Predicting 30-Day Hospital Readmissions for Patients with Diabetes

Y. Zhao[1], W. Wu[1], Y. Jin[1], S. Gu[1], H. Wu[1], J. Wang[1], X. Jiang[2] and H. Xiao[2]

[1] *Computer and Information Science Department, Fordham University, Bronx, NY, USA*

[2] *Department of Pharmaceutical Outcomes and Policy, University of Florida, Gainesville, Florida*

Keywords: Biomedical Informatics, Hospital Readmission, Machine Learning, Ensemble Learning, Predictive Analytics, Risk Factor Analysis.

Abstract: Reducing avoidable hospital readmissions continues to be a high-priority task across the healthcare systems. In our study, we explore machine learning algorithms to predict 30-day hospital readmissions for patients with diabetes. Our work is based on a real-world dataset of 124,678 admission records extracted from the State Inpatient Database (SID) of Florida. We examine the performance of four popular machine learning algorithms and two ensemble learning approaches including the late novel *XGBoost* algorithm. In addition, we analyze the top risk factors associated with 30-day readmissions of diabetic patients. Our experimental results demonstrate that machine learning techniques can be effective in predicting early hospital readmissions based on patients' clinical and demographic characteristics. Furthermore, *XGBoost* outranks other methods with its superior performance, interpretability and scalability. Risk factors consistently identified by our models could serve as a Focus of Attention (FOA) tool for healthcare institutions to establish preventive measures to reduce the readmission rate.

1 INTRODUCTION

Accurately predicting a patient's hospital readmission risk is critical in improving both the quality of life for patients and the financial wellbeing of healthcare institutions. Unnecessary hospitalizations not only expose the patients to potential harm but also incur extra medical expenditures. Early readmission, also known as 30-day readmission, refers to a patient's unplanned hospitalization within 30 days of discharge. In the United States, 30-day all-cause hospital readmissions are associated with about \$41.3 billion in-hospital expenditures each year, while 27% of them are deemed avoidable (Van Walraven et al., 2011). Consequently, early readmission rate is broadly accepted as an indicator of the quality of care for healthcare institutions and considerable effort has been devoted to advancing the capability of assessing the risk of early hospital readmissions.

We focus our research on predicting 30-day readmission risk for patients with diabetes. In particular, the prevalence of diabetes has risen sharply across all age groups, all racial/ethnic groups and both genders over the past two decades (CDC, 2012). According to the National Diabetes Statistics Report (2017) from the Centers for Disease Control and Prevention (CDC), diabetic patients represent about 9.4% of the US population. Furthermore, the percentage of diabetic patients increases with age, reaching a high of 25.2% among those aged 65 years or older (CDC, 2017). Consequently, the number of hospital discharges with diabetes as the primary diagnosis has been increasing at a constant rate of approximately 2% (CDC, 2009) each year. In terms of 30-day readmission risk, diabetic patients carry a 15% chance compared to an average of 8.5% of the entire cohort of discharged patients (Friedman et al., 2008).

In our study, we applied machine learning algorithms to analyze 124,678 patients who had been discharged with a primary diagnosis code of diabetes. Specifically, we aimed to classify the patients into two different risk groups ("Yes" or "No") with respect to their 30-day readmission status based on information in their inpatient and discharge records. For our classification task, we first employed four established machine learning algorithms as our baseline learners. We then explored the power of *heterogeneous* ensemble learning (Dietterich, 2000) by building a meta-learner L on top of the four base learners. Lastly, we examined the *homogeneous* ensemble learning algorithm, *XGBoost* (Chen and Guestrin, 2016), which has gained much attention since its inception in 2016. Our experimental results demonstrate that *XGBoost* is the best model in terms of both overall performance and model stability (see Section 5 for details). More importantly, we conclude that machine learning techniques can be effective in predicting early hospital readmissions based on patients' clinical and demographic characteristics.

An additional motivation to our research is to study various risk factors associated with a patient's early readmission. To this end, we ranked the top predictors in our models and identified the most effective factors. Specifically, we identified leading predictive features (see Section 6 for details) such as a patient's age (*AGE*), total number of previous hospital visits (*TVISITS*), total number of diagnosis in current visit (*NDX*), length of current hospitalization (*LOS*), etc. These risk factors could serve as Focus of Attention (FOA) measures to help healthcare institutions improve their quality of care.

The rest of the paper is organized as follows: in Section 2, we present a brief survey of related work. In Section 3, we describe the acquisition and preprocessing of the dataset we used for our study. We introduce our models and present our findings in Sections 4 and 5 respectively. Finally, we conclude and discuss future work in Section 7.

2 RELATED WORK

Predicting the risk of early hospital readmissions is an active area of research. However, few studies have concentrated on patients with diabetes (Rubin, 2015). The lack of attention paid to readmission of diabetic patients may be related to numerous severe side effects associated with the disease which belie the primary diagnosis of a hospitalization. Futoma etc. (Futoma et al., 2015) provided a comprehensive comparison of four machine learning models (*SVM*, *Random Forest*, *Logistic Regression*, and *Neural Network*) for predicting early hospital readmissions in 280 Diagnosis Related Groups (DRGs). We employed same four models as our baseline models (see Section 4.1) and further explored the power of ensemble learning algorithms. Frizzell, etc. (Frizzell et al., 2017) applied a machine learning approach to predict 30-day all-cause readmissions after a heart failure hospitalization. Golas etc. (Golas et al., 2018) and Mortazavi etc. (Mortazavi et al., 2016) studied machine learning techniques for heart failure related readmissions. Although these studies were not directly conducted on a cohort of diabetic patients, their approaches and experimental results are highly relevant and insightful when compared with ours. As we present in Section 5, using the popular AUC model evaluation metric, our best algorithms achieved an improved AUC score of 0.80 compared to the scores ranging from 0.54 to 0.72 (Frizzell et al., 2017) in the related literature.

Another related study conducted on diabetic patients is Duggal etc. (Duggal et al., 2016) in which the authors compared different classification models to predict early readmission risk based on patients' two-year clinical and administrative data. Compared to their work, our approach is fundamentally different in both the methodology and the underlying data. In particular, data used in (Duggal et al., 2016) are patients' two-year longitudinal observations while ours are clinical and demographic characteristics of individual hospitalization visits.

Our main contribution is an in-depth study of machine learning methods applied to predicting early readmission risk for diabetic patients. In addition to examining individual machine learning algorithms, which is a standard approach in the existing literature, we investigated two additional methods under the ensemble learning paradigm. Another contribution of our paper is the study of top risk factors associated with early readmission of diabetic patients. We discuss our findings on this subject in Section 6.

3 DATA PREPROCESSING

Our dataset is part of the Healthcare Cost and Utilization Project (HCUP). HCUP is a family of healthcare databases developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ). To facilitate various research endeavors, HCUP maintains an array of databases including the National Inpatient Sample (NIS), the Kids' Inpatient Database (KID), the Nationwide Readmissions Database (NRD), the State Inpatient Database (SID), etc. Because diabetes and its related early readmissions tend to have a higher occurrence among the senior population (CDC, 2017), we selected the State Inpatient Database (SID) of Florida from 2012 to 2014 to conduct our study. The SIDs are state-specific databases which contain all inpatient care records with variables capturing lengths of stay, diagnosis and procedure codes, admission sources, disposition after hospitalization and information on demographic characteristics. In addition, SID provides unique encrypted patient identifiers to support readmission study. These identifiers help to link a subsequent admission to its prior hospitalization for each patient and, thus, permit a determination of elapsed time between two visits (more details in Section 3.1).

3.1 Constructing Training Data

The SID database contains over 2.5 million all-cause hospitalization and discharge records per year. To ex-

Table 1: Patient Hospitalization Statistics

Year	Positive Instances[1]	Negative Instances[2]
2012	4,859	35,255
2013	5,285	35,962
2014	5,464	37,853
Total	**15,608**	**109,070**

[1] Patients readmitted within 30-days.
[2] All others.

tract diabetes related instances, we employed Apache Spark, a big data computing engine, to extract the data within a reasonable time. ICD-9 codes and values of primary diagnosis codes (*DX_1*) were used to filter out patients with a primary diagnosis of diabetes. The readmission status for each visit is not directly available from the attributes provided by the SID database. To this end, we first utilized the variable *VisitLink* to identify patients with multiple visits and then calculated the number of days between two consecutive hospital admissions leveraging the variable *DaysToEvent*. Specifically, an SID record contains a randomly assigned unique "start date" for each patient. Another timing variable *DaysToEvent* was calculated for each admission as the difference between the patient's admission date and the "start date". Because the "start date" is unique for each patient, it serves as an encrypted patient identifier. Thus, the number of days between a patient's two consecutive visits can be calculated by first obtaining the difference of the two (in tandem) *DaysToEvent* values and then adjusting the difference by the length of stay (*LOS*) of the first visit, i.e.,

$$\mathit{DaysInBetween} = \mathit{DaysToEvent2} - \mathit{DaysToEvent1} - \mathit{LOS1}$$

All admissions with *DaysInBetween* ≤ 30 are assigned a class label = "Y" and they form the positive training instances; all others form the negative training instances with a class label = "N". As a result, our dataset has a total of 15,608 and 109,070 positive and negative instances respectively. Table 1 presents the statistics of the total number of positive and negative instances in each year.

3.2 Feature Engineering

Our dataset provides more than 450 variables capturing clinical and resource-use information which are included in a typical discharge abstract. A dimensionality reduction on the data is necessary for the purpose of both noise removal and computational feasibility. Based on the description of each variable in the HCUP documentation, we finalized our dataset with 77 variables after eliminating attributes that were irrelevant to our study. For example, we removed variables related to patients' payment information, physician IDs, zip codes, non-primary diagnosis codes, and procedures, etc.

In addition to reducing the dimensionality of our data, we added an attribute (*TVISITS*) reflecting the total number of visits a patient had prior to a hospitalization. Our conjecture is that a higher hospitalization frequency implies a higher likelihood of early readmission. Our experiment results confirm the effectiveness of this feature (see Section 6 for details). Finally, because some of the machine learning algorithms are sensitive to the scales of the variables, we applied normalization to all numeric columns in the dataset.

3.3 Imbalanced Data

A particular challenge in predicting early readmission outcome is that correctly classifying patients of the "Y" class is more important than correctly classifying patients of the "N" class. This is because a "Y" classification engenders closer monitoring. Thus an "N" misclassified as a "Y" only costs more money and time in terms of in-hospital observations, whereas a "Y" misclassified as an "N" (type I error) will result in much more severe financial and operational consequences including penalties from the Centers for Medicare & Medicaid Services (CMS).

From Table 1, we observe that the total number of instances of the "Y" (i.e., "readmitted") class and "N" (i.e., "no-readmitted") class is 15,608 and 109,070 respectively. Standard machine learning algorithms assume that the training samples are equally distributed among the classes (Chawla et al., 2004). A class imbalance occurs when the instances of one class outnumber the instances of the other classes. The class with overwhelming instances is called the majority class while the other called the minority class. Applying standard machine learning algorithms to an imbalanced dataset often leads to insufficient performance on the minority class which is often the more interesting and important class under investigation. Indeed, the primary interest of our classification task is to accurately predict patients of the minority class. To address the class imbalance issue, we employed bootstrap aggregating (a.k.a bagging) with random undersampling technique [1]. In particular, we generated ten "bags" of balanced datasets where each "bag" contained all minority class instances and an equal number of randomly sampled majority class instances. Each subset of the majority instances was sampled with replacement from the entire majority population. For each of our baseline models described in Section

[1] Other class imbalance correction techniques (e.g., SMOTE) were explored and resulted in worse performance.

4.1, ten sub-models were trained using these balanced "bags" of data. The final model outputs were obtained by aggregating the results of these ten sub-models using a majority vote.

4 METHODS

In this section, we describe the methods we used to conduct our study. We present a performance comparison of these models and analysis of the results in Section 5.

4.1 Baseline Models

We selected four established and popular machine learning algorithms as our baseline learners: Support Vector Machine (SVM) (Cortes and Vapnik, 1995), Neural Network (Gurney, 2014), Logistic Regression (Menard, 1995), and Random Forest (Breiman, 2001). These methods were used by most of the literature discussed in Section 2. We describe two additional ensemble learning models in Sections 4.2 and 4.3 respectively.

4.2 Meta-learner L

We employed an ensemble technique (Dietterich, 2000) to integrate information from the four base classifiers described in Section 4.1. Ensemble learning is a family of algorithms that seek to create a "strong" classifier based on a group of "weak" classifiers. In this context, "strong" and "weak" refer to how accurate the classifiers can predict the target variable. Ensemble learning has been proven to produce improved and more robust performance than a single model. Our meta-learner L is an example of *heterogeneous* ensemble because its base learners are obtained from different machine learning algorithms. Our next model, *XGBoost*, explores the efficacy of a *homogeneous* ensemble where the base classifiers are obtained using a single machine learning algorithm.

4.3 XGBoost

We investigated the performance of *XGBoost* (Chen and Guestrin, 2016), an algorithm that has gained much popularity and attention since its inception in 2016 and was the winning algorithm for a number of machine learning competitions. *XGBoost* belongs to the family of *homogeneous* ensemble methods in which the base learners, $L_1, L_2, \ldots, L_n$, are created using a single machine learning algorithm exploiting the concept of "adaptive boosting". (Freund et al.,

Figure 1: ROC Curves

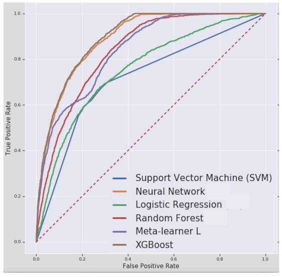

Table 2: AUC Scores

Model	AUC Score
SVM	0.69
Neural Network	**0.80**
Logistic Regression	0.70
Random Forest	0.75
Meta-L	0.71
XGBoost	**0.80**

1999) Specifically, a sequence of classifiers is generated with the new model aiming to correct the errors of the previous model. This correction is typically achieved by boosting the weights of the misclassified instances in the previous model so that the new model will have a higher likelihood of correctly classifying them. New models are added sequentially until no further improvements can be achieved. In *XGBoost*, instead of boosting the weights, the algorithm fits the new model to residuals of the previous model and then minimizes the loss when adding the latest model. The process is equivalent to updating your model with a gradient descent towards a local optimum solution.

5 EXPERIMENTAL RESULTS

All experiments were conducted by running 10-fold cross-validation. Data was divided into ten folds and we iteratively trained on nine folds and tested on the remaining fold. The performance of the model was measured by calculating the average accuracy over all test folds. In addition to overall predictive accuracy, *Sensitivity* and *Specificity* were used to measure the performance in the positive and negative classes respectively. The hyper-parameters in this study were selected via a grid search with the highest Area Under the Curve (AUC) score of a Receiver Operating Characteristic (ROC) curve (Fawcett, 2006) on a validation set. We present the AUC scores of each model and the corresponding ROC curves in Table 2 and Figure 1 respectively.

Table 3 presents the main results of our experiment. As discussed in Section 3.3, our goal is to accurately predict the "Y" (i.e., "readmitted") instances. To this end, we adjusted the threshold in the ROC curve to increase a model's *Sensitivity* at the cost of lowering the *Specificity*. For each threshold displayed in column 1 of Table 3, we present a performance comparison of the six models described in Section 4 using overall accuracy, *Sensitivity* and *Specificity*. Consequently, we can observe the trade-offs between an increase in *Sensitivity* and a decrease in *Specificity*

Table 3: Performance Comparison of Six Models

Threshold	Model	Sensitivity	Specificity	Overall
0.5	SVM	0.61	0.77	0.75
	Neural Network	**0.84**	**0.76**	**0.77**
	Logistic Regression	0.63	0.76	0.75
	Random Forest	0.74	0.75	0.75
	Meta-L^*	0.64	0.78	0.76
	XGBoost	**0.82**	**0.78**	**0.79**
0.45	SVM	0.62	0.76	0.75
	Neural Network	**0.87**	**0.73**	**0.74**
	Logistic Regression	0.70	0.69	0.69
	Random Forest	0.93	0.57	0.60
	Meta-L^*	0.68	0.75	0.74
	XGBoost	**0.87**	**0.74**	**0.76**
0.4	SVM	0.64	0.75	0.74
	Neural Network	**0.89**	**0.70**	**0.72**
	Logistic Regression	0.77	0.60	0.61
	Random Forest	0.99	0.37	0.43
	Meta-L^*	0.76	0.72	0.72
	XGBoost	**0.90**	**0.70**	**0.72**
0.35	SVM	0.64	0.75	0.74
	Neural Network	**0.91**	**0.67**	**0.69**
	Logistic Regression	0.84	0.49	0.52
	Random Forest	1.00	0.14	0.23
	Meta-L^*	0.84	0.65	0.67
	XGBoost	**0.94**	**0.66**	**0.69**
0.3	SVM	0.66	0.73	0.72
	Neural Network	**0.96**	**0.62**	**0.65**
	Logistic Regression	0.89	0.37	0.43
	Random Forest	1.00	0.0	0.12
	Meta-L^*	0.92	0.57	0.60
	XGBoost	**0.97**	**0.61**	**0.65**

Meta-*L*: ensemble of SVM, Neural Network, Logistic Regression and Random Forest.

for each model as we shift the threshold. Since misclassifications in the "N" class incur less cost than misclassifications in the "Y" class (see Section 3.3), a healthcare institution can select a desired threshold depending on its level of tolerance on the insufficient performance of the "N" class (i.e., low *Specificity*). In addition, we observe that:

- *XGBoost* and *Neural Network* are comparable in terms of predictive power. They both deliver superior performance compared to the other four models in our study. Nevertheless, *XGBoost* is capable of achieving a slightly higher *Sensitivity* when we shift the threshold ≤ 0.4. In addition, *XGBoost* is a tree-based algorithm which provides model interpretability. Together with its ability to support parallel computing for big data processing, *XGBoost* is our model of choice for our task.

- Although Meta-learner *L* is an algorithm under the ensemble learning paradigm, it did not outperform all of its baseline learners as expected. One explanation is that ensemble learning is effective only if the baseline learners possess enough diversity. Indeed, from Figure 1, we observe that two (*Logistic Regression* and *SVM*) out of the four base learners do not display enough diversity. This could also explain why *heterogeneous* ensemble, although a well established machine learning technique, is not regularly present in the literature of hospital readmission studies. Our experimental results demonstrate that a *homogeneous* ensemble learner (i.e., *XGBoost*) is an effective choice for our classification task.

- It is worth noting that *SVM* was not very sensitive to the threshold values in our experiment. On the other hand, *Logistic Regression* and *Random Forest* degenerated as we lowered the threshold. This is particularly the case with *Random Forest* where the model started to predict every instance belonging to the "Y" class when the threshold dropped below 0.4. We conclude that *Neural Network*, Meta-learner *L*, and *XGBoost* are the more robust algorithms among our models.

6 Risk Factor Analysis

We next investigated the major factors that contribute to the risk of early readmission. Four linear

Table 4: List of Top 10 Predictive Features

Rank	SVM	Logistic Regression	Random Forest	XGBoost
1	**LOS**	**LOS**	**TVISITS**	**AGE**
2	NPR	**AGE**	**AGE**	**TVISITS**
3	**AGE**	**TVISITS**	CM_DRUG	**NDX**
4	**TVISITS**	NPR	**NDX**	**LOS**
5	**NDX**	**NDX**	**LOS**	NCHRONIC
6	PRDAY1	PRDAY1	NCHRONIC	PRDAY1
7	NCHRONIC	NCHRONIC	PRDAY1	NPR
8	NECODE	NECODE	CM_ANEMDEF	FEMALE
9	HCUP_OS	HCUP_OS	NPR	CM_LYTES
10	CM_ANEMDEF	CM_DRUG	CM_OBESE	CM_ANEMDEF

AGE:	patient's age at admission.
CM_ANEMDEF:	comorbidity measure for deficiency anemias.
CM_DRUG:	comorbidity measure for drug abuse.
CM_LYTES:	comorbidity measure for fluid and electrolyte disorders.
CM_OBESE:	comorbidity measure for obesity.
FEMALE:	indicator for female patient.
HCUP_OS:	have evidence of observation stay (OS) services.
LOS:	length of stay.
NCHRONIC:	total number of chronic condidtions.
NDX:	total number of diagnosis on this record.
NECODE:	total number of external injuries.
NPR:	total number of procedures on this record.
PRDAY1:	number of days from admission to first lab procedure.
TVISITS:	total number of previous hospital visits.

and tree-based algorithms, *SVM*, *Random Forest*, *Logistic Regression* and *XGBoost*, were selected for the study. These models were chosen because their feature importance was well defined. In particular, for linear models, the importance is proportional to the magnitude of the coefficients. For tree-based models, the ranking follows the order of attributes that the algorithm selected to split the branches. Table 4 presents the top ten predictive features identified by each of the four models.

- Examining the top five risk factors associated with each model, we identify four consistent principal predictors (highlighted in bold) across all models, namely a patient's age (*AGE*), total number of previous hospital visits (*TVISITS*), a patient's length of stay (*LOS*), and the total number of diagnosis during the current hospitalization (*NDX*). In particular, *SVM* and *Logistic Regression* are more dependent on variables *LOS* and *AGE*, while *Random Forest* and *XGBoost* rely more on attributes *TVISITS* and *AGE*. *NDX* is the next important risk factor to watch out for because, although mostly towards the lower end, it is ranked within the top five by all models.
- Expanding our investigation to the top 10 risk factors associated with each model, we could identify another three common risk factors across all models, namely a patient's total chronic conditions (*NCHRONIC*), total number of lab procedures (*NPR*), and number of days from admission to the first lab procedure (*PRDAY1*).
- In addition to the above total of seven common risk factors revealed by all models, a patient's comorbidity measure for deficiency anemias is another important attribute which appeared in three (*SVM*, *Random Forest*, *XGBoost*) out of the four models.
- Furthermore, *SVM* and *Logistic Regression* rely on a patient's number of external injuries (*NECODE*) and the availability of observatory stay services (*HUP_OS*) in making their predictions, whereas *Random Forest* and *XGBoost* utilize three additional comorbidity measures, i.e., drug abuse (*CM_DRUG*), fluid and electrolyte disorders (*CM_LYTES*) and obesity (*CM_OBESE*).
- Lastly, a patient's gender (*FEMALE*) appeared as an important predictor for *XGBoost* algorithm. We did find that female patients carry a slightly higher risk as compared to male patients (52.7% vs. 47.3%).

7 CONCLUSION

In this paper, we applied machine learning techniques to predict 30-day hospital readmissions for patients with diabetes based on their clinical and demographic characteristics. We built our model using empirically discharge records of 124,678 patients extracted from the State Inpatient Database (SID) of Florida. We employed four baseline machine learning algorithms and two ensemble learners in our study. We further addressed the data imbalance issue using the bootstrap aggregating method. Our experimental results demonstrate that *XGBoost* and *Neural Network* provide comparable predictive power for our task, and their performances are significantly better than the other four models' in our study. Nevertheless, we recommend *XGBoost* because of its interpretability and computational advantage.

In addition, we examined the top risk factors identified by our linear and tree-based models. We conclude that a patient's age (*AGE*), number of previous hospital visits (*TVISITS*), length of stay in the hospital (*LOS*), and the total number of diagnosis during hospitalization (*NDX*) are the top four predictive variables. In addition, a patient's total number of chronic conditions (*NCHRONIC*), the total number of in-patient lab procedures (*NPR*), and the number of days from admission to first lab procedure (*PRDAY1*) are also significant indicators in assessing a patient's risk of early admission.

Although our work focused on addressing early readmission risk of diabetic patients, our models are directly applicable to other patient cohorts. We foresee one future research direction would be to apply similar study to predict the readmission risk associated with either all-cause or other high stake diseases, such as Chronic Obstructive Pulmonary Disorder (COPD), Heart Failure (HF), Pneumonia (PN), etc. In addition, exploring other state of the art machine learning techniques, such as deep learning, could potentially bring valuable insights and fruitful results.

REFERENCES

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

CDC (2009). Number (in thousands) of hospital discharges with diabetes as any-listed diagnosis, united states, 1988–2009. *Available at:)(Accessed 12 February, 2014)*.

CDC (2012). Increasing prevalence of diagnosed diabetes–united states and puerto rico, 1995-2010. *MMWR. Morbidity and mortality weekly report*, 61(45):918.

CDC (2017). National diabetes statistics report, 2017. *Atlanta, GA: Centers for Disease Control and Prevention*.

Chawla, N. V., Japkowicz, N., and Kotcz, A. (2004). Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*, 6(1):1–6.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM.

Cortes, C. and Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3):273–297.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer.

Duggal, R., Shukla, S., Chandra, S., Shukla, B., and Khatri, S. K. (2016). Predictive risk modelling for early hospital readmission of patients with diabetes in india. *International Journal of Diabetes in Developing Countries*, 36(4):519–528.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.

Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612.

Friedman, B., Jiang, H. J., and Elixhauser, A. (2008). Costly hospital readmissions and complex chronic illness. *Inquiry: The Journal of Health Care Organization, Provision, and Financing*, 45(4):408–421.

Frizzell, J. D., Liang, L., Schulte, P. J., Yancy, C. W., Heidenreich, P. A., Hernandez, A. F., Bhatt, D. L., Fonarow, G. C., and Laskey, W. K. (2017). Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA cardiology*, 2(2):204–209.

Futoma, J., Morris, J., and Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. *Journal of biomedical informatics*, 56:229–238.

Golas, S. B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., Hisamitsu, T., Kojima, G., Felsted, J., Kakarmath, S., et al. (2018). A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC medical informatics and decision making*, 18(1):44.

Gurney, K. (2014). *An introduction to neural networks*. CRC press.

Menard, S. W. (1995). *Applied logistic regression analysis*. Number 04; e-book.

Mortazavi, B. J., Downing, N. S., Bucholz, E. M., Dharmarajan, K., Manhapra, A., Li, S.-X., Negahban, S. N., and Krumholz, H. M. (2016). Analysis of machine learning techniques for heart failure readmissions. *Circulation: Cardiovascular Quality and Outcomes*, pages CIRCOUTCOMES–116.

Rubin, D. J. (2015). Hospital readmission of patients with diabetes. *Current diabetes reports*, 15(4):17.

Van Walraven, C., Bennett, C., Jennings, A., Austin, P. C., and Forster, A. J. (2011). Proportion of hospital readmissions deemed avoidable: a systematic review. *Canadian Medical Association Journal*, 183(7):E391–E402.