



Domain Adaptation Using Convolutional Autoencoder and Gradient Boosting for Adverse Events Prediction in the Intensive Care Unit

Yuanda Zhu¹, Janani Venugopalan², Zhenyu Zhang^{3,4}, Nikhil K. Chanani⁵, Kevin O. Maher⁵ and May D. Wang^{2*}

¹ School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, United States, ² Biomedical Engineering Department, Georgia Institute of Technology, Emory University, Atlanta, GA, United States, ³ Biomedical Engineering Department, Georgia Institute of Technology, Atlanta, GA, United States, ⁴ Department of Biomedical Engineering, Peking University, Beijing, China, ⁵ Pediatrics Department, Emory University, Atlanta, GA, United States

OPEN ACCESS

Edited by:

Enrico Capobianco,
University of Miami, United States

Reviewed by:

Hadi Khorshidi,
The University of Melbourne, Australia
Michael Kirley,
The University of Melbourne, Australia

*Correspondence:

May D. Wang
maywang@gatech.edu

Specialty section:

This article was submitted to
Medicine and Public Health,
a section of the journal
Frontiers in Artificial Intelligence

Received: 12 December 2020

Accepted: 23 February 2022

Published: 11 April 2022

Citation:

Zhu Y, Venugopalan J, Zhang Z,
Chanani NK, Maher KO and
Wang MD (2022) Domain Adaptation
Using Convolutional Autoencoder and
Gradient Boosting for Adverse Events
Prediction in the Intensive Care Unit.
Front. Artif. Intell. 5:640926.
doi: 10.3389/frai.2022.640926

More than 5 million patients have admitted annually to intensive care units (ICUs) in the United States. The leading causes of mortality are cardiovascular failures, multi-organ failures, and sepsis. Data-driven techniques have been used in the analysis of patient data to predict adverse events, such as ICU mortality and ICU readmission. These models often make use of temporal or static features from a single ICU database to make predictions on subsequent adverse events. To explore the potential of domain adaptation, we propose a method of data analysis using gradient boosting and convolutional autoencoder (CAE) to predict significant adverse events in the ICU, such as ICU mortality and ICU readmission. We demonstrate our results from a retrospective data analysis using patient records from a publicly available database called Multi-parameter Intelligent Monitoring in Intensive Care-II (MIMIC-II) and a local database from Children's Healthcare of Atlanta (CHOA). We demonstrate that after adopting novel data imputation on patient ICU data, gradient boosting is effective in both the mortality prediction task and the ICU readmission prediction task. In addition, we use gradient boosting to identify top-ranking temporal and non-temporal features in both prediction tasks. We discuss the relationship between these features and the specific prediction task. Lastly, we indicate that CAE might not be effective in feature extraction on one dataset, but domain adaptation with CAE feature extraction across two datasets shows promising results.

Keywords: intensive care units, clinical decision support, mortality prediction, gradient boosting, convolutional autoencoder, domain adaptation (DA)

INTRODUCTION

Each year, over 5 million patients have admitted to ICUs in the United States (Vranas et al., 2018), with an average mortality between 8 and 10% (Wu et al., 2002). The leading causes of mortality are cardiovascular collapses (Benjamin et al., 2017), multi-organ failures (Marshall et al., 1995), and sepsis (Kissoon et al., 2016). Studies have shown that prolonged

hospital stays lead to increased life-threatening outcomes, which are associated with adverse events and other risks in the ICU environment (Celi et al., 2011; Hunziker et al., 2012), such as severe infection (Fagon et al., 1994; Ribas et al., 2011), cardiac arrest (Nemati et al., 2011), extended invasive ventilation (Fagon et al., 1993; de Rooij et al., 2006), mortality within 1 year (Mandelbaum et al., 2011; Celi et al., 2012), ICU/hospital readmissions (Fialho et al., 2012), acute kidney injury (Mandelbaum et al., 2011; Celi et al., 2012), and hypotension (Lee and Mark, 2010; Hug et al., 2011; Lee et al., 2012). Studies also report that these complications will also lead to a significant increase in the costs incurred (Wheeler et al., 2011; Kwon et al., 2012; Hutchinson et al., 2013). Hence, accurate prediction of adverse patient endpoints would allow for improved resource allocation.

Research on the analysis of patient data to predict adverse events, such as ICU mortality and ICU readmission, has mainly used probabilistic models. Logistic regression (Celi et al., 2012; Fuchs et al., 2012; Lee et al., 2012; Venugopalan et al., 2017), Cox regression (Fuchs et al., 2012), and artificial neural networks (Wong and Young, 1998; Lee and Mark, 2010; Celi et al., 2012) are the most common models used in the analysis of healthcare data (Goldstein et al., 2017). However, these models suffer from inherent issues, particularly, their basis on using a snapshot of the data available to make longitudinal predictions. Patient data itself are temporal in nature; hence, a temporal analysis of these data should be performed for a more appropriate health prediction.

The temporal models commonly seen in the literature include models, such as sequence analysis (Wang et al., 2008; Batal et al., 2011; Tao et al., 2012; Casanova et al., 2015; Syed and Das, 2015), association rule mining (Bellazzi et al., 2011; Casanova et al., 2015; Yang and Yang, 2015), temporal Cox regression (Warner et al., 2013; Cai et al., 2015; McCoy et al., 2015), and clustering (Toddenroth et al., 2014; Choi et al., 2017). Sequence analysis and association rule mining-based studies require extensive user input for identifying specific features whose patterns of correlation can be studied with respect to the target variable. In addition, they are not amenable for discerning relationships and patterns contributing to adverse events, from a large number of features. Regression- (Singh et al., 2015) and clustering-based (Doshi-Velez et al., 2014) studies use the information within a specific time interval for analysis. These studies do not account for the differing length of available data for different patients. Cox regression also does not account for the dependency between the consecutive time points. Graphical methods by Liu et al. use Gaussian Processes (GPs) for time-series analysis (Liu et al., 2013). Their assumption is that the data are piecewise linear and use only the GP coefficients for classification. Such models make the assumptions that ICU data can be approximated using piecewise GPs. Stiglic et al. (2013) used past recordings for a single patient to make predictions about a future time instant using the least absolute shrinkage and selection operator (LASSO) regression. The parameters of these models are trained for each individual patient and do not make use of the information, which can be learned from large databases consisting of multiple patients. Such models not only require the user to train the model for each patient but also tend

to over-fit the data. In addition, these models do not tell the clinicians if the patients are improving over time. Yu et al. (2011) generates individual survival curves by using a series of logistic regression models to calculate the hazard at each time instance. This approach is not only very computationally intensive but also does not account for the variation in the duration of ICU data.

Transfer learning has been widely used in different clinical decision support systems, especially on medical image processing tasks (Han et al., 2018; Lee et al., 2018; Cheplygina et al., 2019; Choudhary et al., 2020). The goal of transfer learning is to use an external dataset to improve classification or segmentation results on the local dataset. There are two major concepts related to transfer learning: domain and task. Medical images are considered from different domains if they are generated by different scanners or follow different scanning protocols. Therefore, datasets collected from multiple sites or providers are considered from different domains. Task refers to the specific machine-learning task, such as disease classification (Chen et al., 2017; Hussein et al., 2017; Li et al., 2018), lesion detection (Hwang and Kim, 2016; Elmahdy et al., 2017; Murthy et al., 2017), and tumor segmentation (Kandemir, 2015; Huynh et al., 2016).

One category of transfer learning is “same domain, different tasks.” Studies in this category would use the same dataset for multiple tasks. A common strategy is feature transfer for multi-task learning that proposed models aim to learn common features across different tasks. Joint learning on task-independent features can effectively mitigate the overfitting problem by regularizing the classifiers and/or increasing sample size. Another category of transfer learning is “different domains, same task.” Imaging data collected from different domains assumably have different sample data distribution; hence, domain adaptation is needed. The objective of domain adaptation is to transfer the knowledge across different domains by learning domain-invariant features transformation. Image-to-image domain transformation (Isola et al., 2018) achieves pixel-level mapping between source and target images using generative models, such as generative adversarial networks (GAN) (Goodfellow et al., 2014). Another approach is latent feature space transformation that learns the domain-invariant features representation by transforming source and target domain images into the shared latent space. Latent feature space transformation has three subcategories of methods: reconstruction-based (autoencoder) (Bousmalis et al., 2016; Ghifary et al., 2016), divergence minimization (divergence metric) (Damodaran et al., 2018; Kang et al., 2019; Rozantsev et al., 2019), and adversarial learning (discriminator) (Lafarge et al., 2017).

Despite its success in image processing (Han et al., 2018; Lee et al., 2018; Choudhary et al., 2020), transfer learning has a limited impact on ICU data for predicting mortality risk and ICU readmission. Gupta et al. (2020) proposed multi-task transfer learning using recurrent neural network (RNN) on Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC)-III data to identify phenotypes and predict in-hospital mortality. The proposed RNN is pretrained on predicting phenotypes, before transferring the learned knowledge on the mortality prediction task. Although the author demonstrated the effectiveness of multi-task transfer learning, there are two shortcomings of the

TABLE 1 | Summary of Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) and Children's Healthcare of Atlanta (CHOA) data.

Data and features	MIMIC II	CHOA	Data type
Sample size	40,416 patient records	5,739 patient records	
Demographics	Gender, age, height, weight, ethnicity, comorbidity	DOB, gender, age, height, weight, ethnicity, religion, date of death, comorbidity with other diseases	Non-temporal
Lab data	Urea, albumin, bilirubin, creatinine, sodium, potassium, calcium	Urea, albumin, bilirubin, creatinine, sodium, potassium, calcium	Temporal
Chart data	Heart rate, blood pressure, NBP, CVP, SaO ₂ , arterial PH, arterial PaCO ₂ , arterial PaO ₂		Temporal
Microbiology		Types of microbes, amount of microbes, dilution	Non-temporal
Medication data		Medication and IV administered, dosage, duration time, concentrations and rate of administration, composition of IV imposed	Non-temporal

work: (1) the proposed RNN model is close to a vanilla model. The complicated RNN models, such as gated recurrent units (GRUs) or long short-term memory (LSTM), could be used to improve the pipeline. (2) There is no preprocessing on the MIMIC-III data, as the author directly worked on the benchmark data. Another way to do transfer learning is to train the model on two datasets to improve prediction performance. Desautels et al. (2017) concatenated MIMIC-III data (external dataset) with UK ICU data (internal dataset) for joint training before evaluating the model on the internal testing set. Multi-source transfer learning is novel on ICU adverse outcome prediction, but the proposed model suffers from very low specificity (0.5917) and poor f1-score (0.1321).

In this study, we propose a retrospective study of adult populations to discover factors indicative of adverse events, such as ICU mortality and 30-day ICU readmissions, using a gradient boosting algorithm for classification and convolutional autoencoder (CAE) for domain adaptation. We apply CAE as domain adaptation for learning domain-invariant latent feature representation from two ICU datasets. We aim to improve model classification performance on our internal dataset after learning from the external dataset.

We summarize our major contribution in several folds:

- We demonstrate that gradient boosting is effective in both the mortality prediction task and the ICU readmission prediction task.
- We use gradient boosting to identify top-ranking temporal and non-temporal features in both the mortality prediction task and the ICU readmission prediction task. We discuss the relatedness of these features with their corresponding prediction task.
- We indicate that CAE might not be effective in feature extraction on one dataset, but domain adaptation with CAE feature extraction across two datasets shows promising results.

We structure the remainder of this article as follows. First, a short description of our data source is followed by a detailed description of the preprocessing and data mining approaches in Section Materials and Methods. Experiments are described in

Section Experiments. Results and discussion are presented in Section Results and Discussion. Finally, the conclusion and future directions are summarized in Section Conclusion.

MATERIALS AND METHODS

In this study, we performed the classification of ICU patients into high risk and low risk for adverse events using random forest, gradient boosting, and CAE. We demonstrated our results using a retrospective data analysis of adult ICU data from the MIMIC-II database and local ICU data from Children's Healthcare of Atlanta (CHOA). We used different machine-learning models to determine patient's factors, which contribute to adverse consequences, such as ICU mortality and 30-day ICU readmission. These endpoints are particularly interesting since they provide the basis for the long-term prediction of adverse events.

Data and Preprocessing

MIMIC Data

Multi-parameter Intelligent Monitoring in Intensive Care-II is a public ICU data repository containing over 40,000 ICU stay (32,331 adults and, 8,085 neonatal) records (Saeed et al., 2011). We performed retrospective data analysis using adult patient data (>16 years). Each ICU stay record consists of the patient's demographic information, diagnosis, chart events, medication intake events, microbiology events, etc. (example features in **Table 1**). Each patient record consists of features which are either static (does not change over the entire duration of the patient's ICU stay) or temporal (changing in time). From the total of over 13,000 features, we ranked the features by the frequency of measurement. Using the top 2,000 most frequent features, we picked 87 features on the basis of clinician judgment (we used only 84 features for ICU readmission to avoid information leakage). The included features that covered clinical measurements, lab results administrative data, comorbidities, and other diagnostic procedures.

For this analysis, the data from each feature were binned into binning intervals of 6 h. After binning the data into intervals

of 6 h, data had $87 \pm 21\%$ of missing data. The missing data were divided into multiple types of missing data known in the statistical literature as “missing completely at random (MCAR),” “missing at random (MAR),” and “missing not at random (MNAR).” Then each type was imputed differently using the techniques described in a previous paper on data imputations (Venugopalan et al., 2019). MNAR data were imputed using Student’s t-copulas and MAR data were imputed using expectation maximization (EM) after clustering.

We then proceeded to perform feature selection and classification on the patient data sequences for identifying patients at risk for adverse events, such as mortality in the ICU and 30-day ICU readmission. In this dataset, there were 2,334 patient records with mortality during the ICU stay and 29,997 patient records of successful discharge from the ICU. Similarly, 7,787 patients’ records had ICU readmission within 30 days and 24,544 patients did not relapse into the ICU within 30 days.

CHOA Data

The other dataset is from CHOA containing 5,739 patient records spanning an 11-month period. The visits spanned pediatric ICU, neonatal ICU, and cardiac ICU. As shown in **Table 1**, each ICU stay record consists of the patient’s demographic information (e.g., gender and age of admission), diagnosis [e.g., International Classification of Diseases (ICD-9) codes], birth-related events (e.g., birth weight, head circumference, gestation weeks), microbiology events (e.g., microbes in blood or serum), chart events (e.g., heart rate), medication intake events, and clinical records (e.g., pulse oximetry) collected from bedside monitors, averaged over each minute.

The data columns are binary, categorical, and quantitative from which we extracted features (9,071 non-temporal and 2,500 time-series features). To be more specific, we used categorical data, such as the disease codes and procedure codes, into the number of times each disease or condition was presented or the procedure performed. This gives us 9,071 non-temporal features that consist of demographics, microbiology, diagnosis codes, and medication data. In this dataset, since the temporal information for microbiology, medication, and pathology was not available, we treated them as non-temporal data and performed aggregates over the duration of the stay. The temporal data we used were from the various lab tests performed. Lab test data had a median sampling interval of 2.05 h, from which we extracted 2,500 features. After removing features with $>80\%$ missing data, we were left with 1,882 lab features, which we binned into 2 h binning intervals. In addition, this dataset had an issue where the tests or values were not recorded for very long-time intervals (\approx several days) in the middle. This could be due to the fact that the patients were no longer in the ICU. We treated these type of data as multiple time-series for each patient visit and did not use the missing period for binning. This gave us a total of 8,489 series. Since non-temporal data have $<1\%$ data missing, no features were removed.

The features extracted in the previous step were either quantitative real numbers or binary. The range of quantitative features had an order of magnitude variation (e.g., respiration rate varied from 10 to 30 breaths per min, blood pressure varied from 90 to 150 mmHg, and blood calcium varied between 8 and

11 mg/dl). To address this issue, we normalized all the features between the ranges 1 and 2. We also converted the binary values into 1 or 2.

Classifiers

Four classifiers are introduced in this section. Gradient boosting and random forest are ensemble models of decision trees. We will also briefly describe the linear regression classifier and support vector machine.

Random Forest

A random forest classifier is an ensemble of decision trees that take the advantage of a large number of uncorrelated trees. Each decision tree in the random forest makes a class prediction and the random forest model takes the votes of the majority. The core theory behind random forest is that a large crowd of decision trees would outperform individual trees. To ensure model generalizability, random forest enables feature randomness that when splitting a node, the model considers all possible features and selects the one leading to the largest separation between left node observations and right node ones. In this work, we used the online package Sklearn¹ to apply the random forest classifier. We used Gini impurity as the criterion to measure the quality of split. To mitigate the overfitting issue, the maximum tree depth was used; it was also the hyperparameter that we tuned in an experiment using grid search.

Gradient Boosting

Boosting is to convert weaker learners into stronger learners. Typically, a weaker learner is a decision tree that classifies the data with a poor performance. Each new tree in boosting is a fit to a modified copy of the original dataset. Boosting can be explained as a numerical optimization problem that the objective is to minimize the loss function defined in the model by using a gradient descent procedure when adding new weak learners.

Gradient boosting (Friedman, 2001) is a stage-wised additive model. It trains many models in a sequential and gradual way. When a new weak learner is added, existing weak learners remain unchanged in the model. Gradient boosting allows the optimization of arbitrary loss functions that are differentiable. In this work, we used the online package Sklearn to apply gradient boosting classifier on ICU data. We used deviance as the loss function, and Friedman mean squared error (Friedman, 1997) to measure the quality of tree split. To mitigate overfitting issue, the maximum tree depth is used; it was also the hyperparameter that we tuned in an experiment using grid search.

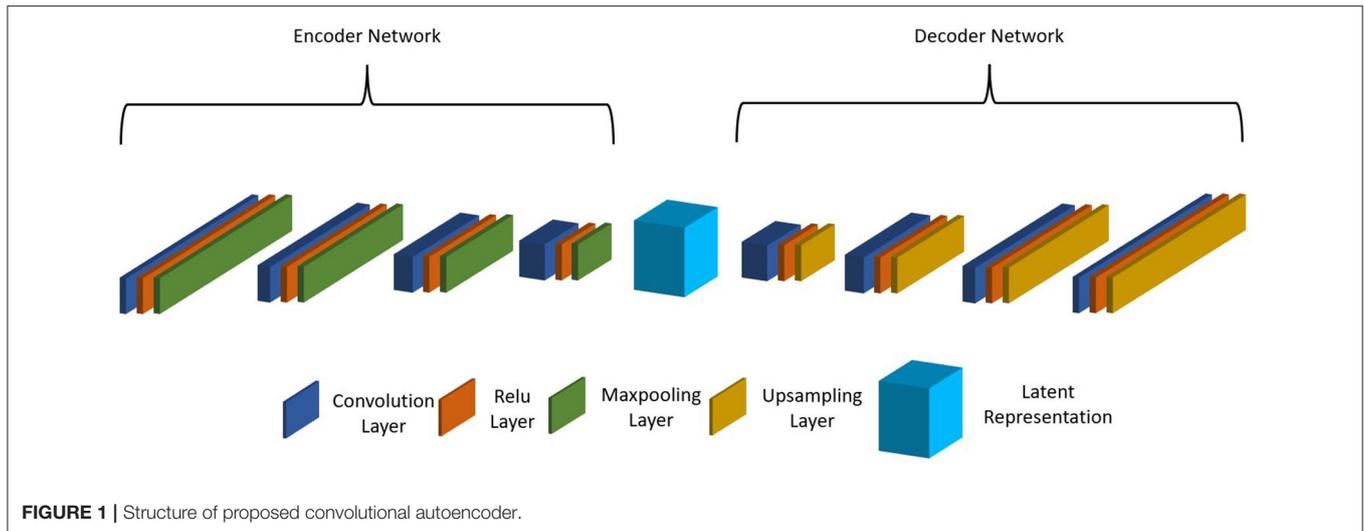
Linear Regression

A Linear regression classifier of Ordinary Least Squares fits a linear model that aims to minimize the residual sum of squares between the prediction (linear approximation) and the ground truth observations. The mathematical formula is:

$$\min_w \|Xw - y\|_2^2.$$

Where X is the input data matrix, w is the weight matrix, and y is the predicted vector. A major limitation of the linear regression

¹<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>



classifier is that it is sensitive to class imbalance. We have used linear regression as a baseline model in this work.

Support Vector Machine

Support vector machine is a kernelized method for classification and outlier detection. It is effective for high-dimensional feature data. In this project, we have used the Radial Basis Function (RBF) kernel for our SVM classifier.

CAE and Domain Adaptation

As the MIMIC dataset has many more patient records than the internal CHOA dataset, we would like to transfer the knowledge learned from the external dataset and apply to the internal dataset to improve model prediction performance. Specifically, we would like to apply domain adaptation to transform data from both domains into a shared feature space and learn domain-invariant feature representation. When inspecting the datasets, we found that the feature names in MIMIC and CHOA datasets cannot be matched since the two datasets are not following the same rules to name their features. This is a major challenge, as we cannot directly learn the domain-independent features across the two datasets. Thus, we proposed to use non-negative matrix factorization (NMF) and CAE for latent feature space transformation. NMF reduces feature dimension so that the CAE model can fit on data from both domains. CAEs are effective in learning latent feature vectors on pathological whole slide images (Zhu et al., 2019); yet for this project, we used CAE with 1D convolutional layers. As shown in **Figure 1**, the stacked CAE consists of an encoder that encodes the original input feature vector into compressed feature representation and a symmetric decoder that projects learned feature representation onto the original feature space. Each block of the encoder includes a 1D convolutional layer, a ReLU activation layer, and a max pooling layer. Similarly, each block of the decoder includes a 1D convolutional layer, a ReLU activation layer, and an upsampling layer. The reconstructed output vector is compared

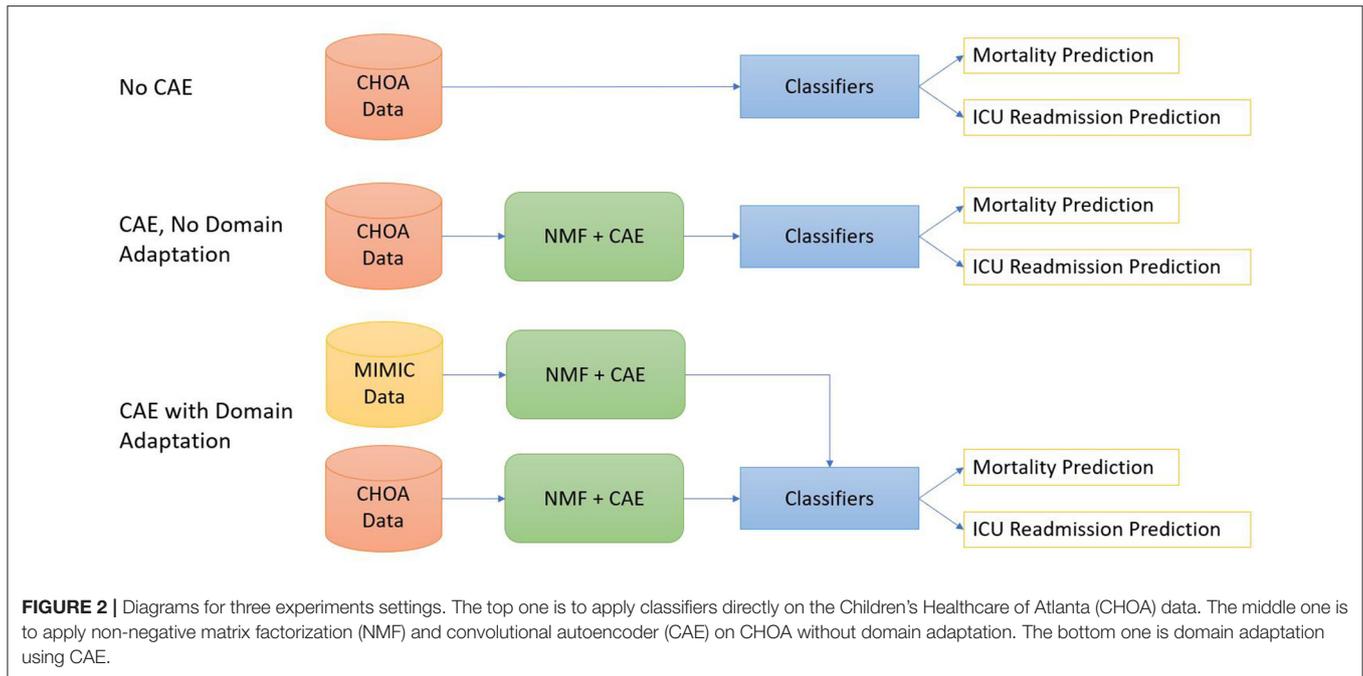
with the original input vector, and the loss is computed and backpropagated during optimization.

EXPERIMENTS

As shown in **Figure 2**, we design three different experimental settings for both mortality prediction and ICU readmission prediction.

1. *Shallow classifiers only.* We directly applied shallow classifiers, such as gradient boosting, random forest, linear regression, and SVM models, on temporal and non-temporal features of CHOA data.
2. *Convolutional autoencoder and shallow classifiers.* We first applied NMF on CHOA data for feature dimensionality reduction and then used a CAE model to learn latent feature representation. Afterward, we applied shallow classifiers on the concatenation of temporal and non-temporal features of the CHOA data.
3. *Convolutional autoencoder, domain adaptation, and shallow classifiers.* We first separately apply NMF on MIMIC and CHOA data for feature dimensionality reduction, then used two separate CAE models to learn latent feature representation from these two datasets. We then pretrain shallow classifiers on the learned latent feature vectors of MIMIC data and fine-tune the classifiers on CHOA features before evaluating the testing set of CHOA data.

Hyperparameter tuning was conducted using grid search and cross validation. The best classification model was automatically selected based on the highest average AUC-ROC score across 5-fold cross-validation. The maximum tree depth was the hyperparameter tuned for the random forest model and gradient boosting model. We also tried a different number of components in NMF. As for CAE, we empirically varied the number of blocks between 3 and 5. The number of filters in each convolutional layer increases by a factor of 2 from outside to the inside of the encoder, the kernel size of the convolutional



layer is 2, and the stride is 1. The CAE was implemented in Keras/TensorFlow.

We performed statistical analysis using two-way analysis of variance (ANOVA) to compare classification results. The two independent variables are experiment settings and classifiers. The null hypothesis is that the mean classification results in different groups are the same. Thus, the alternative hypothesis is that one group mean is different from other groups. In addition to variable-level statistical analysis, we also tried to identify which values of these two variables are significant. We performed a pairwise comparison using the Tukey’s honestly significant difference (HSD) test (Keselman and Rogan, 1978). We would reject the null hypothesis if the p is <0.05 .

RESULTS AND DISCUSSION

Temporal and Non-Temporal Features

We first evaluated the classification performance using different feature sets. As shown in **Table 2** and **Table 3**, the highest area under the curve (AUC)-receiver operating characteristic (ROC) score is achieved using a gradient boosting classifier on the concatenation of both temporal and non-temporal features. On the mortality prediction task, both feature sets have better performance than temporal only feature sets, which are better than non-temporal only feature sets using gradient boosting or random forest classifier. On ICU readmission prediction, the performance between using both feature sets and non-temporal features only is close when using gradient boosting and random forest models; both sets have better performance than temporal feature only. Linear regression and SVM classifiers have lower AUC-ROC scores when using both feature sets than the decision-tree-based models. We argue that non-temperature feature sets

TABLE 2 | Average and standard deviation (SD) of AUC-ROC score for mortality prediction using shallow classifiers with temporal features only, non-temporal features only, and both types of features.

Classifier	Both features	Temporal only	Non-temporal only
Gradient boosting	0.95718 (0.01546)	0.91020 (0.03810)	0.90132 (0.01824)
Random forest	0.94583 (0.01982)	0.92798 (0.03218)	0.87685 (0.02455)
Linear regression	0.71654 (0.08129)	0.80372 (0.05604)	0.74497 (0.06087)
SVM	0.62956 (0.02939)	0.62049 (0.03315)	0.78573 (0.04997)

The highest average score is highlighted in bold and underline.

TABLE 3 | Average and standard deviation (SD) of AUC-ROC score for ICU readmission prediction results using shallow classifiers with temporal features only, non-temporal features only, and both types of features.

Classifier	Both features	Temporal only	Non-temporal only
Gradient boosting	0.76651 (0.03185)	0.60214 (0.03058)	0.75414 (0.03747)
Random forest	0.74103 (0.03934)	0.60882 (0.01912)	0.74421 (0.04731)
Linear regression	0.55374 (0.01909)	0.60293 (0.02486)	0.70807 (0.03570)
SVM	0.54261 (0.03597)	0.46318 (0.01806)	0.66418 (0.04506)

The highest average score is highlighted in bold and underline.

and temporal feature sets contain similar information on adverse events prediction, while the concatenation of both feature sets would achieve the best performance.

Mortality Prediction

Table 4 shows the average and standard deviation (SD) of the mortality prediction results across five-fold on CHOA data under three different experimental settings. We also visualize the

results using a box plot in **Figure 3**. Specifically, the experiment setting of no domain adaptation (directly applying shallow classifiers on CHOA data) achieves better results than the other two experiment settings (pairwise comparison, $p < 0.05$). Domain adaptation using CAE achieves better results than CAE without domain adaptation ($p < 0.05$). This indicates that even though the existing latent feature extraction method using CAE is not effective, domain adaptation between two datasets can significantly improve the performance. Meanwhile, gradient boosting and random forest algorithms are not significantly different in performance ($p > 0.05$), but either of them is better than SVM and linear regression classifier (pairwise comparison, $p < 0.05$).

We also identify the top features for MIMIC and CHOA mortality prediction that include both temporal information and non-temporal information (in **Table 5**). For the MIMIC

dataset, maximum and minimum values of Simplified Acute Physiology Score (SAPS) (Agha et al., 2002) and Sequential Organ Failure Assessment (SOFA) (Arts et al., 2005) score are top-ranking features. SAPS II and SOFA are clinical scores that were designed to assess the severity of illness for ICU patients and to predict their risk of mortality, using lab tests and clinical data. Features, such as hospital length of stay and ICU length of stay, might be controversial for mortality prediction, as the longer length of stay may indicate more severe illness and they fail to predict “ahead of mortality event.” For CHOA dataset, the top-ranking features are lab measurements, clinical events and drug information and ventilator days. These top-ranking features are

TABLE 4 | Mortality prediction results on and Children’s Healthcare of Atlanta (CHOA) data (both temporal and non-temporal features) under three different experimental settings.

Classifier	No domain adaptation	CAE on CHOA only	CAE and domain adaptation
Gradient boosting	0.95718 (0.01546)	0.74864 (0.04162)	0.86375 (0.02267)
Random forest	0.94583 (0.01982)	0.74855 (0.03668)	0.86375 (0.02267)
Linear regression	0.71654 (0.08129)	0.69893 (0.05303)	0.76208 (0.03429)
SVM	0.62956 (0.02939)	0.50896 (0.03349)	0.44725 (0.12380)

The highest average score is highlighted in bold and underline.

TABLE 5 | Top 10 features in mortality prediction using gradient boosting on Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) and Children’s Healthcare of Atlanta (CHOA) data.

Feature ranking	MIMIC	CHOA
1	sofa_max	component_name_art base deficit
2	sofa_min	dx_code_v49.86
3	hospital_los	component_name_nrbc
4	icustay_los	component_name_patient fio2
5	sapsi_max	component_name_plasma free hgb
6	sapsi_min	ventilator_days
7	subject_icustay_total_num	dx_rank
8	cost_weight	component_name_ast (sgot)
9	sofa_first	dx_present_on_admit_yn
10	peptic_ulcer	value_in_range_yn



FIGURE 3 | Boxplot of mortality prediction results using four different classifiers with three experiment settings.

domain-specific; they are very different from the features in the MIMIC dataset. Further validation and interpretation of these top-ranking features are needed to identify potential biomarkers for clinical practice.

ICU Readmission Prediction

Table 6 shows the average and SD of the ICU readmission prediction results across 5-fold on CHOA data under three different experimental settings. We also visualize the results using a box plot in Figure 4. Similar to the results in the mortality prediction task, the experiment setting of no domain adaptation (directly applying shallow classifiers on CHOA data) achieves better results than the other two experiment settings (pairwise comparison, $p < 0.05$). However, domain adaptation using CAE fails to achieve better results than CAE without domain adaptation ($p > 0.05$). If we only focus on the best results models (gradient boosting and random forest), domain adaptation

has better performance than CAE without domain adaptation ($p < 0.05$). This indicates that even though the existing latent feature extraction method using CAE is not effective, domain adaptation between two datasets can significantly improve the performance. Similar to the results in the mortality prediction task, gradient boosting and random forest algorithms are not significantly different in performance ($p > 0.05$), but either of them is better than SVM and linear regression classifier (pairwise comparison, $p < 0.05$).

Different from mortality prediction task, the top-ranking features for ICU readmission prediction on MIMIC data do not have SAPS II or SOFA scores. As shown in Table 7, these

TABLE 6 | ICU readmission prediction results on Children’s Healthcare of Atlanta (CHOA) data (both temporal and non-temporal features) under three different experimental settings.

Classifier	No domain adaptation	CAE on CHOA only	CAE and domain adaptation
Gradient boosting	0.76651 (0.03185)	0.57404 (0.01665)	0.62448 (0.03845)
Random forest	0.74103 (0.03934)	0.59993 (0.01988)	0.63117 (0.03904)
Linear regression	0.55374 (0.01909)	0.57664 (0.00749)	0.58234 (0.01914)
SVM	0.54261 (0.03597)	0.51617 (0.01912)	0.44870 (0.03285)

The highest average score is highlighted in bold and underline.

TABLE 7 | Top 10 features in ICU readmission prediction using gradient boosting on Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) and Children’s Healthcare of Atlanta (CHOA) data.

Feature ranking	MIMIC	CHOA
1	Gender	dx_code_v44.1
2	hospital_los	dx_rank
3	tidal volume (obser)	apgar_score_5_minutes
4	temperature f	gestation_age_weeks
5	icustay_los	dx_type_coded final
6	spo2	nicu_yn
7	carbon dioxide	birth_weight
8	renal_failure	discharge_destination
9	congestive_heart_failure	dx_present_on_admit_exempt_yn
10	fingerstick glucose	birth_length

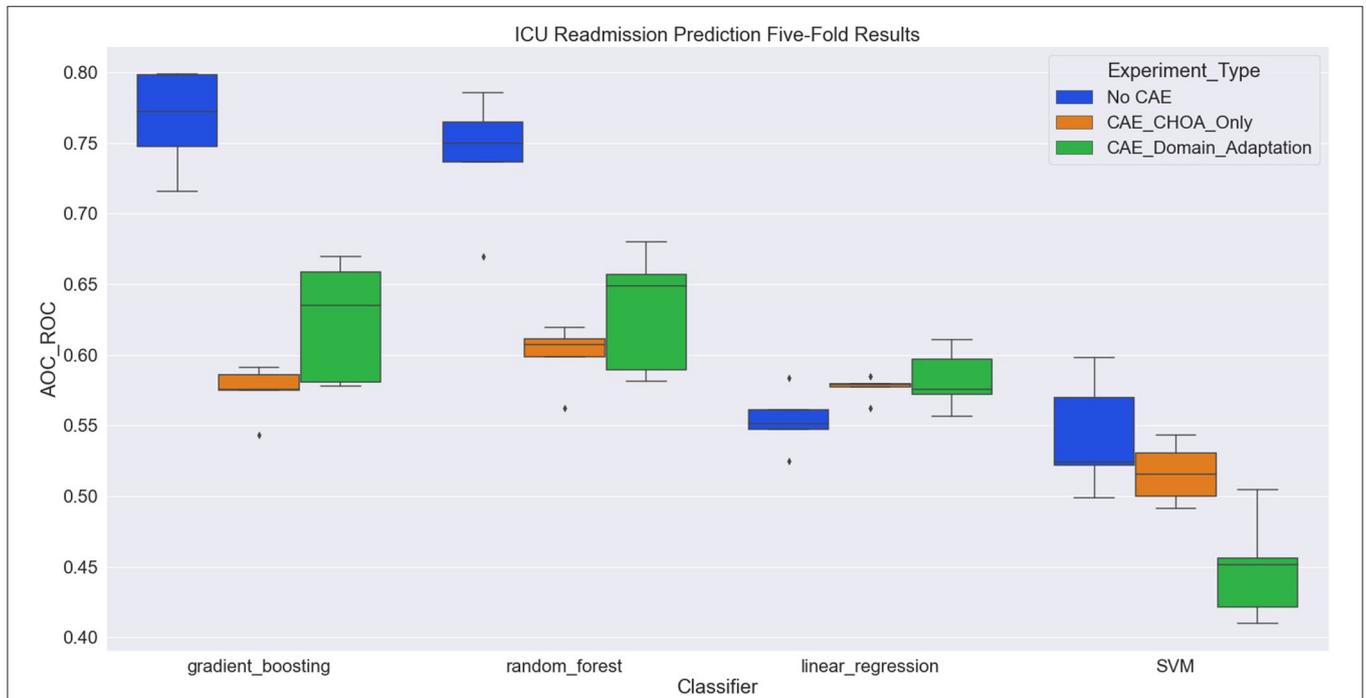


FIGURE 4 | Boxplot of intensive care unit (ICU) readmission prediction results using four different classifiers with three experiment settings.

top-ranking features include lab measurements, chart events, and demographic information. We argue that gradient boosting is effective in identifying top-ranking features for both mortality prediction and ICU readmission prediction, as statistical features of SAPS II and SOFA scores are the majority of top-ranking features in the mortality prediction task, but not in the ICU readmission prediction task. For CHOA data, the top-ranking features are different from those in the mortality prediction task, focusing on drug-related features, but still include lab measurements and clinical events. Some of these top-ranking features are pediatric-specific; they are very different from the features in the MIMIC dataset. Similar to the mortality prediction task, further validation and interpretation on these top-ranking features are needed to identify potential biomarkers for clinical practice.

CONCLUSION

In this work, we extracted temporal and non-temporal features from one public ICU dataset (MIMIC) and a local ICU dataset (CHOA) to build predictive models on mortality risk and ICU readmission. We designed three different experimental settings, implemented CAE to learn latent feature representation, and applied multiple classifiers, such as gradient boosting and random forest for classification. We demonstrated the effectiveness of gradient boosting in both mortality prediction and ICU readmission prediction tasks. In addition, we showed that domain adaptation using CAE across two datasets can significantly improve results against using CAE and classifiers without domain adaptation. We aim to learn domain-invariant latent feature representation and improve prediction performance on the clinical adverse event when the local data set has a very limited sample size.

There are some limitations of this work. First, the temporal features in the CHOA dataset are binned into intervals of 6 h each, which could result in loss of granularity of data and introduction of new missing data points. Second, domain adaptation is designed on the latent feature representation level, not on the deep neural network (CAE) level. This is largely due to the different feature names and feature quantities, as the same deep neural network model with the same hyperparameters cannot be used on both datasets. Third, there is a fundamental bias between the two datasets. For the MIMIC dataset, we extracted ICU data for patients over 16 years old; for the CHOA

dataset, we expect the majority of the patients is children. Consequently, the knowledge learned from the adult patient group may not help the model predictions on the children patient group.

As for future work, we would like to acquire time-stamped temporal information for the CHOA ICU dataset. We will implement deep learning models, such as RNNs, to capture the temporal information of lab tests and microbiology events (instead of aggregating them into non-temporal data). We believe that a combination of static and temporal models could give additional insight into the disease process and improve prediction performance. In addition, we will apply fairness-learning techniques to mitigate biased prediction on age- and birth-related factors. In this way, we can improve the fairness of the domain adaptation model and improve the prediction results. Lastly, we would like to overcome the different feature names of the two ICU datasets so that we can apply the same deep neural network to them to transfer the knowledge between the deep neural network.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YZ designed the experiments, implemented the gradient boosting classifier and convolutional autoencoder (CAE), and drafted the article. JV initiated the project, cleaned and preprocessed the data, and helped draft the article. ZZ helped preprocess the data and literature review. NC and KM annotated the CHOA data and provided clinical guidance on top features. MW guided and oversaw the project and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported in part by the Children's Healthcare of Atlanta (CHOA), the NIH National Center for Advancing Translational Sciences UL1TR000454, the National Science Foundation Award NSF1651360, Microsoft Research, Georgia Institute of Technology PACE, and Hewlett Packard.

REFERENCES

- Agha, A., Bein, T., Fröhlich, D., Höfler, S., Krenz, D., and Jauch, K. W. (2002). "Simplified Acute Physiology Score" (SAPS II) in the assessment of severity of illness in surgical intensive care patients. *Chirurg* 73, 439–442. doi: 10.1007/s00104-001-0374-4
- Arts, D. G. T., de Keizer, N. F., Vroom, M. B., and de Jonge, E. (2005). Reliability and accuracy of sequential organ failure assessment (SOFA) scoring. *Crit. Care Med.* 33, 1988–1993. doi: 10.1097/01.CCM.0000178178.02574.AB
- Batal, I., Valizadegan, H., Cooper, G. F., and Hauskrecht, M. (2011). "A pattern mining approach for classifying multivariate temporal data," in *2011 IEEE International Conference on Bioinformatics and Biomedicine*, 358–365.
- Bellazzi, R., Ferrazzi, F., and Sacchi, L. (2011). Predictive data mining in clinical medicine: a focus on selected methods and applications. *WIREs Data Mining Knowledge Discov.* 1, 416–430. doi: 10.1002/widm.23
- Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., et al. (2017). Heart disease and stroke statistics—2017 update: a report from the American Heart Association. *Circulation* 135, e146–e603. doi: 10.1161/CIR.0000000000000485
- Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., and Erhan, D. (2016). Domain Separation Networks. *arXiv:1608.06019 [cs]*. Available online at: <http://arxiv.org/abs/1608.06019> (accessed January 20, 2022).
- Cai, X., Perez-Concha, O., Coiera, E., Martin-Sanchez, F., Day, R., Roffe, D., et al. (2015). Real-time prediction of mortality, readmission, and length of stay

- using electronic health record data. *J. Am. Med. Inform. Assoc.* 23, 553–561. doi: 10.1093/jamia/ocv110
- Casanova, I. J., Campos, M., Juarez, J. M., Fernandez-Fernandez-Arroyo, A., and Lorente, J. A. (2015). “Using multivariate sequential patterns to improve survival prediction in intensive care burn unit,” in *Artificial Intelligence in Medicine Lecture Notes in Computer Science*, editors J. H. Holmes, R. Bellazzi, L. Sacchi, and N. Peek (Cham: Springer International Publishing), 277–286.
- Celi, L. A., Galvin, S., Davidzon, G., Lee, J., Scott, D., and Mark, R. (2012). A database-driven decision support system: customized mortality prediction. *J. Pers. Med.* 2, 138–148. doi: 10.3390/jpm2040138
- Celi, L. A. G., Tang, R. J., Villarroel, M. C., Davidzon, G. A., Lester, W. T., and Chueh, H. C. (2011). A clinical database-driven approach to decision support: predicting mortality among patients with acute kidney injury. *J. Healthc. Eng.* 2, 97–110. doi: 10.1260/2040-2295.2.1.97
- Chen, S., Qin, J., Ji, X., Lei, B., Wang, T., Ni, D., et al. (2017). Automatic scoring of multiple semantic attributes with multi-task feature leverage: a study on pulmonary nodules in CT images. *IEEE Trans. Med. Imaging* 36, 802–814. doi: 10.1109/TMI.2016.2629462
- Cheplygina, V., de Bruijne, M., and Pluim, J. P. W. (2019). Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Anal.* 54, 280–296. doi: 10.1016/j.media.2019.03.009
- Choi, E., Schuetz, A., Stewart, W. F., and Sun, J. (2017). Medical Concept Representation Learning From Electronic Health Records and its Application on Heart Failure Prediction. *arXiv:1602.03686 [cs]*. Available online at: <http://arxiv.org/abs/1602.03686> (accessed December 8, 2020).
- Choudhary, A., Tong, L., Zhu, Y., and Wang, M. D. (2020). Advancing medical imaging informatics by deep learning-based domain adaptation. *Yearb. Med. Inform.* 29, 129–138. doi: 10.1055/s-0040-1702009
- Damodaran, B. B., Kellenberger, B., Flamary, R., Tuia, D., and Courty, N. (2018). “DeepJDOT: deep joint distribution optimal transport for unsupervised domain adaptation,” in *Computer Vision – ECCV 2018 Lecture Notes in Computer Science*, editors V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer International Publishing), 467–483.
- de Rooij, S. E., Govers, A., Korevaar, J. C., Abu-Hanna, A., Levi, M., and de Jonge, E. (2006). Short-term and long-term mortality in very elderly patients admitted to an intensive care unit. *Intensive Care Med.* 32, 1039–1044. doi: 10.1007/s00134-006-0171-0
- Desautels, T., Das, R., Calvert, J., Trivedi, M., Summers, C., Wales, D. J., et al. (2017). Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach. *BMJ Open* 7, e017199. doi: 10.1136/bmjopen-2017-017199
- Doshi-Velez, F., Ge, Y., and Kohane, I. (2014). Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. *Pediatrics* 133, e54–e63. doi: 10.1542/peds.2013-0819
- Elmahdy, M. S., Abdeldayem, S. S., and Yassine, I. A. (2017). “Low quality dermal image classification using transfer learning,” in *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 373–376.
- Fagon, J.-Y., Chastre, J., Hance, A. J., Montravers, P., Novara, A., and Gibert, C. (1993). Nosocomial pneumonia in ventilated patients: a cohort study evaluating attributable mortality and hospital stay. *Am. J. Med.* 94, 281–288. doi: 10.1016/0002-9343(93)90060-3
- Fagon, J.-Y., Novara, A., Stephan, F., Girou, E., and Safar, M. (1994). Mortality attributable to nosocomial infections in the ICU. *Infect. Control Hosp. Epidemiol.* 15, 428–434. doi: 10.2307/30148490
- Fialho, A. S., Cisondi, F., Vieira, S. M., Reti, S. R., Sousa, J. M. C., and Finkelstein, S. N. (2012). Data mining using clinical physiology at discharge to predict ICU readmissions. *Expert Syst. Appl.* 39, 13158–13165. doi: 10.1016/j.eswa.2012.05.086
- Friedman, J. H. (1997). On bias, variance, 0/1-loss, and the curse-of-dimensionality. *Data Min. Knowl. Discov.* 1, 55–77. doi: 10.1023/A:1009778005914
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi: 10.1214/aos/1013203451
- Fuchs, L., Chronaki, C. E., Park, S., Novack, V., Baumfeld, Y., Scott, D., et al. (2012). ICU admission characteristics and mortality rates among elderly and very elderly patients. *Intensive Care Med.* 38, 1654–1661. doi: 10.1007/s00134-012-2629-6
- Ghifary, M., Kleijn, W. B., Zhang, M., Balduzzi, D., and Li, W. (2016). Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation. *arXiv:1607.03516 [cs, stat]*. Available online at: <http://arxiv.org/abs/1607.03516> (accessed January 20, 2022).
- Goldstein, B. A., Navar, A. M., Pencina, M. J., and Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J. Am. Med. Inform. Assoc.* 24, 198–208. doi: 10.1093/jamia/ocw042
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative adversarial nets,” in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.). Available online at: <https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html> (accessed January 20, 2022).
- Gupta, P., Malhotra, P., Narwariya, J., Vig, L., and Shroff, G. (2020). Transfer learning for clinical time series analysis using deep neural networks. *J. Healthc. Inform. Res.* 4, 112–137. doi: 10.1007/s41666-019-00062-3
- Han, D., Liu, Q., and Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Syst. Appl.* 95, 43–56. doi: 10.1016/j.eswa.2017.11.028
- Hug, C. W., Clifford, G. D., and Reisner, A. T. (2011). Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension. *Crit. Care Med.* 39, 1006–1014. doi: 10.1097/CCM.0b013e31820eab8e
- Hunziker, S., Celi, L. A., Lee, J., and Howell, M. D. (2012). Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients. *Critical Care* 16, R89. doi: 10.1186/cc11351
- Hussein, S., Cao, K., Song, Q., and Bagci, U. (2017). “Risk stratification of lung nodules using 3D CNN-based multi-task learning,” in *Information Processing in Medical Imaging Lecture Notes in Computer Science*, editors M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap, et al. (Cham: Springer International Publishing), 249–260.
- Hutchinson, S. G., Mesters, I., van Breukelen, G., Muris, J. W., Feron, F. J., Hammond, S. K., et al. (2013). A motivational interviewing intervention to PREvent PAssive Smoke Exposure (PREPASE) in children with a high risk of asthma: design of a randomised controlled trial. *BMC Public Health* 13, 177. doi: 10.1186/1471-2458-13-177
- Huynh, B. Q., Li, H., and Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *JMI* 3, 034501. doi: 10.1117/1.JMI.3.3.034501
- Hwang, S., and Kim, H.-E. (2016). Self-Transfer Learning for Fully Weakly Supervised Object Localization. *arXiv:1602.01625 [cs]*. Available online at: <http://arxiv.org/abs/1602.01625> (accessed January 20, 2022).
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2018). Image-to-Image Translation with Conditional Adversarial Networks. *arXiv:1611.07004 [cs]*. Available online at: <http://arxiv.org/abs/1611.07004> (accessed January 20, 2022).
- Kandemir, M. (2015). “Asymmetric transfer learning with deep gaussian processes,” in *Proceedings of the 32nd International Conference on Machine Learning (PMLR)*, 730–738. Available online at: <https://proceedings.mlr.press/v37/kandemir15.html> (accessed January 20, 2022).
- Kang, G., Jiang, L., Yang, Y., and Hauptmann, A. G. (2019). “Contrastive adaptation network for unsupervised domain adaptation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA: IEEE), 4888–4897.
- Keselman, H. J., and Rogan, J. C. (1978). A comparison of the modified-Tukey and Scheffe methods of multiple comparisons for pairwise contrasts. *J. Am. Stat. Assoc.* 73, 47–52. doi: 10.1080/01621459.1978.10479996
- Kissoon, N., Daniels, R., van der Poll, T., Finfer, S., and Reinhart, K. (2016). Sepsis—the final common pathway to death from multiple organ failure in infection. *Crit. Care Med.* 44, e446. doi: 10.1097/CCM.0000000000001582
- Kwon, S., Florence, M., Grigas, P., Horton, M., Horvath, K., Johnson, M., et al. (2012). Creating a learning healthcare system in surgery: Washington State’s Surgical Care and Outcomes Assessment Program (SCOAP) at 5 years. *Surgery* 151, 146–152. doi: 10.1016/j.surg.2011.08.015
- Lafarge, M. W., Pluim, J. P., Eppenhof, K. A., Moeskops, P. and Veta, M. (2017). “Domain-adversarial neural networks to address the appearance variability of histopathology images,” in *Deep Learning in Medical Image Analysis and*

- Multimodal Learning for Clinical Decision Support* (Cham: Springer), 83–91. doi: 10.1007/978-3-319-67558-9_10
- Lee, J., Kothari, R., Ladapo, J. A., Scott, D. J., and Celi, L. A. (2012). Interrogating a clinical database to study treatment of hypotension in the critically ill. *BMJ Open* 2, e000916. doi: 10.1136/bmjopen-2012-000916
- Lee, J., and Mark, R. G. (2010). An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care. *Biomed. Eng. Online* 9, 62. doi: 10.1186/1475-925X-9-62
- Lee, K.-H., He, X., Zhang, L., and Yang, L. (2018). *CleanNet: Transfer Learning for Scalable Image Classifier Training With Label Noise*, 5447–5456. Available online at: https://openaccess.thecvf.com/content_cvpr_2018/html/Lee_CleanNet_Transfer_Learning_CVPR_2018_paper.html (accessed December 8, 2020).
- Li, Z., Wang, C., Han, M., Xue, Y., Wei, W., Li, L.-J., et al. (2018). *Thoracic Disease Identification and Localization With Limited Supervision*, 8290–8299. Available online at: https://openaccess.thecvf.com/content_cvpr_2018/html/Li_Thoracic_Disease_Identification_CVPR_2018_paper.html (accessed January 20, 2022).
- Liu, Z., Wu, L., and Hauskrecht, M. (2013). “Modeling clinical time series using gaussian process sequences,” in *Proceedings of the 2013 SIAM International Conference on Data Mining (Society for Industrial and Applied Mathematics)*, 623–631.
- Mandelbaum, T., Scott, D. J., Lee, J., Mark, R. G., Malhotra, A., Waikar, S. S., et al. (2011). Outcome of critically ill patients with acute kidney injury using the AKIN criteria. *Crit. Care Med.* 39, 2659–2664. doi: 10.1097/CCM.0b013e3182281f1b
- Marshall, J. C., Cook, D. J., Christou, N. V., Bernard, G. R., Sprung, C. L., and Sibbald, W. J. (1995). Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Crit. Care Med.* 23, 1638–1652. doi: 10.1097/00003246-199510000-00007
- McCoy, T. H., Castro, V. M., Cagan, A., Roberson, A. M., Kohane, I. S., and Perlis, R. H. (2015). Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: an electronic health record study. *PLoS ONE* 10, e0136341. doi: 10.1371/journal.pone.0136341
- Murthy, V., Hou, L., Samaras, D., Kurc, T. M., and Saltz, J. H. (2017). “Center-focusing multi-task CNN with injected features for classification of glioma nuclear images,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 834–841.
- Nemati, S., Malhotra, A., and Clifford, G. D. (2011). T-wave alternans patterns during sleep in healthy, cardiac disease, and sleep apnea patients. *J. Electrocardiol.* 44, 126–130. doi: 10.1016/j.jelectrocard.2010.10.036
- Ribas, V. J., López, J. C., Ruiz-Rodríguez, J. C., Ruiz-Sanmartín, A., Rello, J., and Vellido, A. (2011). “On the use of decision trees for ICU outcome prediction in sepsis patients treated with statins,” in *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 37–43.
- Rozantsev, A., Salzmann, M., and Fua, P. (2019). Beyond sharing weights for deep domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 801–814. doi: 10.1109/TPAMI.2018.2814042
- Saeed, M., Villarreal, M., Reisner, A. T., Clifford, G., Lehman, L.-W., Moody, G., et al. (2011). Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database. *Crit. Care Med.* 39, 952–960. doi: 10.1097/CCM.0b013e31820a92c6
- Singh, A., Nadkarni, G., Gottesman, O., Ellis, S. B., Bottinger, E. P., and Guttag, J. V. (2015). Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration. *J. Biomed. Inform.* 53, 220–228. doi: 10.1016/j.jbi.2014.11.005
- Stiglic, G., Davey, A., and Obradovic, Z. (2013). “Temporal evaluation of risk factors for acute myocardial infarction readmissions,” in *2013 IEEE International Conference on Healthcare Informatics*, 557–562.
- Syed, H., and Das, A. K. (2015). “Identifying chemotherapy regimens in electronic health record data using interval-encoded sequence alignment,” in *Artificial Intelligence in Medicine Lecture Notes in Computer Science*, editors J. H. Holmes, R. Bellazzi, L. Sacchi, and N. Peek (Cham: Springer International Publishing), 143–147.
- Tao, C., Wongsuphasawat, K., Clark, K., Plaisant, C., Shneiderman, B., and Chute, C. G. (2012). “Towards event sequence representation, reasoning and visualization for EHR data,” in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium IHI '12*. (New York, NY: Association for Computing Machinery), 801–806.
- Toddenroth, D., Ganslandt, T., Castellanos, I., Prokosch, H.-U., and Bürkle, T. (2014). Employing heat maps to mine associations in structured routine care data. *Artif. Intell. Med.* 60, 79–88. doi: 10.1016/j.artmed.2013.12.003
- Venugopalan, J., Chanani, N., Maher, K. and Wang, M. D. (2017). “Combination of static and temporal data analysis to predict mortality and readmission in the intensive care,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE)*, 2570–2573.
- Venugopalan, J., Chanani, N., Maher, K., and Wang, M. D. (2019). Novel data imputation for multiple types of missing data in intensive care units. *IEEE J. Biomed. Health Inform.* 23, 1243–1250. doi: 10.1109/JBHI.2018.2883606
- Vranas, K. C., Jopling, J. K., Scott, J. Y., Badawi, O., Harhay, M. O., Slatore, C. G., et al. (2018). The association of ICU with outcomes of patients at low risk of dying. *Crit. Care Med.* 46, 347–353. doi: 10.1097/CCM.0000000000002798
- Wang, T. D., Plaisant, C., Quinn, A. J., Stanchak, R., Murphy, S., and Shneiderman, B. (2008). “Aligning temporal data by sentinel events: discovering patterns in electronic health records,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '08*. (New York, NY: Association for Computing Machinery), 457–466.
- Warner, J. L., Zollanvari, A., Ding, Q., Zhang, P., Snyder, G. M., and Alterovitz, G. (2013). Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications. *J. Am. Med. Assoc.* 309, e281–e287. doi: 10.1136/amaiajn-2013-001861
- Wheeler, D. S., Giaccone, M. J., Hutchinson, N., Haygood, M., Bondurant, P., Demmel, K., et al. (2011). A hospital-wide quality-improvement collaborative to reduce catheter-associated bloodstream infections. *Pediatrics* 128, e995–e1007. doi: 10.1542/peds.2010-2601
- Wong, L. S. S., and Young, J. D. (1998). A comparison of ICU mortality prediction using the Apache II scoring system and artificial neural network. *Anaesthesia*. 54, 1048–54. doi: 10.1049/ic:19980797
- Wu, A. W., Pronovost, P., and Morlock, L. (2002). ICU incident reporting systems. *J. Crit. Care* 17, 86–94. doi: 10.1053/jcrrc.2002.35100
- Yang, H., and Yang, C. C. (2015). Using health-consumer-contributed data to detect adverse drug reactions by association mining with temporal analysis. *ACM Trans. Intell. Syst. Technol.* 6, 55:1–55:27. doi: 10.1145/2700482
- Yu, C.-N., Greiner, R., Lin, H.-C., and Baracos, V. (2011). Learning patient-specific cancer survival distributions as a sequence of dependent regressors. *Adv. Neural Inf. Process. Syst.* 24, 1845–1853.
- Zhu, Y., Tong, L., Deshpande, S. R., and Wang, M. D. (2019). “Improved prediction on heart transplant rejection using convolutional autoencoder and multiple instance learning on whole-slide imaging,” in *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 1–4.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhu, Venugopalan, Zhang, Chanani, Maher and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.