Using Deep Feedforward Neural Networks to Predict Hospital Readmission in Diabetic Patients

Amy Howe
CKME136

Ryerson University

Table of Contents

Introduction

Diabetes mellitus (DM) is a widespread metabolic disorder affecting 10.5% of the US population in 2018 (Centers for Disease Control and Prevention [CDC], 2020). In the US, the rate of hospital readmission within 30 days is almost twice as high for people with diabetes when compared to the general population (Ostling et al., 2017). Rates of hospital readmission are often used as an indicator of quality of care; reducing preventable readmissions is a common target for policies and quality improvement initiatives (Wadhera et al., 2019). Predicting at the point of hospital discharge whether a patient will be readmitted can allow for allocation of resources to these high risk individuals to reduce readmissions. Additionally, identifying pre-discharge risk factors with the largest impact on readmissions in diabetic patients may highlight the need for targeted interventions.

Artificial neural networks have been shown to outpredict classic models like logistic regression in binary classification (Liu et al., 2020). While these have been used before in prediction of all-cause readmission for patients, they have not focused specifically on diabetic patients or attributes pertinent to these cases (Liu et al., 2020; Jamei et al., 2017). Deep feedforward neural network algorithms will be used as binary classifiers to predict whether a patient will be readmitted to hospital within 30 days based on pre-discharge factors such as demographics, diagnoses, change in diabetic medications, and reason for discharge. Model hyperparameters will be tuned and measured using 10-fold cross validation to identify the best-performing predictive neural network model. The performance of this model will be compared against several traditional classifiers: k-nearest neighbours, random forest, and logistic regression. The best-performing model will be used to identify risk factors that have the largest impact on readmission.

Literature Review

While there are studies looking at risk factors for readmission in diabetic patients, most haven't focused on prediction (Kim et al., 2011; Strack et al., 2014; Ostling et al., 2017). Those that did had limited success, generally using multivariable logistic regression (Dungan, 2012; Rubin et al., 2017; Karunakaran et al., 2019). Other studies found greater success using neural networks in predicting readmission, although did not focus specifically on diabetic patients, nor did they use real-time data which reduces the practical applications of these models (Jamei et al., 2017; Liu et al., 2020). Factors modifiable during hospital stay were often missed, such as specific diabetic medications being taken (e.g., insulin), missing an important area for targeted interventions (Kansagara et al., 2011; Kim et al., 2011; Jamei et al., 2017; Ostling et al., 2017; Rubin et al., 2017; Karunakaran et al., 2019; Liu et al., 2020). Some studies were also performed at single health care organizations, limiting the generalizability of findings (Ostling et al., 2017; Rubin et al., 2017; Karunakaran et al., 2019).

The researchers who compiled the dataset to be used explored the importance of HbA1c (glycated hemoglobin) assessment in hospital for diabetic patients as a predictor for 30-day hospital readmission

(Strack et al., 2014). Using multivariable logistic regression, they compared HbA1c measurement with readmissions while controlling for other factors like demographics. It was found that in the majority of cases (81.6%), HbA1c was not measured in hospital, and that in cases where it was measured, readmission was less likely to occur. The study itself did not focus on prediction of readmission, or on other predictor variables available.

Other studies have looked at risk factors for readmission in diabetic patients as well, albeit without focusing on prediction. Some key risk factors for readmission identified by Karunakaran et al. (2019) included length of stay, recent hospital visit, lack of visit after discharge, leaving against medical advice (AMA), demographics, diagnosis, and lab values on admission. Ostling et al. (2017) found that if patients had diabetes as a primary diagnosis, the most common reason for readmission would be diabetes-related. They also found that patients followed by health system diabetes services had a lower rate of emergency department use, but no change in inpatient readmission rates. Both studies were completed at single organizations, and as such, results may not be generalizable to other health care facilities. Kim et al. (2011) looked at hospital readmissions for diabetic patients grouped into scheduled and unscheduled visits. It was found that the majority of readmissions were unscheduled (87.2%), with predictors varying between the two groups. Predictors identified through logistic regression for unscheduled readmissions included comorbidities, having public insurance, living in a low income area, being an ethnic minority, and having a recent history of hospitalization. This study only looked at adults aged 50 and older, and therefore results may not be generalizable to people of all ages with diabetes. In all three studies, modifiable risk factors such as changes in diabetic medications were not reviewed, despite being a potentially useful avenue for targeted interventions prior to hospital discharge.

Multiple studies did focus on prediction of readmission in patients with diabetes, but with limited success, often using multivariable logistic regression. Rubin et al. (2017) used logistic regression to predict readmission in diabetic adults, but only those who were specifically admitted for cardiovascular disease. The c-statistic measured was 0.71, and further identified predictors of readmission included socioeconomic factors such as previous education and employment, as well as address within five miles, recent hospital visit, admission lab values, previous diabetes therapy and complications, and mental illness. A meta-analysis of articles was conducted reviewing the link between diabetes and hospital readmissions (Dungan, 2012). While confirming predictors for readmission include demographics, recent admissions, comorbidities, and lack of follow-up after discharge, it was also noted that predictive models have had limited success, with c-statistics ranging from 0.658 to 0.68. Additionally, hospital glycemic control was identified as an area of need for continued research to determine its impacts on readmission.

Several studies have focused on predicting hospital readmissions using neural networks, albeit not for diabetic patients specifically. Liu et al. (2020) used two deep feedforward neural network variants to predict hospital readmissions using claims data. It was found that both neural networks used outperformed the logistic regression variants created, with the best neural network performance including the embedding of diagnosis codes. Unfortunately, claims data is not available in real time, which limits the use of such a predictor in practice. Jamei et al (2017) compared shallow feed-forward neural network algorithm performance with a common manual tool (LACE) in predicting the risk of hospital readmission. The optimized neural network outperformed LACE with a 20% higher precision and a c-statistic of 0.78. Features were supplemented by census data, which would likely not be readily available in practice for real time predictions. Additionally, Kansagara et al. (2011) completed a systematic review of prediction models created for the risk of all-cause hospital readmission. Overall, predictive power of these statistical models was found to be poor, ranging from 0.55 - 0.83 c-statistic, with only five of the 26 models able to be used at the point of discharge. No models included use of medications as variables, and researchers also noted the lack of variables associated with severity of illness and the social determinants of health.

Overall, studies have created prediction models for readmission of diabetic patients with limited success, often using data that would not be available in real-time. Most specific to diabetes were also focused on patients in a certain age group, and missed modifiable attributes in hospital that may provide targets for intervention. Using artificial neural networks to predict readmission in all-ages diabetic patients considering factors available at the point of discharge could improve model performance, provide a practical application for model use, and identify modifiable risk factors in hospital.

Dataset

The dataset used spans 101,766 encounters with 50 attributes across 130 hospitals in the United States (US), and was compiled by Strack et al. for their 2014 study (available in their Supplementary Materials) and is available in the UCI Machine Learning Repository (2014). It is also available, along with all coding for this project, on GitHub (https://github.com/amyahowe/diabetes_readmission). See Table 1 for a description of each attribute included in the original dataset. Researchers gathered the data from Cerner's Health Facts Database in the US (Strack et al., 2014). The data spans the years 1999-2008, and originates from hospitals and health centres all over the US ranging from bed sizes less than 100 to greater than 500. Of note, data from out-of-network providers is not available. The criteria for data extracted were that records captured an inpatient encounter for a patient with diabetes documented as a diagnosis, with a length of stay between one and 14 days, where at least one lab test was performed and at least one medication was administered.

Table 1: Original Attributes from Raw Dataset with Descriptions

Attribute(s)	Data Type	Description
Encounter ID	Numeric	ID associated with the patient's unique encounter (hospital visit)
Patient Number	Numeric	Patient ID; each patient has a consistent unique ID to

		identify them in the system
Race	Categorical	Race of the patient
Gender	Categorical	Gender of the patient
Age	Ordinal	Age of patient in years, grouped into categories of ten years
Weight	Ordinal	Weight of patient in pounds, grouped into categories of 25 pounds
Admission Type ID	Categorical	Type of admission (e.g., urgent)
DIscharge Disposition ID	Categorical	Reason for discharge, or where the patient was discharged to (e.g., discharged home)
Admission Source ID	Categorical	Method of admission to hospital (e.g., through the emergency department)
Time in Hospital	Numeric	Number of days from admission to discharge; also known as length of stay
Payer Code	Categorical	Code indicating who the payer is (e.g., private insurance, Medicaid)
Medical Specialty	Categorical	Specialty the patient is admitted under (e.g., cardiology)
Number of Lab Procedures	Numeric	Number of lab tests completed during stay
Number of Procedures	Numeric	Number of procedures excluding lab tests completed during stay
Number of Medications	Numeric	Number of distinct medications the patient was administered during stay
Number Outpatient	Numeric	Number of outpatient encounters in the previous year
Number Emergency	Numeric	Number of emergency encounters in the previous year

Number Inpatient	Numeric	Number of inpatient encounters in the previous year
Diagnosis 1, 2, 3	Categorical	Three attributes: primary, secondary, and tertiary diagnosis. All diagnoses are expressed in ICD-9-CM codes
Number of Diagnoses	Numeric	Total number of diagnoses recorded
Max Serum Glucose	Ordinal	Maximum blood glucose level during stay (mg/dL) split into groups: None, Normal, >200, >300
A1C Result	Ordinal	Glycated hemoglobin (HbA1c) test result, measured in percentage, split into groups: None, Normal, >7, >8
Specific Medications	Categorical	23 attributes: Data indicates whether a diabetic medication was increased, decreased, the dosage stayed the same, or the medication was not prescribed for the following medications: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, citoglipton, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, metformin-pioglitazone
Change in Medication	Categorical	Indicates whether or not there has been a change in any diabetic medication during the patient's stay
Diabetes Medication	Categorical	Indicates whether the patient is taking any diabetic medications or not
Readmitted	Categorical	Indicates whether the patient was readmitted within 30 days, readmitted after 30 days, or if there was no record or readmission.

For the purpose of this project, out of scope cases were removed. It was identified that of the 101,766 encounters, there were only 71,518 unique patient IDs. In order to focus on early identification of risk factors for readmission, and avoid violating statistical assumptions of independence for analysis, any encounter record that was not a patient's first was removed. Additionally, any record where the patient died or was discharged to a hospice was removed, as these patients are not eligible for readmission. This brought the total of records from 71,518 down to 69,973.

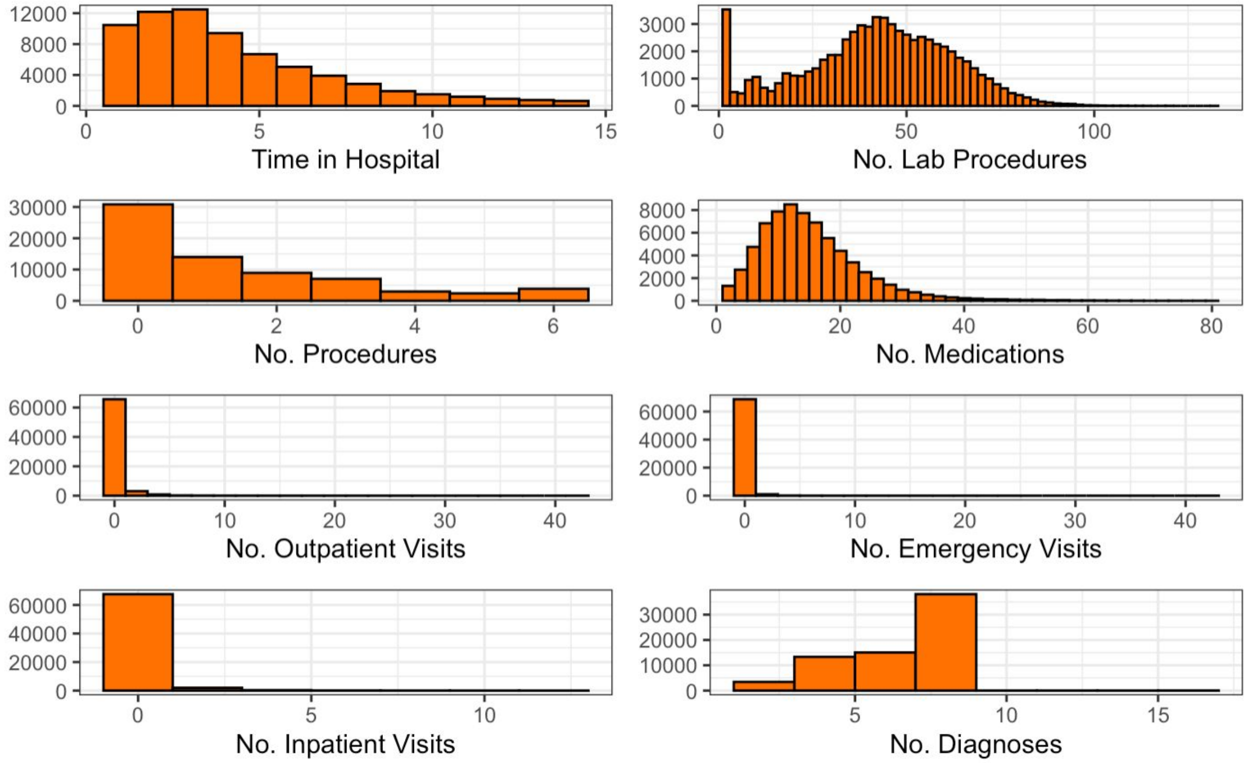
Attributes with a large number of missing values were removed: Weight (96.0% missing), Medical Specialty (48.1% missing) and Payer code (43.5% missing). Any attribute in which one category spanned 95% of records or greater was also removed due to low variance. This included: Max Serum Glucose, and 16 of the 23 medications (Repaglinide, Nateglinide, Chlorpropamide, Acetohexamide, Tolbutamide, Acarbose, Miglitol, Troglitazone, Tolazamide, Examide, Citoglipton, Glyburide-metformin, Glipizide-metformin, Glimepiride-pioglitazone, Metformin-rosiglitazone, and Metformin-pioglitazone).

Summary statistics for the numeric attributes are available in Table 2. Histograms of numeric attributes are shown in Figure 1. Encounter ID and Patient Number are excluded from analysis, as they provide no further information. There were no missing values for any of the numeric attributes, and all numeric attributes were discrete (whole numbers).

Table 2: Numeric Attribute Summary Statistics

Attribute	Summary Statistics						
	Min	Q1	Median	Mean	Q3	Max	Standard Deviation
Time in Hospital	1	2	3	4.273	6	14	2.934
Number of Lab Procedures	1	31	44	42.88	57	132	19.895
Number of Procedures	0	0	1	1.426	2	6	1.757
Number of Medications	1	10	14	15.67	20	81	8.287
Number Outpatient	0	0	0	0.2795	0	42	1.064
Number Emergency	0	0	0	0.1039	0	42	0.512
Number Inpatient	0	0	0	0.1763	0	12	0.602
Number of Diagnoses	1	6	8	7.224	9	16	2.001

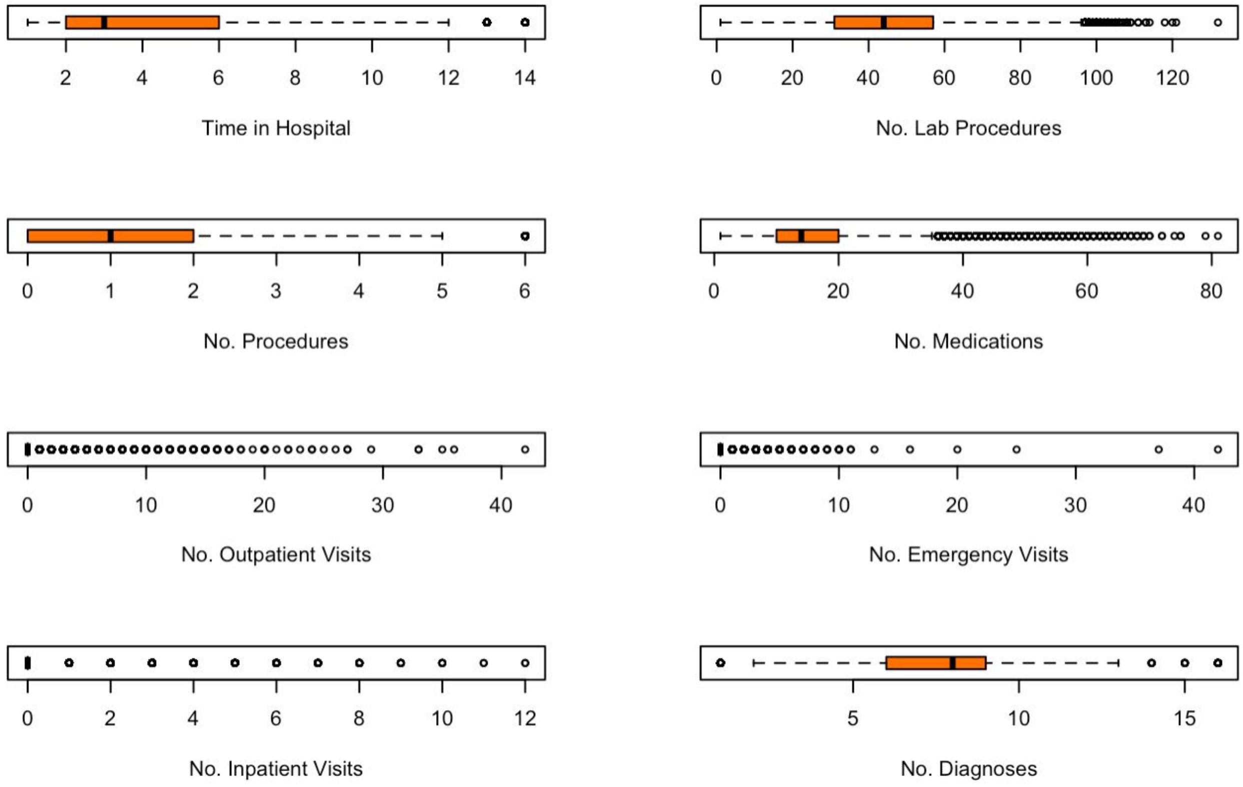
Figure 1: Numeric Attribute Histograms



As is visible in the attribute histograms [Figure 1], no numeric attributes appear to be normally distributed, although Number of Lab Procedures and Number of Medications may be close. Most are skewed right, with the exception of the Number of Diagnoses. The Number of Diagnoses attribute presents an interesting case, as the distribution is skewed left from values 1 to 9, with a small number of cases (30 or less) for each number above 9 from 10 to 16. Since there is no biological reason people would stop developing disease conditions after their ninth, it is likely that the way diagnoses are coded makes it unusual or difficult to code any more than nine.

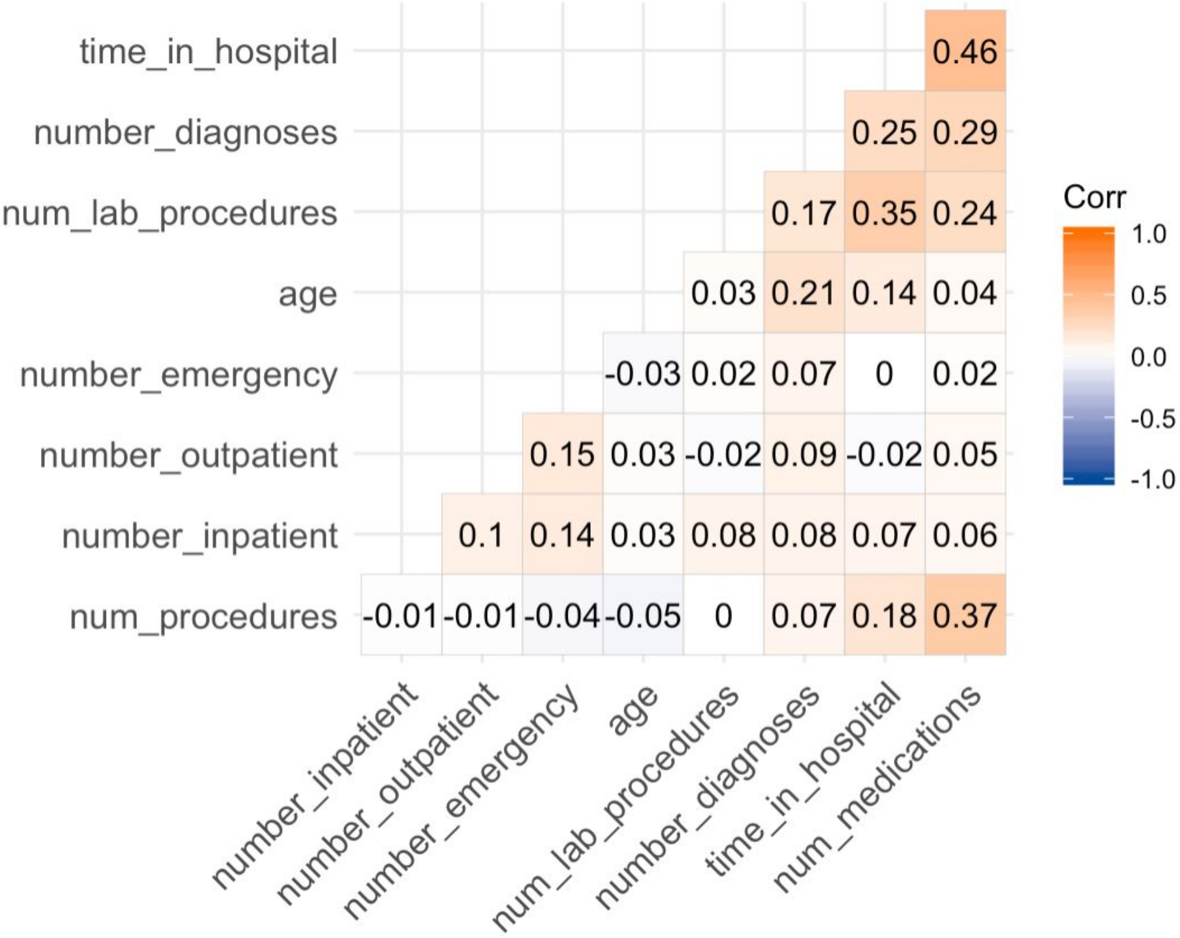
The numeric attributes have a number of outliers [Figure 2]. All outliers reviewed appear to be within reason, and therefore may represent actual cases. Given that neural networks may be robust to the presence of some outliers, and that 33.1% of records in the sample include at least one outlier, these were kept in for modeling.

Figure 2: Numeric Attribute Boxplots



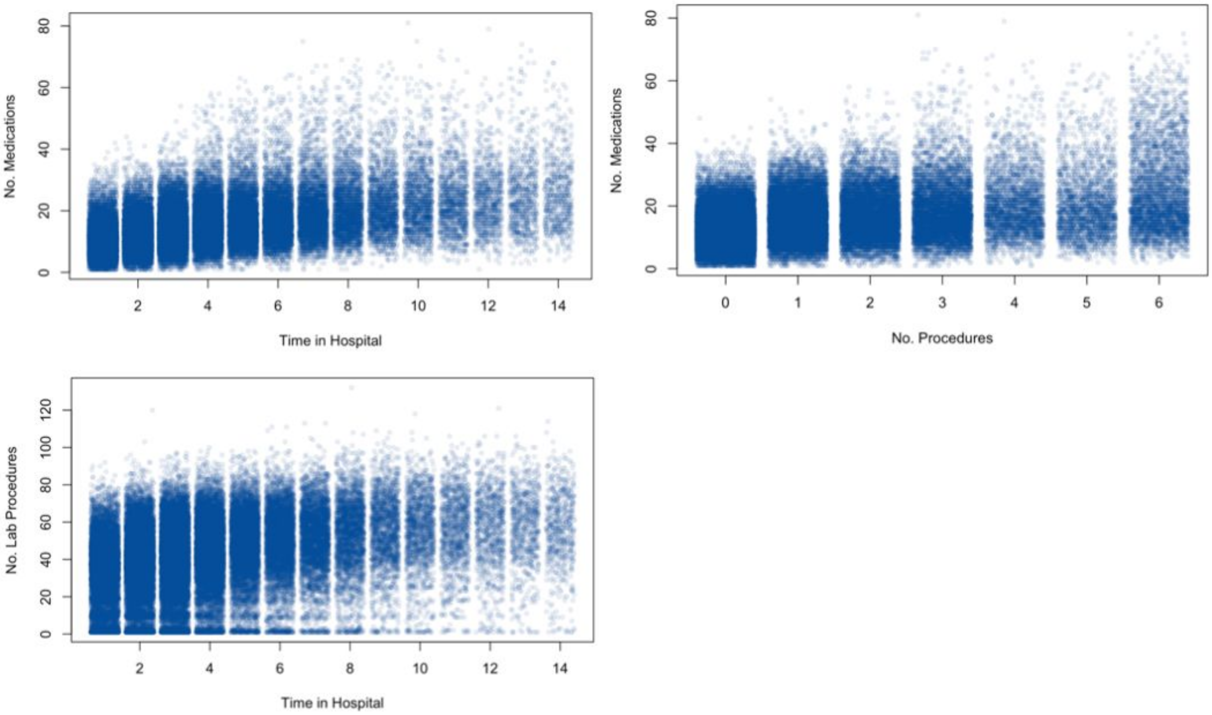
Spearman correlation was used to assess correlation between numeric attributes and the one remaining ordinal attribute (Age). None of the attributes are strongly correlated with one another [Figure 3]. Time in Hospital is moderately positively correlated with both Number of Medications (r_s = 0.46) and Number of Lab Procedures (r_s = 0.35). These are both expected, as the longer someone is in hospital, the more time there is to administer medications and lab tests. Additionally, there is a moderate positive correlation between Number of Procedures and Number of Medications (r_s = 0.37). This is also expected, as some procedures require additional medications. Two cases of zero correlation were found, but neither were statistically significant at an alpha of 0.05: Number of Procedures and Number of Lab Procedures (p = 0.309), and Time in Hospital and Number Emergency (p = 0.283). All other correlations were statistically significant.

Figure 3: Spearman Correlation of Numeric and Ordinal Attributes



Scatterplots for the three moderately correlated numeric variables are available in Figure 4. The points were made transparent and jitter was added to x-axis variables in order to better visualize the trends. Time in Hospital against Number of Medications appears to have a positive linear relationship. Although, as the number of days in hospital increases, there appear to be more outliers in the number of medications present. Similarly, Number of Procedures against Number of Medications also appears to have a positive linear relationship. For Time in Hospital against Number of Lab procedures, the positive monotonic relationship is visible, although it may be better represented by a concave down, increasing curve. This indicates that as the time in hospital increases, the increase in number of lab procedures completed decreases in rate. This is an expected finding, as more lab tests would be completed soon after admission, when the medical issue(s) are being initially investigated, with fewer being completed later.

Figure 4: Scatterplots for Numeric Variable Combinations with Moderate Correlation



Of the 50 attributes, 36 of them were categorical and four were ordinal. As mentioned above, the three with high incidences of missing values were removed (Weight, Medical Specialty, Payer Code). The attribute A1C Result was converted from ordinal to categorical by reducing the "<7" and "<8" categories into one category for abnormal results. The missing values ("None") indicating the test was not taken was kept as a category for analysis due to its medical significance. There were eight categorical attributes that contained missing values. Missing values were imputed into the majority class. These are: Race (2.7% missing) imputed to Caucasian, Gender (0.004% missing) imputed to Female, Admission Type ID (11.3% missing) imputed to Emergency, Discharge Disposition ID (4.6% missing) imputed to Discharge Home, Admission Source ID (7.1% missing) imputed to Emergency Room (ER), Diagnosis 1 (0.001% missing) imputed to 390-459 (circulatory disorders), Diagnosis 2 (0.4% missing) imputed to 390-459 (circulatory disorders), and Diagnosis 3 (1.8% missing) imputed to 390-459 (circulatory disorders). For the Diagnosis attributes, it may be possible that data was missing because the patient doesn't have a diagnosis; however, there is no way to distinguish between this case and the information existing but being missing.

All remaining categorical and ordinal attributes that were not removed due to missing values or low variance are shown in Table 3 with their category names, frequency and percentage frequency of each category.

Table 3: Categorical and Ordinal Attribute Categories with Frequency and Percentage Frequency

Attribute	Categories	Frequency	Percentage Frequency
Race	Caucasian African American Hispanic Asian Other	54,210 12,652 1,500 488 1,150	77.5% 18.0% 2.1% 0.7% 1.6%
Gender	Female Male	37,323 32,741	53.2% 46.8%
Age	[0-10) [10-20) [20-30) [30-40) [40-50) [50-60) [60-70) [70-80) [80-90) [90-100)	153 534 1,121 2,692 6,828 12,349 15,684 17,750 11,102 1,760	0.2% 0.8% 1.6% 3.8% 9.8% 17.6% 22.4% 25.4% 15.9% 2.5%
Admission Type ID	Emergency Elective Urgent Other	43,359 13,785 12,802 27	62.0% 19.7% 18.3% 0.03%
Discharge Disposition ID	Discharge Home Discharge to Skilled Nursing Facility (SNF) Discharge Home with Home Health Services Discharge to Other Facility Other	47,569 8,784 8,362 4,787 471	68.0% 12.6% 12.0% 6.8% 0.7%
Admission	Emergency Room (ER)	42,328	60.5%

Source ID	Physician Referral	21,746	31.1%
	Transfer from External Facility	5,880	8.4%
	Other	19	0.03%
Diagnosis 1	390-459: “Diseases Of The Circulatory System” (International Classification of Diseases [ICD], 2020)	21,326	30.5%
	460-519: “Diseases Of The Respiratory System”	6,446	9.2%
	520-579: “Diseases Of The Digestive System”	6,325	9.0%
	250.0-250.9: “Diabetes Mellitus”	5,748	8.2%
	780-799: “Symptoms, Signs, And Ill-Defined Conditions”	5,503	7.9%
	800-999: “Injury And Poisoning”	4,694	6.7%
	710-739: “Diseases Of The Musculoskeletal System And Connective Tissue”	4,064	5.8%
	Other	15,867	22.7%
Diagnosis 2	390-459: “Diseases Of The Circulatory System”	22,075	31.5%
	250.0-250.9: “Diabetes Mellitus”	9,700	13.9%
	460-519: “Diseases Of The Respiratory System”	6,445	9.2%
	240-279 (not250): “Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders”	5,605	8.0%
	580-629: “Diseases Of The Genitourinary System”	5,042	7.2%
	Other	21,106	30.2%
Diagnosis 3	390-459: “Diseases Of The Circulatory System”	21,848	31.1%
	250.0-250.9: “Diabetes Mellitus”	12,546	17.9%
	240-279 (not250): “Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders”	6,403	9.2%
	460-519: “Diseases Of The Respiratory System”	4,245	6.1%
	580-629: “Diseases Of The Genitourinary System”	3,785	5.4%
	Other	21,146	30.2%
A1C Result	None	57,128	81.6%
	>7	9,104	13.0%
	Normal	3,741	5.3%

Metformin	No Steady Up Down	55,070 13,634 834 435	78.7% 19.5% 1.2% 0.6%
Glimepiride	No Steady Up Down	66,276 3,331 230 136	94.7% 4.8% 0.3% 0.2%
Glipizide	No Steady Up Down	60,966 8,063 573 371	87.1% 11.5% 0.8% 0.5%
Glyburide	No Steady Up Down	62,198 6,744 613 418	88.9% 9.6% 0.9% 0.6%
Pioglitazone	No Steady Up Down	64,710 5,004 178 81	92.3% 7.2% 0.3% 0.1%
Rosiglitazone	No Steady Up Down	65,312 4,455 132 74	93.3% 6.4% 0.2% 0.1%
Insulin	No Steady Up Down	34,258 21,617 6,777 7,321	49.0% 30.9% 9.6% 10.5%
Change in Medication	No Change Change	38,482 31,491	55.0% 45.0%
Diabetes Medication	Yes No	53,293 16,680	76.2% 23.8%
Readmitted	Not readmitted in <30 days Readmitted in <30 days	63,696 6,277	91.0% 9.0%

Attributes with large numbers of categories had their categories condensed prior to analysis. The condensed versions are what is shown in Table 3. Categories were grouped together based on domain knowledge where possible. If they did not fit together, any category with less than 5% of records was inserted into an "Other" category.

The number of categories were reduced for the following attributes for ease of analysis: Admission Type ID, Discharge Disposition ID, Admission Source ID, Diagnosis 1, 2, and 3, A1c Result, and Readmitted. The mapping for the original categories for Admission Type ID, Discharge Disposition ID, and Admission Source ID was available in a separate document provided by Strack et al. (2014) in their Supplementary Materials entitled "id_mapping.csv". Diagnosis attributes 1, 2, and 3 initially had 717, 749, and 790 distinct categories respectively, based on ICD-9-CM codes. Categories were initially grouped based on standardized ICD-9-CM groupings, and then any representing less than 5% of records were grouped into an "Other" category (ICD, 2020). For the target attribute, Readmitted, the categories for not readmitted and readmitted after 30 days were amalgamated into one. For A1c Result, the categories for <7 and <8 were amalgamated into one category to represent abnormally high values, as there is no domain-related reason to keep them separate for analysis.

All variables were compared against the target Readmitted variable to identify if there were any significant differences between the group of patients readmitted within 30 days and the group not readmitted within 30 days. One-sided sample t-tests were completed for each numeric variable, with the mean and standard deviation of each group identified [Table 4]. All findings were statistically significant except for Number of Procedures by Readmitted (p=0.4847). The mean was found to be higher in the readmitted within 30 days group than the not readmitted within 30 days group for every other numeric variable. A one-sided Wilcoxon rank sum test was completed for the one remaining ordinal variable to identify if there was a statistically significant difference between the center of each distribution [Table 5]. The findings were statistically significant that the center of the distribution for the readmitted within 30 days group was higher than the not readmitted within 30 days group (p<2.2 x 10-16). Chi-squared tests were completed for each categorical variable against Readmitted to identify if the variables were independent of one another [Table 6]. The findings for all categorical variables were statistically significant (i.e., each variable has an association with the Readmitted variable), except for Gender (p=0.5856), Admission Source ID (p=0.0555), Glimepiride (p=0.235), Glyburide (p=0.6068), Pioglitazone (p=0.3622), and Rosiglitazone (p=0.7032).

Table 4: Numeric Variables by Readmitted Group

Attribute	Readmitted <30 Group		Readmitted Not <30 Group		P-value for One-sided T-test
	Mean	Standard Deviation	Mean	Standard Deviation	
Time in Hospital	4.797	3.058	4.222	2.916	<2.2 x 10^{-16} *
Number of Lab Procedures	44.915	19.339	42.675	19.937	<2.2 x 10^{-16} *
Number of Procedures	1.425	1.730	1.426	1.760	0.4847
Number of Medications	16.625	8.324	15.571	8.278	<2.2 x 10^{-16} *
Number of Outpatient Visits	0.308	1.045	0.277	1.066	0.01096 *
Number of Emergency Visits	0.150	0.582	0.099	0.504	2.016 x 10^{-11} *
Number of Inpatient Visits	0.369	0.983	0.157	0.546	<2.2 x 10^{-16} *
Number of Diagnoses	7.513	1.847	7.195	2.014	<2.2 x 10^{-16} *

* Indicates statistical significance at the alpha = 0.05 level

Table 5: Ordinal Variable by Readmitted Group

Attribute	Readmitted <30 Group	Readmitted Not <30 Group	P-value for One-sided WIlcoxon Rank Sum Test
	Median	Median	
Age	[70-80)	[60-70)	<2.2 x 10^{-16} *

* Indicates statistical significance at the alpha = 0.05 level

Table 6: Categorical Variable by Readmitted Group

Attribute	P-value for Chi-squared Test
Race	0.0311 *
Gender	0.5856
Admission Type ID	0.008417 *
Discharge Disposition ID	<2.2 x 10^{-16} *
Admission Source ID	0.0555
Diagnosis 1	<2.2 x 10^{-16} *
Diagnosis 2	0.03454 *
Diagnosis 3	3.032 x 10^{-6} *
A1C Result	0.03305 *
Metformin	0.002862 *
Glimepiride	0.235
Glipizide	0.003262 *
Glyburide	0.6068
Pioglitazone	0.3622
Rosiglitazone	0.7032
Insulin	5.876 x 10^{-11} *
Change in Medication	0.0001085 *
Diabetes Medication	1.953 x 10^{-13} *

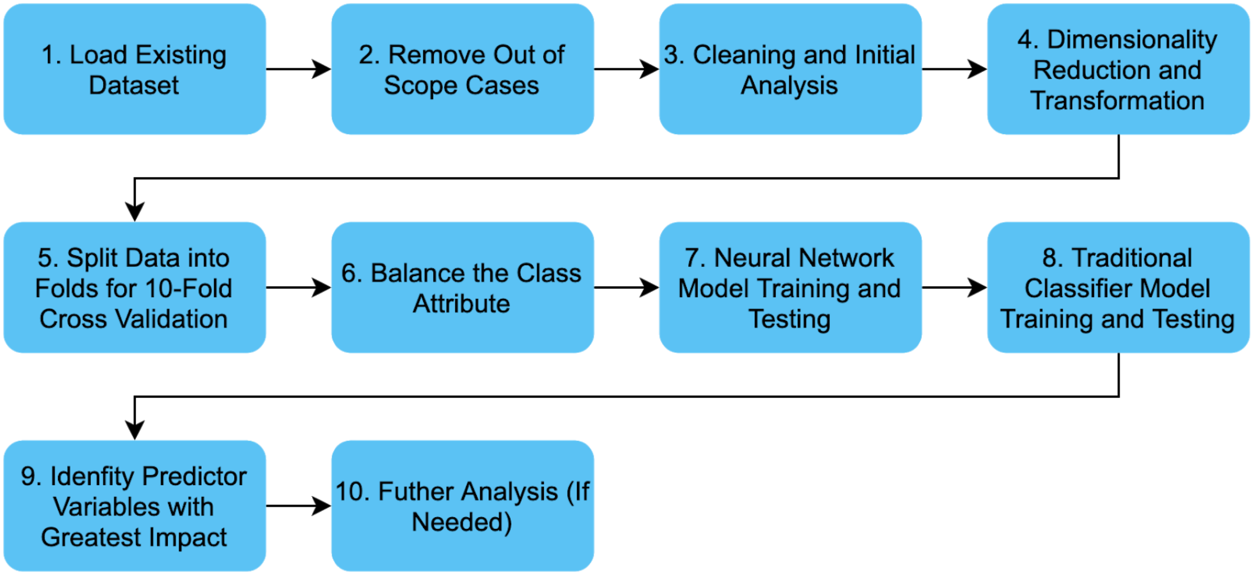
* Indicates statistical significance at the alpha = 0.05 level

For the categorical variables, percent contribution towards the chi-squared test statistic was calculated to identify which combinations of levels within each two variables had the biggest impact. Residuals were reviewed to identify if the association with high percent contribution was negative or positive. Overall, the biggest positive associations were identified between being readmitted within 30 days and being Caucasian, having an emergency admission, being discharged to a skilled nursing facility or other facility, having a primary diagnosis of a circulatory disorder or injury/poisoning, having a tertiary diagnosis of a respiratory disease, being on glipizide, having a dosage increase in insulin, having had a change in any diabetic medication, and being on any diabetic medication. This means that there were more people than expected readmitted within these groups. Negative associations were identified between being readmitted within 30 days and being African American, Hispanic, or "Other" race, having an elective admission, being discharged home, having a secondary diagnosis of diabetes mellitus, having an abnormal A1C result, taking metformin or having a metformin dosage increase, and not being on diabetes medication. This means that there were fewer people than expected in the readmitted group for each of these conditions.

After completion of exploratory data analysis, the final 28 variables were transformed for modeling. Numeric attributes were scaled using min-max scaling so that they may all be compared on the same scale between 0 and 1. The remaining ordinal variable (age) was integer encoded (e.g., [0-10) was assigned the number 1, [10-20) was assigned the number 2, etc.), and then scaled using min-max scaling as well. The four categorical variables with two categories each were converted to binary variables: Gender (Female: 0, Male: 1), Change in Medication (No change: 0, Change: 1), Diabetes Medication (No: 0, Yes: 1), and Readmitted (Not readmitted in <30 days: 0, Readmitted in <30 days: 1). All other categorical variables were one-hot encoded to create dummy variables with the number of variables equalling the number of categories in each original variable. The transformations prior to modeling resulted in a final dataset of 82 variables, with all values between 0 and 1.

Approach

Figure 5: Project Approach


1. Load Existing Dataset
2. Remove Out of Scope Cases
3. Cleaning and Initial Analysis
4. Dimensionality Reduction and Transformation
5. Split Data into Folds for 10-Fold Cross Validation
6. Balance the Class Attribute
7. Neural Network Model Training and Testing
8. Traditional Classifier Model Training and Testing
9. Idenfity Predictor Variables with Greatest Impact
10. Futher Analysis (If Needed)

Step 1: Load Existing Dataset

Load existing data from the Supplementary Materials of the Strack et al. 2014 study into RStudio.

Step 2: Remove Out of Scope Cases

The original dataset includes multiple encounters for the same patients. In order to focus on early identification of readmission risk factors, as well as avoid violating the statistical assumption of independence for analysis, encounters that were not a patient's first encounter were removed. Additionally, patients who died or were discharged to hospice are not eligible for readmission, and therefore these records were excluded as well.

Step 3: Cleaning and Initial Analysis

Convert attributes to correct data types, assign factor levels based on the id_mapping document that came with the Supplementary Materials from Strack et al. (2014). Complete univariate analysis on each attribute. Review and handle missing values and outliers as needed. Collapse target variable factor into two levels: readmitted within 30 days and not readmitted within 30 days. Collapse the categories for Discharge Disposition ID, Admission Source ID, the three Diagnosis attributes, et al. into meaningful groupings that may be processed better (e.g., Diagnosis 1 initially has 717 factor levels). Complete bivariate and multivariate analysis; check for correlations in numeric attributes.

Step 4: Dimensionality Reduction and Transformation

Remove attributes irrelevant for modeling: Encounter Number and Patient ID. Remove attributes with large number of missing values: Payer Code, Medical Specialty, Weight. Remove any attribute with extremely low variance (one category includes more than 95% of cases). One-hot encode remaining categorical variables (e.g., Diagnosis). Integer encode ordinal variables (e.g., Age). Scale any non-binary attributes with min-max scaling to improve model performance.

Step 5: Split Data into Folds for 10-Fold Cross Validation

Set a seed so that the findings may be reproducible. Split the data randomly into 10 folds stratified by class attribute (Readmitted). Each model was trained on each set of 9 folds and tested on the 10th in order to ensure consistency in results. The best performing algorithm was to be identified using the average of its 10 scores.

Step 6: Balance the Class Attribute

The class attribute (Readmitted) was initially imbalanced, with 9% of patients being readmitted with 91% not. In order to avoid model bias in favour of the majority class, the minority class was randomly oversampled in each fold to reach a balance between classes.

Step 7: Neural Network Model Training and Testing

Using the Keras Deep Learning library in R, variants of deep feedforward neural network algorithms (multilayer perceptrons) were built. Hyperparameters were tuned (including number of hidden layers, number of units per hidden layer, activation function, and optimization function) on each training set, and tested on each test set. Performances between variants were compared focusing on the accuracy, recall, precision, F1 score, and the c-statistic (area under the receiver operating characteristic curve [AUC]) per fold, with the scores averaged for comparison. The best performing variant was selected for comparison to other methods.

Step 8: Traditional Classifier Model Training and Testing

The following traditional classifiers were trained and tested on the same data preparation for comparison to the neural network performance: random forest, k-nearest neighbours (kNN), and logistic regression. The same measures as the neural networks (accuracy, recall, precision, F1 score, and AUC) were recorded and compared.

Step 9: Identify Predictor Variables with Greatest Impact

Using the best-performing model, each predictor variable was systematically removed to identify the effect on model performance. The removal of attributes resulting in the greatest reduction in performance were identified as the predictor variables with the greatest impact on readmission.

Step 10: Further Analysis (If Needed)

Review model performance, and findings of predictor variables with the greatest impact. Identify if any further analysis is warranted based on findings.

Results

The transformed dataset was used to model predictions for the Readmitted attribute using a deep feedforward neural network. Additional traditional classifiers were modeled for comparison, specifically k-nearest neighbours (kNN), random forest, and logistic regression. Evaluation metrics recorded for each model included accuracy, precision, recall, the F1 score (harmonic mean of precision and recall), and the area under the receiver operating characteristic (ROC) curve (AUC, or c-statistic).

All neural network algorithm variations tested included an input layer of 81 neurons, and an output layer of 1 neuron using the sigmoid activation function so that it may express the output between 0 and 1. Hyperparameters tuned included the number of fully-connected hidden layers, number of neurons per hidden layer, batch size, number of epochs, the addition of dropout layers with different rates of dropout, optimizer, weight regularization, and activation function for the hidden layers. The best performing configuration [Figure 6] included two dense hidden layers with 16 neurons each using the ReLU activation function, with a dropout layer (50% dropout) after the input layer and each of the hidden layers. It used the Adam optimizer with a batch size of 128 and it was found that 50 epochs were sufficient for the validation set accuracy to reach its peak. No weight regularization was used. See Figure 7 for a visualization of the training and validation set loss and accuracy during training (using folds 2 - 10 for training).

Figure 6: Structure of Optimized Neural Network

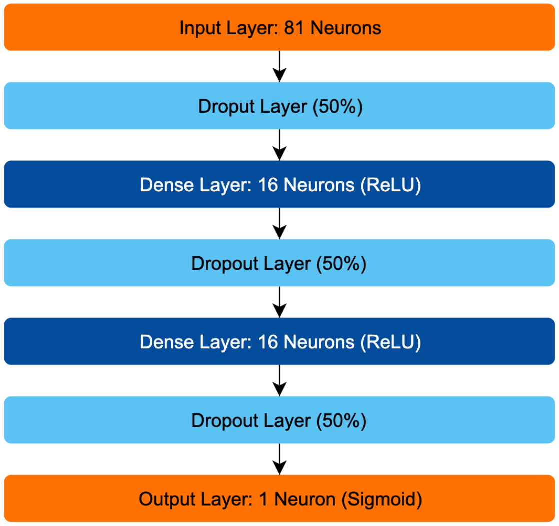
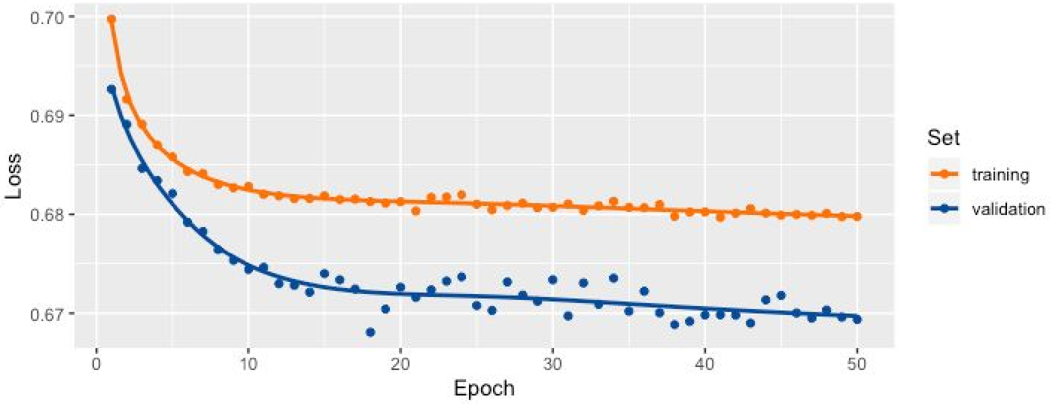
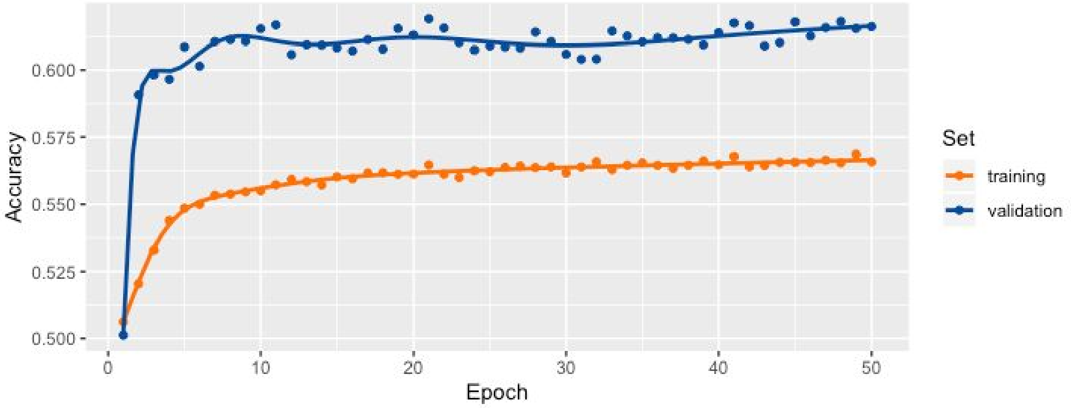

Input Layer: 81 Neurons
Droput Layer (50%)
Dense Layer: 16 Neurons (ReLU)
Dropout Layer (50%)
Dense Layer: 16 Neurons (ReLU)
Dropout Layer (50%)
Output Layer: 1 Neuron (Sigmoid)

Figure 7: Training and Validation Set Loss and Accuracy during Neural Network Training



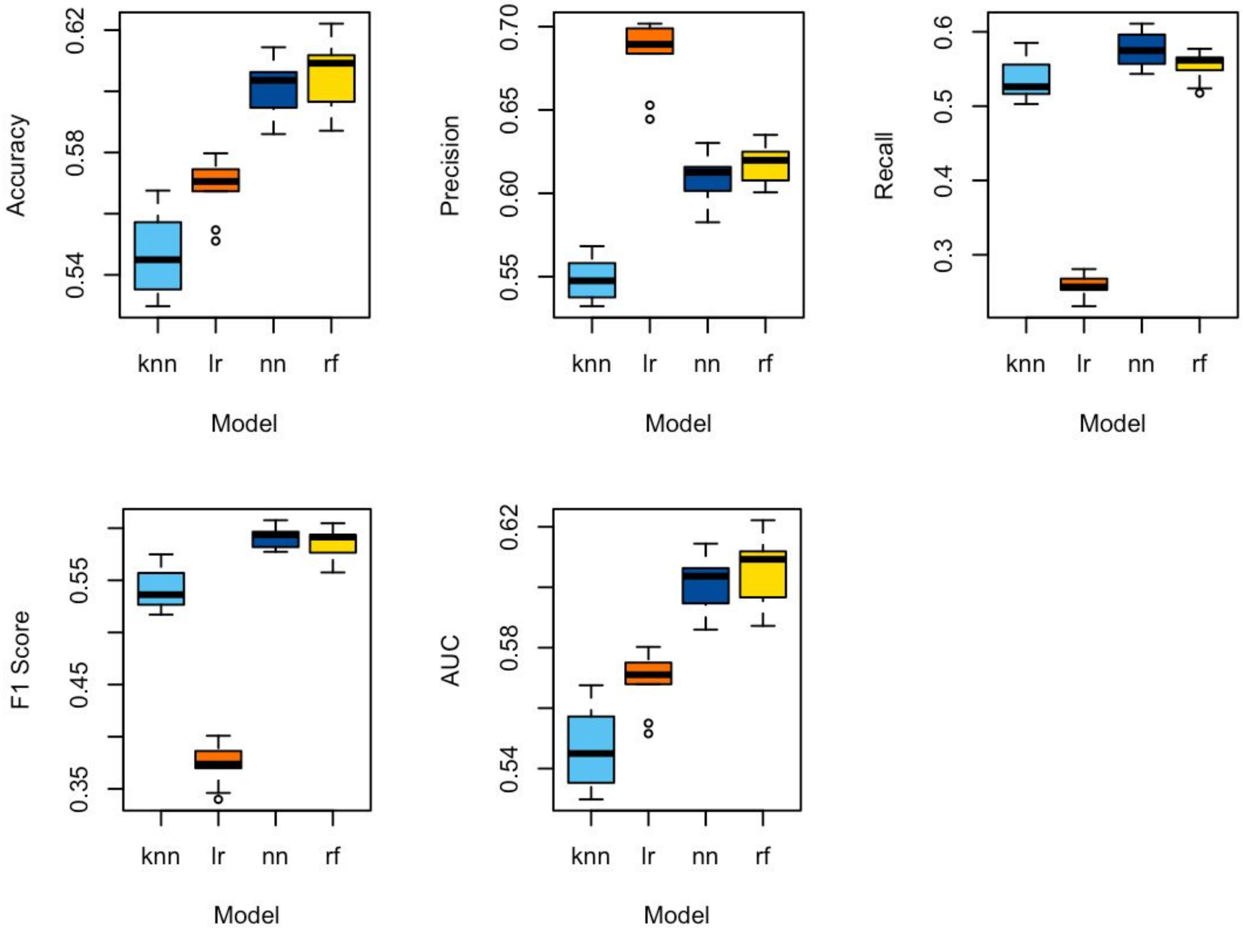


For the kNN algorithm, a *k* value of 85 was chosen after testing values in increments of 10 from five to 105 to identify the best performer. For the random forest model, optimized parameters included 100 trees with six variables per tree, a node size of 500 (i.e., there had to be at least 500 data points for a node to be created), and a maximum of 17000 terminal nodes. The best performing logistic regression model included all variables, except for one dummy variable from each original categorical variable (to eliminate the multicollinearity). The means and standard deviations of evaluation metrics across folds by model are available in Table 7. Figure 8 displays the boxplots for evaluation metrics by model. In Figure 8, the model "knn" refers to k-nearest neighbours; "lr" is logistic regression; "nn" is neural network, and "rf" is random forest.

Table 7: Mean and Standard Deviation of Evaluation Metric Scores across Folds by Model

Model	Accuracy (%)		Precision (%)		Recall (%)		F1 Score (%)		AUC (%)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
kNN	54.74	1.35	54.92	1.26	53.52	2.61	54.20	1.91	54.75	1.35
Logistic Regression	56.89	0.93	68.45	2.01	25.76	1.60	37.42	1.95	56.94	0.94
Neural Network	60.36	0.93	60.86	1.41	57.63	2.27	59.16	0.93	60.18	0.94
Random Forest	60.53	1.13	61.81	1.11	55.38	1.95	58.41	1.55	60.54	1.13

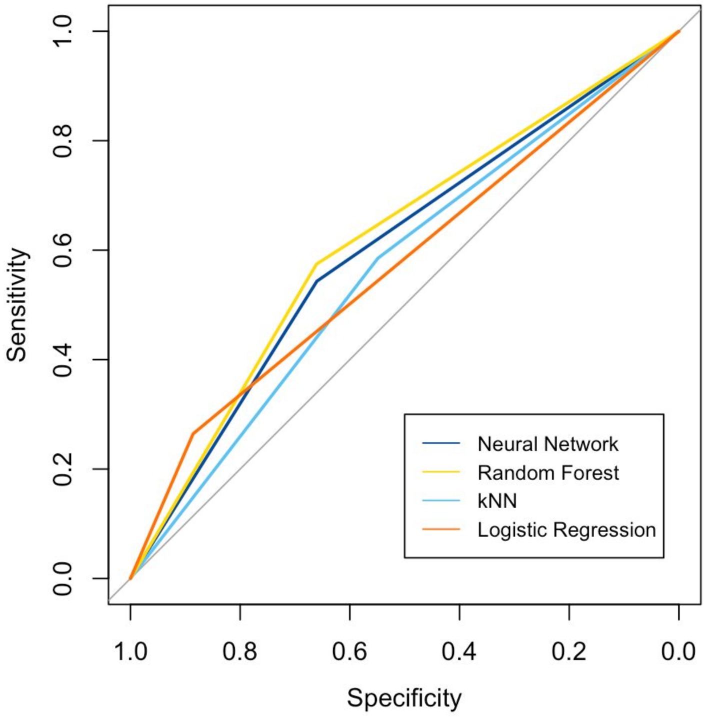
Figure 8: Evaluation Metric Boxplots by Model



The neural network and random forest algorithms appeared to perform similarly in all evaluation metrics recorded, and had the highest mean values of the four models, except in precision. Logistic regression had the highest mean precision (0.6845), although its recall was much lower than all other models (0.2576). The evaluation metrics recorded for kNN were all consistently lower, between 0.53 and 0.55.

Comparison of model performance may also be visualized in the receiver operating characteristic (ROC) curves [Figure 9]. The ROC curves plot the specificity (true negative rate) against the sensitivity (true positive rate/recall) for each model. The further each model's line is away from the diagonal grey (specificity = sensitivity) line, towards the top left corner, the better it is performing. The random forest model appears to be performing the best here, with the neural network following closely, which is confirmed by the AUC values [Table 7]..

Figure 9: Receiver Operating Characteristic (ROC) Curves for Each Model



Recorded evaluation metric percent values of the optimized models all peaked around 60%, suggesting that the information provided by the predictor variables was not sufficient to accurately predict whether a patient would be readmitted. This finding is consistent with other studies reviewed in the literature review, which found AUC values of 55% - 83% in studies predicting all-cause readmission (Kansagara et al., 2011). Access to additional features shedding more light on demographics (as they relate to the social determinants of health), such as insurance type and proximity of address to hospital, as well as in hospital features such as whether the patient stayed in intensive care (ICU) may support a more accurate prediction of early readmission. However, it may not even be possible to predict patient readmission with complete accuracy, as it is a complex problem affected by health, social, psychological, and environmental factors.

Each of the five evaluation metrics (accuracy, precision, recall, F1 score, AUC) across folds for each model were compared using statistical techniques. First, the Shapiro-Wilk test was completed for each distribution (per evaluation metric per model) to identify normality. All distributions were identified as normal at an alpha of 0.05 level (all p > 0.05), except for precision in the logistic regression model (p = 0.01). Therefore, parametric tests were used for all comparisons of accuracy, recall, F1 score, and AUC, whereas non-parametric tests were used for comparison of precision between models. Then, a two-way block design analysis of variance with the folds as blocks (Friedman test for precision) was used to identify if any of the means (medians for precision) of evaluation metrics between models differed. Findings were highly significant for each of the five evaluation metrics that at least one mean or median differed (accuracy $p = 2.4 \times 10^{-14}$, precision $p = 3.5 \times 10^{-6}$, recall $p < 2 \times 10^{-16}$, F1 score $p < 2 \times 10^{-16}$, AUC $p = 2.6 \times 10^{-14}$).

Subsequently, pairwise two-sided paired t-tests (Wilcoxon signed-rank test for precision) were completed for each type of evaluation metric to identify which models differed from one another in mean or median. For all five evaluation metrics, the means or medians did not significantly differ from one another between the neural network and the random forest (accuracy p = 1.0, precision p = 0.22, recall p = 0.43, F1 score p = 1.0, AUC p = 1.0). Additionally there was no statistically significant difference between the means in recall for random forest and kNN (p=0.090). Otherwise, all other pairings had a statistically significant difference from one another.

Given that the neural network and random forest models performed equally to one another for each evaluation metric recorded, the variables of importance were identified from each of these models for analysis. For the optimized neural network model, the model was trained and evaluated with each variable systematically excluded, and the reduction in AUC from baseline was identified. The variable with the highest reduction in AUC with removal is the variable of greatest importance identified by the model. AUC was selected, as this measure was commonly used for comparison in the articles discussed in the literature review. See Table 8 for the top 10 variables of importance identified from the neural network model.

Table 8: Top 10 Variables of Importance Used in Neural Network

Rank	Variable	Reduction in AUC
1	Number of Inpatient Visits*	- 0.01229
2	Discharge Disposition ID: Discharge to Other Facility*	- 0.00805
3	Race: Asian	- 0.00728
4	Pioglitazone: Steady	- 0.00725
5	Discharge Disposition ID: Discharge home with Home Health Services	- 0.00698

6	Time in Hospital*	- 0.00564
7	Insulin: Steady	- 0.00510
8	Diagnosis 3: Other	- 0.00401
9	Glipizide: Steady	- 0.00359
10	Glyburide: No	- 0.00356

* Indicates variable is also in the top 10 for the random forest model

Variable importance in the random forest model was evaluated using mean decrease in accuracy, which is how much the mean accuracy decreases with the exclusion of each variable. Mean decrease in Gini, another common measure of variable importance in random forest models, was not used as it is generally biased in favour of continuous variables. This was evident in results, as all eight continuous variables were in the top 12 for highest mean decrease in Gini. See Table 9 for the top 10 variables of importance in the random forest model identified using mean decrease in percent accuracy.

Table 9: Top 10 Variables of Importance Used in Random Forest

Rank	Variable	Mean Decrease in Percent Accuracy		
		Not Readmitted	Readmitted	Total
1	Number of Inpatient Visits*	24.190	23.842	26.690
2	Age	10.987	15.827	19.501
3	Number of Medications	2.193	17.625	19.464
4	Number of Lab Procedures	7.713	19.833	19.268
5	Time in Hospital*	10.366	13.786	17.675
6	Discharge Disposition ID: Discharge Home	14.551	9.453	17.097
7	Number of Diagnoses	5.584	16.069	16.517
8	Discharge Disposition ID: Discharge to Other Facility*	7.783	15.510	16.498
9	Number of Procedures	5.993	15.058	15.331

10	Number of Emergency Visits	8.804	12.652	13.390

* Indicates variable is also in the top 10 for the neural network model

There are three variables in common within the top 10 variables of importance between the two models, further confirming their importance: Number of Inpatient Visits, Time in Hospital, and Discharge Disposition ID: Discharge to Other Facility. The Number of Inpatient Visits was ranked as number one in variable importance for both the neural network model and random forest model, further underscoring its importance in determining whether a patient will be readmitted within 30 days or not. While these may contribute significantly to model performance, they are not modifiable risk factors in practice.

Even though the mean decrease in accuracy was used instead of mean decrease in Gini, there still appears to be a bias in favour of continuous variables for the random forest model. Seven of the eight continuous variables are still in the top 10 variables of importance, in addition to Age, which was collected as an ordinal variable that was later integer encoded and scaled. It may not be the information itself, but rather how many options of values are present for each variable (binary [2] vs continuous [many]) that resulted in the identification of continuous variables as those of greatest importance. Regardless, most are risk factors that aren't able to be altered during hospital stay, except for the Number of Medications. A potential target for intervention may be completing a full pharmacist review of medications and discontinuation of medically unnecessary medications in order to potentially reduce the chances of readmission and improve outcomes. Polypharmacy (the use of more medications than medically necessary) is a common health issue in older adults, which can lead to problems such as non-adherence, drug interactions, cognitive impairment, falls, and functional decline (Maher et al., 2013). This can be at least partially addressed in hospital.

The neural network also identifies being on pioglitazone, insulin, glipizide, and not being on glyburide as variables of importance ranking in the top 10. Insulin, glipizide, and glyburide are considered higher risk medications: insulin requires close daily monitoring of blood sugar levels, and nonadherence can lead to hypo- or hyperglycemia; glipizide and glyburide are in the sulfonylurea class of medications, which are associated with increased risk of cardiovascular death as well as episodes of severe hypoglycemia (Douros et al., 2018). These risk factors are modifiable, and can be potentially mitigated through changes in medication or dosage, patient education, and outpatient follow-up.

Conclusion

Of the four models created, the neural network and random forest models performed the best over all five evaluation metrics measured: accuracy, precision, recall, F1 score, and AUC. There was no statistically significant difference between the performance of these two models across any of these evaluation metrics. The highest mean percent accuracy (60.53% for random forest) and AUC (60.54% for random forest) were just above 60%, indicating that the information provided by the predictor variables used was not sufficient in accurately predicting whether a patient with diabetes would be readmitted to hospital within 30 days or not. The inclusion of additional features may improve model performance, such as further demographic information and details of care during hospital stay. Analysis of variable importance in the neural network and random forest models both identified the number of inpatient visits in the previous year, time in hospital, and being discharged to a facility other than a skilled nursing facility as key variables impacting model performance. Modifiable variables also identified as key predictors include use of pioglitazone, insulin, and glipizide, and lack of use of glyburide by the neural network model and number of medications by the random forest model.

The features selected were limited to those already extracted by Strack et al. for their 2014 study and were affected by their initial preprocessing choices resulting in some information loss (e.g., converting Age to an ordinal variable with groups of 10). The data came from Cerner's Health Facts database of US clients, and as such is limited to organizations partnering with Cerner in the US who chose to participate. Additionally, the data used excluded out-of-network providers, and therefore may fail to capture the full scope of readmissions. Future studies may consider using additional features related to demographics, such as payer code and proximity to hospital, as well as factors present in hospital, such as ICU stay. Additional focused study may be warranted to implement interventions related to diabetic medications, particularly insulin and sulfonylureas, and identify the effects on readmission.

References

Centers for Disease Control and Prevention [CDC]. (2020, February 11). *National diabetes statistics report, 2020*. Retrieved from https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

Douros, A., Dell'Aniello, S., Yu, O. H. Y., Filion, K. B., Azoulay, L., & Suissa, S. (2018). Sulfonylureas as second line drugs in type 2 diabetes and the risk of cardiovascular and hypoglycaemic events: Population based cohort study. *British Medical Journal, 2018*(362). doi: https://doi.org/10.1136/bmj.k2693

Dungan, K. M. (2012). The effect of diabetes on hospital readmissions. *Journal of Diabetes Science and Technology, 6*(5). Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/193229681200600508

International Classification of Diseases [ICD]. (2020). *ICD-9-CM Chapters*. Retrieved from https://icd.codes/icd9cm

Jamei, M., Nisnevich, A., Wetchler, E., Sudat, S., & Liu, E. (2017). Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *Public Library of Science One*, *12*(7):e0181173. Doi: 10.1371/journal.pone.0181173

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. *Journal of the American Medical Association, 306*(15):1688-1698. doi: 10.1001/jama.2011.1515

Karunakaran, A., Zhao, H., & Rubin, D. J. (2019). Pre- and post-discharge risk factors for hospital readmission among patients with diabetes. *Med Care, 56*(7): 634-642. doi: 10.1097/MLR.0000000000000931

Kim, H., Ross, J. S., Melkus, G. D., Zhao, Z., & Boockvar, K. (2011). Scheduled and unscheduled hospital readmissions among diabetes patients. *American Journal of Managed Care, 16*(10): 760–767. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024140/

Liu, W., Stansbury, C., Singh, K., Ryan, A. M., Sukul, D., Mahmoudi, E., Waljee, A., Zhu, J., & Nallamothu, B. K. (2020). Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding. *Public Library of Science One*, *15*(4):e0221606. doi: 10.1371/journal.pone.0221606

Maher, R. L., Hanlon, J. T., & Hajjar, E. R. (2013). Clinical Consequences of Polypharmacy in Elderly. *Expert Opinion on Drug Safety, 13*(1). doi: 10.1517/14740338.2013.827660

Ostling, S., Wyckoff, J., Ciarkowski, S. L., Pai, C., Choe, H. M., Bahl, V., & Gianchandani, R. (2017). The relationship between diabetes mellitus and 30-day readmission rates. *Clinical Diabetes and Endocrinology, 3*(3). doi:10.1186/s40842-016-0040-x

Rubin, D. J., Golden, S. H., McDonnell, M. E., & Zhao, H. (2017). Predicting readmission risk of patients with diabetes hospitalized for cardiovascular disease: A retrospective cohort study. *Journal of Diabetes and its Complications, 31*(8): 1332-1339. doi: 10.1016/j.jdiacomp.2017.04.02131(8):1332-1339

Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., & Clore, J. N. (2014). Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed Research International*, *2014*(781670). doi: https://doi.org/10.1155/2014/781670

UCI Machine Learning Repository. (2014, May 3). *Diabetes 130-US hospitals for years 1999-2008 data set*. Retrieved from https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008

Wadhera, R. K., Maddox, K. E. J., Kazi, D. S., Shen, C., & Yeh, R. W. (2019). Hospital revisits within 30 days after discharge for medical conditions targeted by the Hospital Readmissions Reduction Program in the United States: National retrospective analysis. *British Medical Journal*, *366*(l4563). Doi: https://doi.org/10.1136/bmj.l4563