be sufficient for many cases of readmission prediction.

The top features as determined by mutual information for each readmission timeframe contained a mix of demographic and clinical variables. Three of the five predictors used in the SWIFT score appeared in the top 10 ranked features for all readmission timeframes. The top 10 ranked features for anytime-readmission was dominated by lab values - this is not surprising given that lab values for hemoglobin, bicarbonate, creatinine, urea are have previously been found to be good predictors for ICU mortality (Ding et al. (2015), Iqbal et al. (2015)), suggesting some level of generalizability for these predictors across different datasets.

There are a number of possible avenues to improve the results of this preliminary study. First of all, in the case of ICU readmission the classes (readmitted or not readmitted) are highly imbalanced with a prevalence of about 2-6%, making it a difficult classification task. Previous studies examining ICU readmission have tackled this imbalance by combining prediction of ICU readmission and mortality into one class: adverse outcomes after ICU discharge within a given timeframe (Desautels et al. (2017), Badawi and Breslow (2012)). Other methods that could be employed include under sampling and oversampling, and more advanced techniques such as SMOTE which involves creating new examples of the minority class by interpolating between existing ones (Chawla et al. (2002)).

Second, much like any hospital ICU, the MIMIC-III ICU database contains a rich clinical notes dataset which can be used to improve classifier performance. The narrative in clinical notes is able to provide clinicians with the most important aspects of a patients care and also contains various variables not found in structured fields such as clinical trends and past medical history which are known predictors of ICU readmission(Gajic et al. (2008)). Furthermore, previous studies have demonstrated that using representations of clinical notes

(Ghassemi et al. (2014)) or convolutional neural networks for clinical notes (Grnarova et al. (2016)) along with standard physiological measurements can improve performance in tasks such as ICU mortality considerably, making it a promising option for improving performance for readmission prediction tasks as well.

Lastly, when considering short term ICU readmission, for example within 48 hours of discharge, physiological trends play a key role in identifying whether the patient is likely to be readmitted or not (Churpek et al. (2016)). Recurrent neural networks (RNNs) have shown tremendous success in time series modeling for clinical tasks including disease prediction from longitudinal lab values (Razavian et al. (2016) ICU intervention prediction (Suresh et al. (2017)) and pediatric ICU mortality (Aczon et al. (2017)). Therefore, RNNs are a method that may be useful for time series modeling using physiological variables for readmission prediction tasks.

Ultimately, ICU readmission is far more complex problem than, for example, ICU length of stay prediction. Discharging a patient from the ICU involves assessing whether a patients main condition has been resolved sufficiently while relying on investigatory results and monitoring observations over time. Therefore, in order to improve over a physician's judgment requires an algorithm that can take into account subtle and multidimensional time trends.

Nonetheless, these results are an interesting start for using deep neural networks to predict risk of readmission in ICU patients and future work to expand this will enable more robust models.

References

M Aczon, D Ledbetter, L Ho, A Gunny, A Flynn, J Williams, and Randall Wetzel. Dynamic mortality risk predictions in pediatric critical care using recurrent neural networks. 01 2017.

Anand Avati, Kenneth Jung, Stephanie Harman, Lance Downing, Andrew Ng, and Nigam H. Shah. Improving palliative care with deep learning. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017. doi: 10.1109/bibm.2017.8217669.

Omar Badawi and Michael J. Breslow. Readmissions and death after icu discharge: Development and validation of two predictive models. *PLoS ONE*, 7(11), Jul 2012. doi: 10.1371/journal.pone.0048758.

Sydney E. S. Brown, Sarah J. Ratcliffe, and Scott D. Halpern. An empirical derivation of the optimal time interval for defining icu readmissions. *Medical Care*, 51(8):706?714, 2013. doi: 10.1097/mlr.0b013e318293c2fa.

James D. Chalmers and Euan Black. Critical care transition and prevention of icu readmissions. *Critical Care Medicine*, 42(1):216?217, 2014. doi: 10.1097/ccm.0b013e3182a520a0.

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357, June 2002. ISSN 1076-9757. URL `http://dl.acm.org/citation.cfm?id=1622407.1622416`.

Liddy M. Chen, Claudio M. Martin, Sean P. Keenan, and William J. Sibbald. Patients readmitted to the intensive care unit during the same hospitalization. *Critical Care Medicine*, 26(11):1834?1841, 1998. doi: 10.1097/00003246-199811000-00025.

Matthew M. Churpek, Richa Adhikari, and Dana P. Edelson. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*, 102:1?5, 2016. doi: 10.1016/j.resuscitation.2016.02.005.

Sergio Curto, Joao P. Carvalho, Catia Salgado, Susana M. Vieira, and Joao M. C. Sousa. Predicting icu readmissions based on bedside medical text notes. *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016. doi: 10.1109/fuzz-ieee.2016.7737956.

Thomas Desautels, Ritankar Das, Jacob Calvert, Monica Trivedi, Charlotte Summers, David J Wales, and Ari Ercole. Prediction of early unplanned intensive care unit readmission in a uk tertiary care hospital: a cross-sectional machine learning approach. *BMJ Open*, 7(9), 2017. doi: 10.1136/bmjopen-2017-017199.

Yew Y. Ding, Boris Kader, Cindy L. Christiansen, and Dan R. Berlowitz. Hemoglobin level and hospital mortality among icu patients with cardiac disease who received transfusions. *Journal of the American College of Cardiology*, 66(22):2510?2518, 2015. doi: 10.1016/j.jacc.2015.09.057.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *J. Mach. Learn. Res.*, 9:1871–1874, June 2008. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1390681.1442794`.

A.s. Fialho, F. Cismondi, S.m. Vieira, S.r. Reti, J.m.c. Sousa, and S.n. Finkelstein. Data mining using clinical physiology at discharge to predict icu readmissions. *Expert Systems with Applications*, 39(18):13158–13165, 2012. doi: 10.1016/j.eswa.2012.05.086.

Ognjen Gajic, Michael Malinchoc, Thomas B. Comfere, Marcelline R. Harris, Ahmed Achouiti, Murat Yilmaz, Marcus J. Schultz, Rolf D. Hubmayr, Bekele Afessa, J Christopher Farmer, and et al. The stability and workload index for transfer score predicts unplanned intensive care unit patient readmission: Initial development and validation*. *Critical Care Medicine*, 36(3):676?682, 2008. doi: 10.1097/ccm.0b013e318164e3b0.

Marzyeh Ghassemi, Tristan Naumann, Finale Doshi-Velez, Nicole Brimmer, Rohit Joshi, Anna Rumshisky, and Peter Szolovits. Unfolding physiological state: Mortality modelling in intensive care units. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 14*, 2014. doi: 10.1145/2623330.2623742.

Paulina Grnarova, Florian Schmidt, Stephanie L. Hyland, and Carsten Eickhoff. Neural document embeddings for intensive care patient mortality prediction, 2016.

Neil A. Halpern and Stephen M. Pastores. Critical care medicine in the united states 2000?2005: An analysis of bed numbers, occupancy rates, payer mix, and costs*. *Critical Care Medicine*, 38(1):65?71, 2010. doi: 10.1097/ccm.0b013e3181b090d0.

A Iqbal, I Welters, R Kolamunnage-Dona, C Toh, and C Downey. Developing a laboratory-based score to predict mortality in patients admitted to the icu. *Critical Care*, 19(Suppl 1), 2015. doi: 10.1186/cc14627.

Alistair E.w. Johnson, Tom J. Pollard, Lu Shen, Li-Wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, Roger G. Mark, and et al. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3:160035, 2016. doi: 10.1038/sdata.2016.35.

Marc Kastrup, Robert Powollik, Felix Balzer, Susanne Röber, Robert Ahlborn, Vera Von Dossow-Hanfstingl, Klaus D. Wernecke, and Claudia D. Spies. Predictive ability of the stability and workload index for transfer score to predict unplanned readmissions after icu discharge*. *Critical Care Medicine*, 41(7):1608?1615, 2013. doi: 10.1097/ccm.0b013e31828a217b.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL `http://arxiv.org/abs/1412.6980`.

Andrew A. Kramer, Thomas L. Higgins, and Jack E. Zimmerman. The association between icu readmission rate and patient outcomes*. *Critical Care Medicine*, 41(1):24?33, 2013. doi: 10.1097/ccm.0b013e3182657b8a.

Narges Razavian, Jake Marcus, and David Sontag. Multi-task prediction of disease onsets from longitudinal lab tests. *Proceedings of the 1st Machine Learning for Healthcare Conference*, 2016.

Kirsi Reini, Mats Fredrikson, and Anna Oscarsson. The prognostic value of the modified early warning score in critically ill patients. *European Journal of Anaesthesiology*, 29(3): 152?157, 2012. doi: 10.1097/eja.0b013e32835032d8.

J. Renton, D. V. Pilcher, J. D. Santamaria, P. Stow, M. Bailey, G. Hart, and G. Duke. Factors associated with increased risk of readmission to intensive care in australia. *Intensive Care Medicine*, 37(11):1800?1808, 2011. doi: 10.1007/s00134-011-2318-x.

Regis Goulart Rosa, Cintia Roehrig, Roselaine Pinheiro De Oliveira, Juçara Gasparetto Maccari, Ana Carolina Peçanha Antônio, Priscylla De Souza Castro, Felippe Leopoldo Dexheimer Neto, Patrícia De Campos Balzano, and Cassiano Teixeira. Comparison of unplanned intensive care unit readmission scores: A prospective cohort study. *Plos One*, 10(11), 2015. doi: 10.1371/journal.pone.0143127.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533, 1986.

Harini Suresh, Nathan Hunt, Johnson Alistair, Leo Anthony Celi, Peter Szolovits, and Marzyeh Ghassemi. Clinical intervention prediction and understanding using deep networks, 2017.

Annemarie L. Woldhek, Saskia Rijkenberg, Rob J. Bosman, and Peter H.j. Van Der Voort. Readmission of icu patients: A quality indicator? *Journal of Critical Care*, 38:328?334, 2017. doi: 10.1016/j.jcrc.2016.12.001.

Harry Zhang. The optimality of nave bayes. In *In FLAIRS2004 conference*, 2004.

Appendix A.

Variable	**Variable**
Patient Variables	Age, Gender, Ethnicity, Height, Marital Status, Religion, Weight
Admission variables	Admission type, Admission location, Length of ER stay, Length of ICU stay
Labs and Vitals (First, Last, Max, Min, Mean Values)	Anion gap, Albumin, Bicarbonate,Bilirubin, Blood urea nitrogen, Creatinine,Chloride, Glucose, Diastolic blood pressure, Heart rate, Hematocrit,Hemoglobin, Lactate, Mean blood pressure, Platelet, Potassium, Partial thromboplastin time (PTT), International Normalized Ratio (INR), Peripheral Oxygen Saturation, Prothrombin time,Respiratory rate, Sodium, Systolic blood pressure, Temperature, White blood cell count, Pulse oximetry, Oxygen saturation, Partial pressure of oxygen, Partial pressure of carbon dioxide, PaO2/FiO2 Ratio,Blood pH, Base excess, Total CO2, Carboxyhemoglobin
Scores	Elixhauser scores, Glasgow coma scale (total, verbal, motor, eyes), SAPS-II scores, OASIS scores, qSOFA scores
Other	Medical service under which patient is admitted, number of unique drugs given,Did patient receive: dialysis, ventilation, vasopressors or intubation, Total urine output