PREDICTING RENAL REPLACEMENT THERAPY IN CHRONIC KIDNEY DISEASE PATIENTS

Viktor Basharkevich, David Chang, Sabrina Chunn, Nishi Gupta,
Jasmin Langomas, Chaeyoung Lee, Xander Marshall, Ava Ostrem,
Matthew Quispe, Skylar Walters, Charles Yu, Kaitlyn Zelvin

Advisor: Dr. Minjoon Kouh
Assistant: David Van Dongen

ABSTRACT

Chronic kidney disease (CKD) is a potentially fatal disease that affects a large percentage of the world population. Once CKD has progressed to the end stage, patients must receive treatment to survive, which often comes in the form of renal replacement therapy (RRT). However, predicting when to start RRT can be challenging. A recent study explored the possibility of only using patients' comorbidities to train ML algorithms to predict whether a patient would need to start RRT within a given period. The purpose of our project was to replicate this study by applying ML techniques to data about the comorbidities of patients with CKD to predict whether they will need RRT within the next 6 months. We used data from Taiwan's National Health Insurance Database (NHIRD) to train eight machine learning algorithms. These models were then evaluated by their accuracy, sensitivity, specificity, and area under the receiver operating curve (AUC). Four out of the eight algorithms that we tested had AUC greater than 0.70, with XGBoost being most effective for this dataset. Improvements in computing power, access to more diverse data, and further tailored algorithms will allow ML to be used in hospitals in the future to predict the time of RRT.

INTRODUCTION

Chronic Kidney Disease

The kidney is an organ that rids the bloodstream of toxins and transforms the body's liquid waste into urine. One disease that affects the kidneys is chronic kidney disease (CKD). CKD causes a gradual loss of kidney function and a backup of waste in the kidneys. If left untreated, it can be fatal. The disease often starts with pre-existing damage to the kidneys, which can spawn from a multitude of factors such as smoking, excessive drinking, and diabetes. To compensate for this damage, the body enters a state of "renal hyperfiltration" where the kidneys' filtration cells (glomerular) produce an excessive amount of pro-urine, leading to a drop in the body's levels of creatinine, a waste product of the muscles. As a result of this drop, the kidneys' lose their ability to filter waste and become inflamed (1).

CKD is classified into five stages based on the symptoms leading toward kidney failure; each stage becomes increasingly fatal (Table I) (2). The stages are heavily dictated by a patient's estimated glomerular filtration rate (eGFR), which is an estimate of the kidney's ability to filter waste taking into account factors such as age, sex, race, and creatinine levels. The average eGFR is greater than or equal to 60; any eGFR level below 60 puts a patient at an increased risk of CKD (3).

Table I: Stages of CKD

Stage	Description	eGFR levels	Symptoms
1	Kidneys work efficiently. Little to no damage.	>90	No health altercations
2	Kidneys work efficiently. Minor kidney damage.	60-89	No health altercations
3	Health complications begin (moderate kidney damage). Waste starts building up as kidneys work less efficiently.	30-59 Early stage 3: 45-59 Late stage 3: 44-30	Early stage 3: Back pain, change in urination frequency. Late stage 3: High blood pressure, anemia
4	Severe kidney damage. 2-5 weeks from kidney failure.	15-29	More severe stage 3 symptoms
5	Fatal kidney damage. Kidney failure is inevitable as waste enters the bloodstream. Life expectancy drops to days if not treated.	<15	Chaotic urination cycle, trouble breathing, sharp cramps, sharp back pain

When kidneys lose their ability to filter dangerous fluids, electrolytes, and wastes, the body enters end-stage renal disease (ESRD). Once diagnosed with ESRD, patients must undergo supportive care, kidney transplantation, or renal replacement therapy (RRT) (4). The optimal solution is a kidney transplant, since it restores the body's urinary system. If not available, patients can opt out for supportive care, in which they take medicine to limit the CKD's symptoms on the body. However, supportive care can only temporarily lengthen the lifespan and proves to be the most ineffective method of the three. The final treatment option for CKD is RRT, commonly known as dialysis, which replaces the work of the kidney by filtering a patient's blood with a dialyzer, a machine that pushes the blood through a series of hollowed, thin tubes. The blood is directed out of and into the patient via needles and tubes (5).

ICD-9 Codes

The International Classification of Diseases (ICD) is a universal coding system that hospitals use to store disease statistics for patients. Its purposes include locating disease outbreaks and cataloguing causes of deaths or illnesses, along with noting which patients have a certain antimicrobial resistance when prescribed antibiotics (6). ICD utilizes uniform values, making patient data easily accessible in the event of domestic and international hospital transfers, and can be used to monitor where certain adverse events are happening. The ICD database is vital to hospitals because it makes mortality statistics more useful by consolidating related conditions, selecting a cause of death from a recorded chronological sequence of conditions, and highlighting certain categories over others to establish importance.

The ICD-9 code for general CKD is 585, with 585.1 - 585.6 referencing specific stages of CKD. This allows doctors to easily identify the severity of the disease and make decisions such as when to start RRT (7). Other ICD-9 codes that are related to CKD are shown in Table II.

Table II: ICD-9 codes related to CKD

ICD-9 Code	Description
V420	Kidney transplant
258.9	Anemia
564.0	Constipation
272.4	Hyperlipidemia (elevated levels of fat in the blood)

Often, several ICD-9 codes will be noted so doctors can have a better idea of the patient's symptoms. For example, in a subset of Taiwan's National Health Insurance Research Database (NHIRD) containing data for 2001 patients, there were 461 patients with code 585; 50 of whom also had codes for unspecified anemia and 34 with codes for unspecified hyperlipidemia. A majority of the patients had 3 different ICD-9 codes filed to indicate comorbidities.

Benefits of Machine Learning

Having access to thousands of patients' comorbidities and ICD-9 codes allow for the implementation of computer science algorithms to predict the proper start time of RRT. Machine learning (ML) is a subset of artificial intelligence that is used to emulate the process of human learning. ML is frequently applied to fields outside of computer science to solve problems with a high degree of efficiency and accuracy. Creating a ML algorithm involves collecting, preprocessing, and manipulating data, ultimately allowing us to predict outcomes based on the set of training data. Many ML applications use the Python programming language because of its

simplicity and versatility. Python uses libraries such as Numpy, which creates multi-dimensional arrays to manipulate data, and Matplotlib, which creates graphs. In addition, the Pandas library allows for data manipulation and analysis, and the Sci-kit Learn library contains classification, regression, and clustering algorithms used for ML.

ML and data science are becoming increasingly critical in the medical research and healthcare fields. ML has been used to streamline recordkeeping practices, repurpose drugs towards COVID-19, predict disease outcomes, and more (8). For example, a team from the University of Massachusetts Medical School employed ML to identify how and why HIV-1 grows resistant to darunavir, a common antiviral used in the treatment (9). Additionally, algorithmic learning has been used to predict the progression and diagnosis of Parkinson's Disease, with the most effective algorithm having an accuracy rate of over 90% (10). Machine learning has also provided scientists with a greater understanding of cancer and its underlying mechanisms; one study employed cryo-EM, which incorporates ML in the data processing procedure, to better understand the hedgehog proteins that drive many human cancers (11). With such advances in how we approach healthcare, ML will undoubtedly have a profound impact on patient care and medicine efficiency.

Furthermore, the continuous influx of CKD patients creates time limitations to starting RRT, especially for those who might also suffer from comorbidities. Additionally, treatment is expensive and CKD patients may lack understanding of the illness they are suffering from. As these issues prevail, unnecessary fatalities ensue, which could be avoided if CKD is caught early and treated accordingly. Having a method to properly determine when a CKD patient should undergo RRT would allow patients and doctors to properly prepare for treatment. By training ML software on CKD ICD-9 codes to become accustomed to ESRD diagnosis, ML may be able to predict when to provide a patient with RRT, saving both lives and healthcare funding.

Prior Study Replication

This project is a replication of a recent study which predicted the onset of RRT 3, 6, and 12 months from the time that the patient was first diagnosed with CKD using 10 ML algorithms (12). The main metric that the study used to measure performance was Area Under the Radio Operator Characteristic Curve (AUC). The study found the Logistic Regression ML algorithm to yield the highest accuracy rate, followed by the XGBoost algorithm and the SGD algorithm. All of these algorithms obtained AUC > 0.7, indicating significant performance in predicting patient outcomes.

Data sharing greatly benefits scientific discovery because it allows for more efficient collaboration and increases confidence in findings. Study replication is also important because finding results that are similar to the original study can give greater validity to the findings (13).

Replication studies confirm if future researchers should build upon previous findings or if there are mistakes that need to be caught before the information is widely distributed and built upon. For example, in 2011 a group of researchers at the OPERA particle detector observed a neutrino that moved faster than the speed of light. It was only when researchers re-checked their work that they realized that the fluke was because of a loose fiber optic cable (14).

Our replication of the study used the same data from the NHIRD and the same metrics to measure performance as used by the original study. In the study we are replicating, the authors mainly focused on a 12-month prediction period. Therefore, we focused on a 6-month prediction period after the initial diagnosis to expand upon the existing research. In addition, we selected 8 algorithms to test, including the most efficient algorithms from the original study and a few new algorithms to see if they performed better.

METHODS

Data Preprocessing

Original Dataset

One country significantly burdened by the prevalence of CKD is Taiwan; the condition has been reported to affect as high as 15.5% of the country's population according to one study (15). Taiwan has the highest concentration of ESRD in Asia due to environmental and genetic factors (15). The US has similarly high CKD prevalence, at an estimated 14.2% (16). The dataset used in this study comes from Taiwan's National Health Insurance Research Database (NHIRD). It is one of the most comprehensive databases available in the country, containing data from over 99% of the population—around 23 million people. Each patient in the database is well-documented with a medical history consisting of diagnoses upon each doctor visit using the ICD-9 code system. In the US, there is no such analog, as the health insurance databases are often limited to individual states or insurance companies (17). In this study, a subset of the NHIRD dataset from 1997 to 2011 was analyzed.

Selecting Data

The data from the NHIRD database was preprocessed following similar filtering criteria to the original study (Fig. 1). The data was first filtered for only patients with a CKD diagnosis, which corresponds to an ICD-9 code of '585'. Then, patients who had RRT before their CKD diagnosis were eliminated, because their data could not be used to predict the time of RRT after an initial CKD diagnosis. Finally, the data was filtered for only patients with six months of health data after their CKD diagnosis. This was done so the ML model could be used to predict whether a patient would need RRT within six months of their CKD diagnosis.

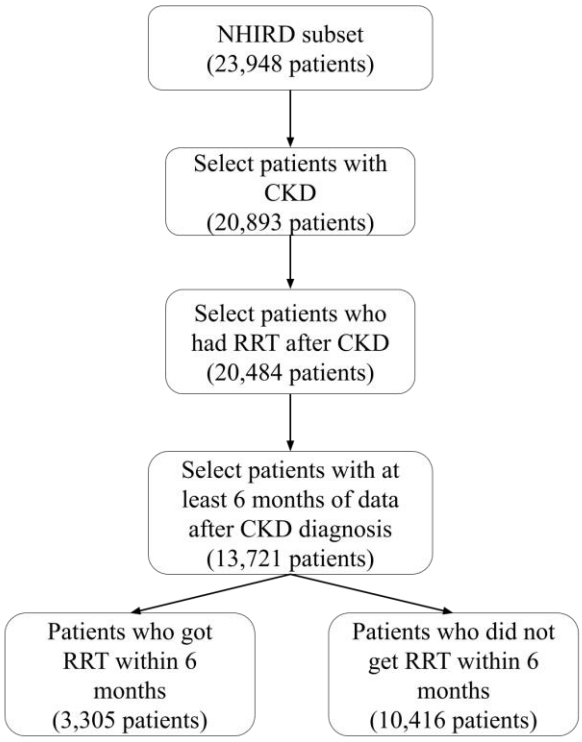


Fig 1: Overview of data selection process. All patient data from the original dataset was filtered to contain only data relevant to the ML algorithms.

The time of RRT was determined by the first appearance in a patient's medical records of either the ICD-9 code V420, which indicates that the patient received a kidney transplant, or an index key corresponding to one of several other RRT procedures, including hemodialysis and peritoneal dialysis. These dialysis drug codes were 58001C, 58019C, 58020C, 58021C, 58022C, 58023C, 58024C, 58025C, 58027C, 58029C, 58030B, 58002C, 58009B, 58010B, 58011C, 58012B, 58017C, and 58028C.

The diagnoses that each patient had received prior to their CKD diagnosis were added to the input file. Only pre-CKD diagnoses are used so the ML model could be used to predict how long after the initial CKD diagnosis a patient should start RRT. The ultimate goal of this study is for doctors to be able to use ML algorithms to start planning treatment for CKD patients at the time of their CKD diagnosis. Accurate predictions would allow hospitals to allocate resources more efficiently and provide better care to CKD patients.

Input and output datasets

The preprocessing code produced two separate files: an input file and an output file. The input file (Table III) is a matrix of size 13721 x 8038. There are 13721 rows for each patient, and

8038 columns for each ICD9 code. The columns contain a 1 if the patient received this ICD9 code before CKD diagnosis and 0 if the patient did not. The output file (Table IV) is a vector of size 13721 x 1. Each row represents a patient and each patient has a binary number indicating whether they received RRT treatment within six months of CKD diagnosis. Overall, our preprocessed data contains data for 13,721 patients. 3,305 of those patients received RRT within six months of their CKD diagnosis and 10,416 did not.

Table III: Sample input matrix.

Diagnosis	**#1**	**#2**	**#3**	**...**
Patient #1	0	0	1	...
Patient #2	1	0	0	...
Patient #3	0	0	0	...
...	...	...	...	...

Table IV: Sample output matrix.

	RRT status
Patient #1	0
Patient #2	0
Patient #3	1
...	...

Preliminary Analysis of NHIRD Data

Before applying ML algorithms, we conducted statistical analysis of the input and output datasets. First, we plotted the distributions of the number of ICD-9 codes of two different groups: all CKD patients and CKD patients who received RRT within six months (Fig. 2).

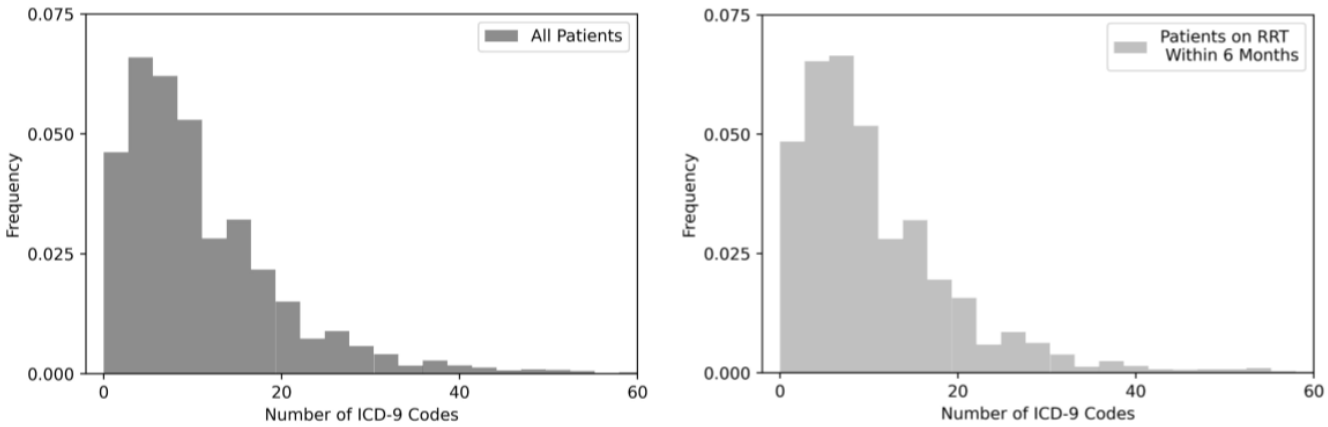


Fig. 2: Distributions of the number of ICD-9 codes. The first histogram shows the distribution for all CKD patients and the second histogram shows the distribution for CKD patients that received RRT within six months.

Additionally, we determined the rate of certain ICD-9 diagnoses in patients receiving RRT in comparison to those without RRT (Fig. 3). Notably, codes 4019 (hypertension) and 4659 (acute upper respiratory infection) have the highest frequencies in RRT patients at 42.6% and 32.3% respectively, while codes such as 25000 (diabetes), 5829 (chronic glomerulonephritis), and 586 (renal failure) are significantly more frequent in RRT patients relative to those who did not have RRT. Table V shows the most frequent codes and their corresponding medical names. This information could be valuable for identifying specific diagnoses that contribute toward the RRT outcome and limit the number of features being processed by the ML model.

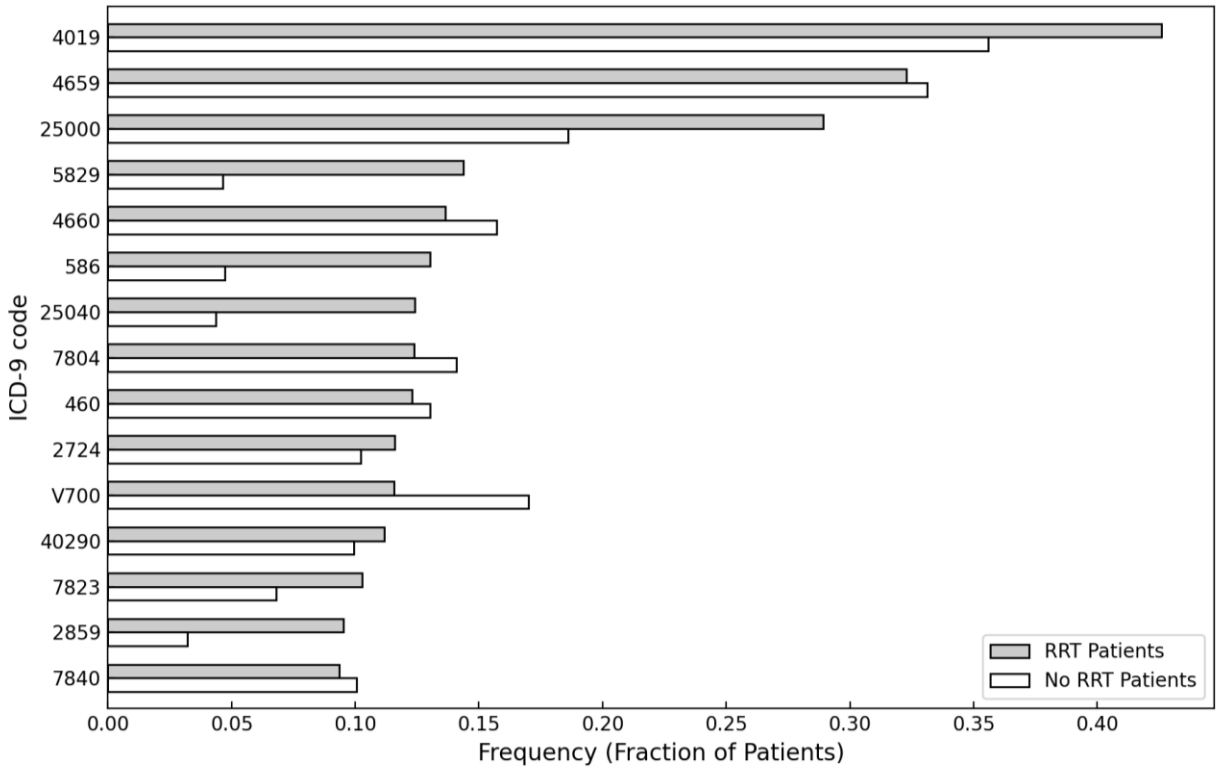


Fig. 3: Frequency of ICD-9 Diagnoses in Patients. Consists of the top 15 ICD-9 diagnoses with the highest frequencies in patients with RRT within 6 months after a CKD diagnosis**.** For those ICD-9 codes, a comparison is drawn to CKD patients without RRT within the same timeframe. Frequency for a particular ICD-9 code in a certain category of patients is measured from the number of patients having that diagnosis code divided by the number of patients. Note that the frequencies were taken from data prior to a patient's CKD diagnosis.

Table V: ICD-9 Codes from Figure 3

ICD-9 Code	Description
4019	Unspecified essential hypertension
4659	Acute upper respiratory infections of unspecified site
25000	Diabetes mellitus without mention of complication
5829	Chronic glomerulonephritis with unspecified pathological lesion in kidney
4660	Acute bronchitis
586	Renal Failure
25040	Type 2 diabetes mellitus with other diabetic kidney complication
7804	Dizziness and giddiness
460	Acute nasopharyngitis
2724	Unspecified hyperlipidemia
V700	Routine general medical examination at a health care facility
40290	Unspecified hypertensive heart disease without heart failure
7823	Edema
2859	Anemia, unspecified
7840	Headache

We also computed the Pearson Correlation Coefficient (PCC) for each ICD-9 code to determine which ICD-9 codes have the strongest positive correlation with the RRT outcome (Table VI). Our results show that 5829 has the strongest positive correlation with a PCC of 0.1631, with 36202 coming in second with a PCC of 0.1559.

Table VI: Five ICD-9 codes that have the highest positive Pearson correlation coefficient with RRT outcome.

ICD-9 Code	**Pearson Correlation Coefficient (PCC)**
5829 (Chronic glomerulonephritis with unspecified pathological lesion in kidney)	0.1631
36202 (Proliferative diabetic retinopathy)	0.1559
25040 (Type 2 diabetes mellitus with other diabetic kidney complication)	0.1414
586 (Renal Failure)	0.1413
2859 (Anemia, unspecified)	0.1271

Splitting Testing and Training Data

Designing an ML algorithm requires the input of data. Feeding this data into the algorithm refines the program, allowing it to recognize certain elements as more data is provided; this is known as the training data, which is instrumental to creating an accurate algorithm. To estimate the performance of our ML model, we also must input test data into the algorithm. Should the model for the test data predict the results with a high degree of precision, then the test data would confirm the validity and accuracy of the algorithm. This is similar to when a teacher gives practice problems to a class (train data), and then gives a graded test (test data).

When training the algorithms we randomly split our dataset into training and testing data to perform cross validation. This is particularly crucial because cross-validation is a strong deterrent to overfitting. Overfitting occurs when an algorithm is tuned so closely to the training data that it cannot accurately predict other data, making the algorithm useless. Thus, avoiding this is crucial for a high performing algorithm. Furthermore, splitting the data also forces the algorithms to learn patterns within the data instead of just memorizing outcomes.

To split the data, we used *k*-fold, a cross validation technique where the full data is randomly split into *k* equally sized pieces (Fig. 4). After this, an algorithm is trained *k* times, with 1 piece of data being reserved as the testing data each time. For instance, with $k = 5$, there will be 5 chunks of data, and the algorithm in question will be trained 5 times, with a different chunk of data being reserved as testing data each time. In our testing, we used $k = 10$ because this led to higher accuracies for the algorithms.

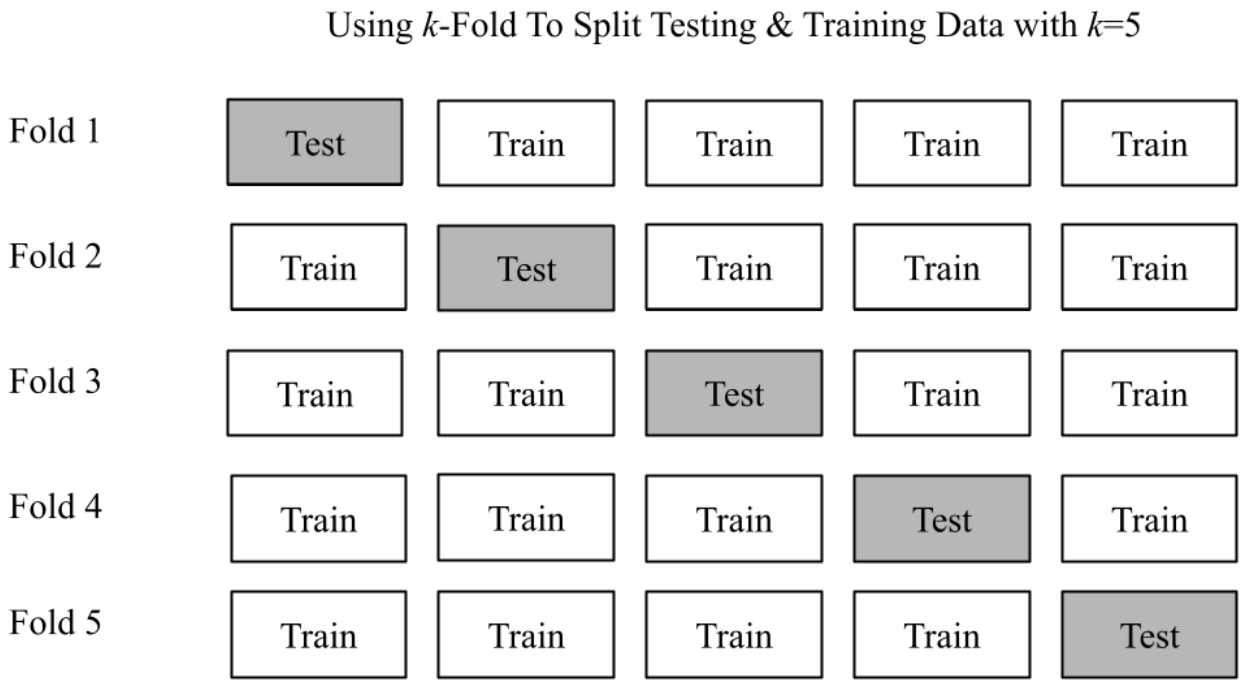


Fig. 4: K-fold. *k*-fold is a cross validation technique where the full data is randomly split into *k* equally sized pieces.

Logistic Regression

Logistic regression uses a logistic function to model the probability of an event that has two possible outcomes, such as win/lose or yes/no. Logistic regression uses a best-fitting logistic S-shaped (sigmoid) curve to predict an association between two variables. This curve has a midpoint at the x-axis. Anything to the right of the x-axis approaches y = 1, predicting the 1 outcome. Anything to the left of the x-axis approaches y = 0, predicting the 0 outcome. This logistic function can be modeled with the general form:

$$y = 1/(1 + e^{-t(x-x_0)}) \quad (1)$$

Here x_0 is the x-value of the midpoint of the sigmoid curve, 1 is the maximum value of the function, and *t* is the logistic growth rate, or the steepness of the curve.

Logistic regression is ideal for the context of the RRT investigation. In a logistic curve, the value y = 0 can represent not needing RRT six months after a CKD diagnosis, whereas the value y = 1 can represent needing RRT after six months. A value that is closer to y = 1 on the logistic curve would suggest a higher probability of needing RRT, and vice versa for y = 0.

Other Algorithms

While logistic regression was the primary means by which we evaluated the data, a variety of other algorithms were also utilized for direct comparison. Out of the other algorithms we utilized, the following were the most efficient and will be explained further below: Decision Tree, Random Forest, Gaussain Naive Bayes, Bernoulli Naive Bayes, and XGBoost algorithms.

Decision Tree utilizes a binary tree structure to decide outcomes based on several features in the dataset. The algorithm can be likened to a flow chart, where depending on the input, the output will be yes, no, or continue through the flowchart (18).

Random Forest algorithm implements multiple decision trees and averages their outputs, which is used to classify the input data (19).

Adaboost is a boosting algorithm, which means it starts with a weak model and continuously improves by utilizing the mistakes of previous models. All of these models have a weight value, which the algorithm determines based on the correctness of a model's prediction in addition to how much the model affects the final output. Adaboost consists of decision trees called stumps, which only have one node (20).

XGBoost is another boosting algorithm that is similar to Adaboost. Unlike Adaboost, XGBoost's trees have multiple nodes, along with extra tools, such as regularization, tree pruning, and more that enable the program to run efficiently. The amount of trees used and aspect of the individual trees can be altered to fit the algorithm to the data (21).

The Gaussian Naive Bayes algorithm operates on Bayes' theorem for conditional probability, assuming all features are mutually independent in addition to a normal/Gaussian distribution. This leads to high performance with continuous features (22) .

The Bernoulli Naive Bayes algorithm also assumes all features are mutually independent. However, it is specifically programmed to work with binary data (23).

Multilayer Perceptron Neural Network is a simple feedforward network where data is inputted into layers of neurons that transform the data using arithmetic operations, eventually outputting a probability for each classification. The algorithm self-learns through backpropagation, where each neuron is finetuned via gradient descent (24).

Other algorithms were tested, but they were comparably more inefficient with default parameters and did not predict patients any better than the previously described algorithms. These included the Support Vector Machine (SVM), Gradient Boosting Classifier, Gaussian

Process Classifier, and Quadratic Discriminant Analysis algorithms. In total, eight different algorithms were utilized, including the previously described logistic regression algorithm.

RESULTS

Assessing the Goodness of a Model

The simplest way to assess the goodness of a model is by calculating accuracy, which is a single percentage that represents how correct a model is. The accuracy is found by counting the number of correctly predicted values and dividing by the total number of predictions. While accuracy is simple to calculate and compare, it is not an appropriate measure when the data is not balanced. For instance, consider a dataset where 90% of patients ended up needing RRT within 6 months. If a model simply learns to always predict that patients need RRT within 6 months, it will have an accuracy of 90%. While the accuracy of the model on paper is high, it fails to predict anything meaningful. Thus, other metrics are often needed to truly evaluate how well a model performs.

The Receiver Operating Characteristic (ROC) curve, a probability metric that is used for binary classification, addresses this problem by explaining the correlation between the sensitivity (true positive rates) versus the specificity (true negative rates). The true positive rates in our study would represent people who need RRT who were correctly predicted to need RRT, while true negative rates represent people who do not need RRT that were correctly predicted to not need RRT. The Area Under the Curve (AUC) represents the measure of separability of the classes. In other words, if the AUC is higher, the model is more likely to predict a class of 0 being 0 and a class of 1 being 1. If AUC = 1, the model predicts all positives as positives and all negatives as negatives, but if AUC = 0, the model predicts all positives as negatives and all negatives as positives. If AUC = 0.5, the model is making predictions entirely randomly. In our study, AUC = 1 would indicate that the model correctly predicted which people need RRT and which do not, and AUC = 0.5 would indicate that the model is not finding any correlation between the data and the outcomes and is making random predictions as a result.

Results of ML Algorithms

In this experiment, we tested the predictions from all of the algorithms. The results are shown in Table VII and Figure 5.

Table VII: Algorithm performance

Algorithm Name	Accuracy	Sensitivity	Specificity	AUC
XGBoost	0.78	0.17	0.98	0.73
Bernoulli Naive Bayes	0.77	0.24	0.94	0.72
AdaBoost Classifier	0.78	0.26	0.95	0.72
Random Forest Classifier	0.78	0.23	0.95	0.72
MLP NN	0.73	0.39	0.84	0.68
Logistic Regression	0.74	0.36	0.86	0.67
Decision Tree	0.71	0.34	0.83	0.60
Gaussian Naive Bayes	0.37	0.82	0.23	0.52

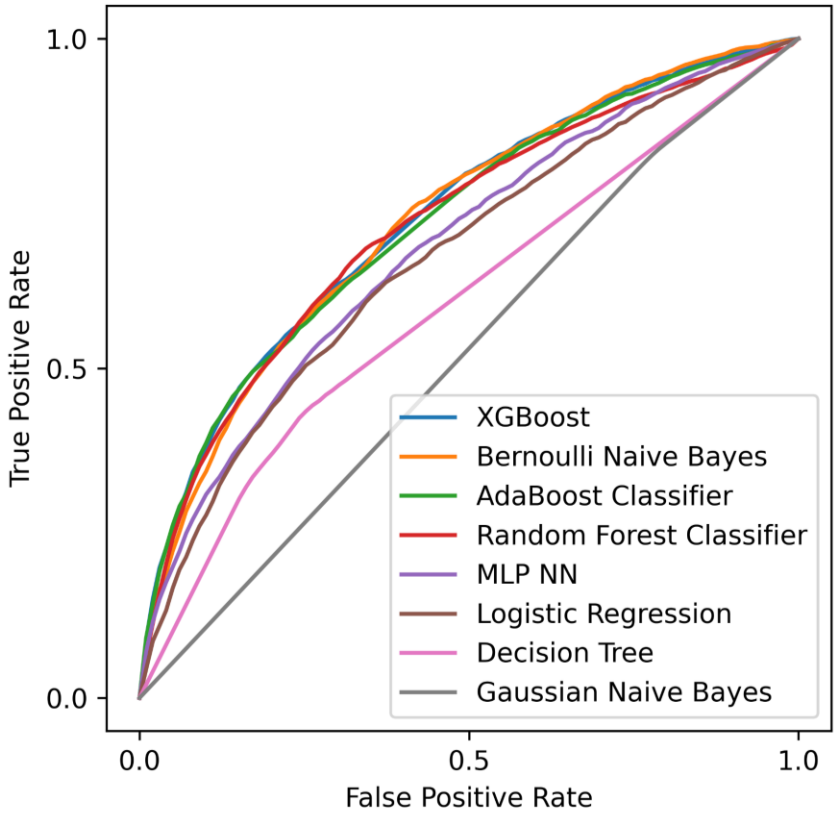


Figure 5: ROC curves with results of different algorithms. The key shows the algorithm performances in order from best to worst. The XGBoost, Bernoulli Naive Bayes, and AdaBoost Classifier were most successful, while the Gaussian Naive Bayes was the least successful.

In particular, AUC values lie on the interval [0.5,1]. This indicates that the more diagonal a curve, the lower the AUC value and thus is a worse model. However, curves that are above a straight diagonal line indicate that the model is performing better. From this observation, we can directly note that the Gaussian Naive Bayes was the worst performing algorithm, with Decision Tree closely following. The best performing algorithm was XGBoost, followed by Bernoulli Naive Bayes and AdaBoost Classifier.

Because many of these algorithms are curved above the straight diagonal line, we can verify that ML can be used to predict if a person diagnosed with Chronic Kidney Disease received RRT. Most of the algorithms tested are successfully finding a pattern in the data and are not making random predictions.

DISCUSSION

Complexities of ML in Healthcare

Despite the increasing applications of ML, critical issues still remain. Bias typically creeps into systems through inadequate datasets; if training data lacks information on a specific feature to clearly classify, then the ML network will have difficulty identifying and working with said feature. In our investigation of ICD-9 codes, many key demographics may have been left out of the data set. It is possible that the underserved population may not have easy access to medical care and therefore may not be represented enough within the dataset. Additionally, individuals without healthcare, the privilege to afford checkups, or other barriers to healthcare would not be included in the ICD-9 set. Creating a more representative set of ICD-9 codes could be remedied by gradually enacting systemic change in how individuals access healthcare. Additionally, groups could work together to incentivize individuals at risk of CKD to seek medical care and diagnoses; however, the issue of healthcare access is incredibly nuanced, as individuals see different merit in creating universal healthcare systems. Particularly in the United States, which operates without a universal healthcare system, pushback to expanding access to the healthcare system is rampant. Thus, including these underrepresented demographics in the data presents numerous challenges. However, if the ICD-9 data were more balanced (e.g. encompassing greater demographics), then we may have revealed a link missing from the rest of the population between RRT and a specific illness.

Privacy remains a significant concern in the digitization of healthcare work and application of ML algorithms. “Privacy in the age of medical big data” proposes that the use of the data for predictive studies comes down to reciprocity: if a patient will benefit from their data being used, medical professionals should take it (25). This represents just one of many opinions on the matter. In fact, it strongly applies to this investigation of predicting RRT. In this study, while individuals who had their information used and later developed RRT did not reap a major

benefit, individuals who have not yet developed RRT will benefit, as they helped refine the algorithm that may eventually predict their need for the therapy. Utilizing predictive technology, such as RRT predictions, presents additional controversy in relation to health insurance. If an individual is predicted to need RRT within six months and their health insurance provider gains access to this information, they may change their rates, threatening their access to medical care over time.

The most critical concern in our study is that of false negatives. Here, a false negative would occur if the algorithm predicts an individual would not need RRT within 6 months when in reality they would need RRT. This can be particularly dangerous, as it may hinder access to life-saving dialysis. No model can be completely accurate; thereby, relying on a computational algorithm to predict medical care when lives are at risk brings up a number of ethical questions. One approach toward deciding whether to use the algorithm compares the accuracy rate of the algorithm to the accuracy rate of a doctor. If the algorithm predicts if a patient needs RRT more effectively than the doctor, then they will use the algorithm. However, it is important to be mindful that there are limitations in the automated predictions and not to stake choices too heavily on the outcome.

Importance in Predicting RRT

In recent years, Chronic Kidney Disease has impacted more and more of the population, currently affecting roughly 10% of the world's population, or around 750 million people (26). In the United States, 1 in 7 people have CKD (27). These increasing rates are the result of a number of factors, including the commonality of risk factors such as diabetes or difficulty accessing healthy foods. As such, properly diagnosing and treating CKD is essential to improving public health. The ability to predict this need for dialysis using machine learning can dramatically improve patient outcomes.

ACKNOWLEDGEMENTS

We would like to thank the authors of the paper "Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients," Dovgan et al., for generously allowing us to utilize their data.

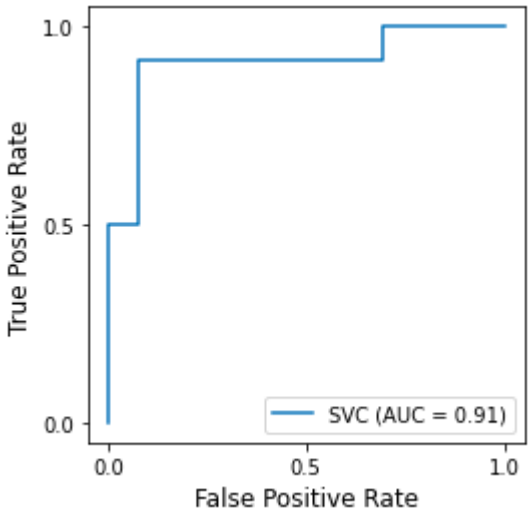
REFERENCES

1. Fraser S, Blakeman T. Chronic kidney disease: Identification and management in primary care. Pragmatic and Observational Research. 2016 [accessed 2021 Jul 29];Volume 7:21–32. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5087766/. doi:10.2147/por.s97310
2. Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management. JAMA. 2019 [accessed 2021 Jul 29];322(13):1294.

https://jamanetwork.com/journals/jama/article-abstract/2752067. doi:10.1001/jama.2019.14745
3. Delanaye P, Glassock RJ, Pottel H, Rule AD. An Age-Calibrated Definition of Chronic Kidney Disease: Rationale and Benefits. 2016 [accessed 2021 Jul 29];37(1): 17–26. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4810758/#.
4. Queeley GL, Campbell ES. Comparing treatment modalities for end-stage renal disease: A meta-analysis. American Health & Drug Benefits. 2018 May [accessed 2021 Jul 29]:118–127. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5973249/. doi:10.1016/j.jval.2014.03.1690
5. Vadakedath S, Kandi V. Dialysis: A review of the mechanisms underlying complications in the management of chronic renal failure. Cureus. 2017 [accessed 2021 Jul 29]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5654453/. doi:10.7759/cureus.1603
6. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: Icd code accuracy. Health Services Research. 2005;40(5p2):1620–1639. doi:10.1111/j.1475-6773.2005.00444.x
7. Buck CJ. 2015 ICD-9-CM for Hospitals, volumes 1, 2, & 3. St. Louis, MO: Elsevier; 2015.
8. Mohapatra S, Nath P, Chatterjee M, Das N, Kalita D, Roy P, Satapathi S. Repurposing therapeutics for COVID-19: RAPID prediction of commercially available drugs through machine learning and docking. PLOS ONE. 2020;15(11). doi:10.1371/journal.pone.0241543
9. Whitfield TW, Ragland DA, Zeldovich KB, Schiffer CA. 2020. Characterizing protein-ligand binding using atomistic simulation and machine learning: Application to drug resistance in HIV-1 protease. J Chem Theory Comput. 16(2):1284–1299.
10. Sriram TV, Rao MV, Narayana GS, Kaladhar DS, Vital TP. 2013. Intelligent Parkinson disease prediction using machine learning algorithms. International Journal of Engineering and Innovative Technology (IJEIT). 3(3): 1568-1572.
11. Jiang Y, Benz TL, Long SB. Substrate and PRODUCT complexes reveal mechanisms of Hedgehog Acylation by hhat. Science. 2021 [accessed 2021 Jul 30];372(6547):1215–1219. https://science.sciencemag.org/content/372/6547/1215. doi:10.1126/science.abg4998
12. Dovgan E, Gradišek A, Luštrek M, Uddin M, Nursetyo AA, Annavarajula SK, Li Y-C, Syed-Abdul S. 2020. Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients. PLoS One. 15(6):e0233976.
13. Popkin G. 2019. Data sharing and how it can benefit your scientific career. Nature. 569(7756):445–447.
14. Vandette K. 2018. Replication studies are vital, so why are there so few of them?. Earth.com. https://www.earth.com/news/replication-studies-vital-few/
15. Tsai M-H, Hsu C-Y, Lin M-Y, Yen M-F, Chen H-H, Chiu Y-H, Hwang S-J. 2018. Incidence, prevalence, and duration of chronic kidney disease in Taiwan: Results from a community-based screening program of 106,094 individuals. Nephron. 140(3):175–184.
16. Murphy D, McCulloch CE, Lin F, Banerjee T, Bragg-Gresham JL, Eberhardt MS, Morgenstern H, Pavkov ME, Saran R, Powe NR, et al. Trends in prevalence of chronic kidney disease in the United States. Annals of Internal Medicine. 2016 [accessed 2021 Jul 29];165(7):473. https://www.acpjournals.org/doi/10.7326/M16-0273?url_ver=Z39.88-

2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dpubmed&. doi:10.7326/m16-0273
17. Agency for Healthcare Research and Quality, 2017. *Inventory and Prioritization of Measures To Support the Growing Effort in Transparency Using All-Payer Claims Databases*. Rockville, MD.
18. Kingsford C, Salzberg SL. What are decision trees? Nature Biotechnology. 2008 [accessed 2021 Jul 29];26(9):1011–1013. https://www.nature.com/articles/nbt0908-1011. doi:10.1038/nbt0908-1011
19. Breiman L. Random Forests. Machine Learning. 2001 [accessed 2021 Jul 29];45(1):5–32. https://link.springer.com/article/10.1023/A:1010933404324#citeas. doi:10.1023/a:1010933404324
20. Wang R. AdaBoost for feature Selection, classification and its relation with SVM, a review. Physics Procedia. 2012 [accessed 2021 Jul 29];25:800–807. https://www.sciencedirect.com/science/article/pii/S1875389212005767. doi:10.1016/j.phpro.2012.03.160
21. Chen T, Guestrin C. Xgboost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 [accessed 2021 Jul 29]. https://dl.acm.org/doi/abs/10.1145/2939672.2939785. doi:10.1145/2939672.2939785
22. Pérez A, Larrañaga P, Inza I. Supervised classification with CONDITIONAL Gaussian NETWORKS: Increasing the structure complexity from naive bayes. International Journal of Approximate Reasoning. 2006 [accessed 2021 Jul 29];43(1):1–25. https://www.sciencedirect.com/science/article/pii/S0888613X0600003X. doi:10.1016/j.ijar.2006.01.002
23. McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. AAAI/ICML-98, 41–48.
24. Murtagh F. Multilayer perceptrons for classification and regression. Neurocomputing. 1991 [accessed 2021 Jul 29];2(5-6):183–197. https://www.sciencedirect.com/science/article/abs/pii/0925231291900235. doi:10.1016/0925-2312(91)90023-5
25. Price WN 2nd, Cohen IG. 2019. Privacy in the age of medical big data. Nat Med. 25(1):37–43.
26. Haileamlak A. 2018. Chronic Kidney Disease is on the Rise. Ethiop J Health Sci. 28(6):681–682.
27. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, Van Lente F, Levey AS. Prevalence of chronic kidney disease in the United States. JAMA. 2007 [accessed 2021 Jul 29];298(17):2038. https://jamanetwork.com/journals/jama/fullarticle/209357. doi:10.1001/jama.298.17.2038

APPENDICES



Appendix A: Sample ROC curve. This sample plot using demo data plots an ROC curve. The Area Under The Curve is 0.91, indicating that the model is fairly accurate in predicting positives as positives and negatives as negatives. Instead of plotting the sensitivity and specificity directly, the sensitivity is compared to 1-specificity.