TechStat Health Solutions

Executive Summary: Final Report

Chronic Kidney Disease Detection and Prevention Plan

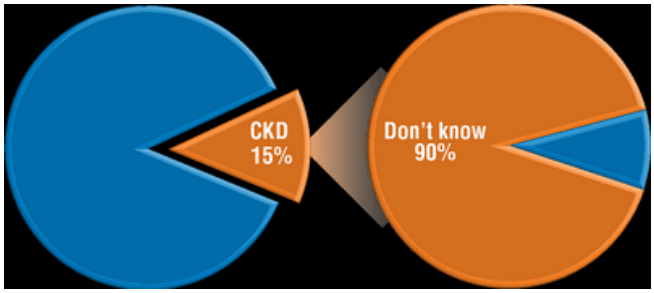
May 23, 2023

Team Members: Sylvia Chung, Kati LiPetri, Nick Malone, Spencer Puterbaugh, and Josh Squires

Table of Contents

Problem Overview

According to the United States Centers for Disease Control and Prevention (CDC), there are approximately 37 million adults in the US (15% of the adult population) who have chronic kidney disease (CKD), but 9 out of 10 of those adults are not aware that they have CKD[1].



Kidneys are an important organ of the human body that function to filter waste and toxins from blood, filtering all blood in a person's body every 30 minutes. CKD is a medical condition with varying levels of severity in which a person's kidneys are damaged and are unable to filter blood as well as they should. In a person with CKD, waste and toxins are not fully filtered and removed from the blood, meaning they remain in the body and cause additional health problems like heart disease, stroke, anemia, increased infections, mineral imbalance in the body, loss of appetite, and lower quality of life[2].

If left untreated, CKD eventually turns into kidney failure, also known as end-stage renal disease (ESRD). By this point dialysis or kidney transplant are needed for the person to survive. It is not possible to reverse damage already done to kidneys, but if caught earlier, steps can be implemented to slow the progression of CKD[2].

TechStat Health Solutions is a healthcare consulting company dedicated to improving the health and lives of people around the world. TechStat's goals are to develop a solution for earlier detection and prevention of CKD and ESRD in the US that is easy to access and use by healthcare providers and patients.

The Initial Findings report outlines TechStat's initial findings to better detect and prevent CKD and ESRD. The most significant factors in detection of CKD from health will be determined along with non-health factors that could play a role in whether a person develops CKD, such as environmental pollution levels (air and water), food desert locations, and socioeconomic status throughout the US.

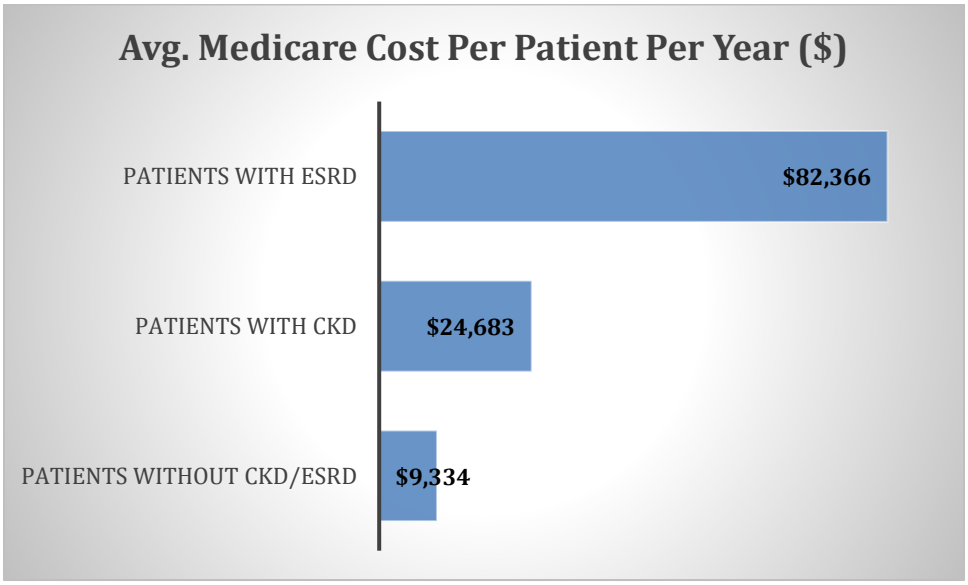
---

[1] "Using Electronic Health Records to Identify Patients with Chronic Kidney Disease." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, February 18, 2022. https://www.cdc.gov/kidneydisease/publications-resources/electronic-health-records.html

[2] "Chronic Kidney Disease Basics." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, February 28, 2022. https://www.cdc.gov/kidneydisease/basics.html

Financial Impact

Chronic Kidney Disease not only impacts people's health and livelihood individually, but the collective economic burden of CKD on the US healthcare system is substantial. According to data on Medicare spending in 2020, the average cost per patient per year was 2.6x more for patients with CKD and 8.8x more for patients with ESRD when compared to patients without CKD or ESRD[3].



In 2020, 13.9% of Medicare fee-for-service patients aged 66 or older were diagnosed with CKD, and this group accounted for about 25% of total Medicare fee for service spending (or $75 Billion). In addition, Medicare fee for service spending for patients of all ages diagnosed with CKD accounted for 23.5% of total Medicare fee for service spending (or $85.4 Billion)[3].

Implementation of early detection and prevention of CKD has the potential to save Medicare up to $53 Billion per year based on costs per patient per year and total spending on CKD in 2020.

[3] United States Renal Data System. 2022 *USRDS Annual Data Report: Epidemiology of kidney disease in the United States.* National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2022.

CKD Predictive Modeling

The CKD dataset was used to develop a model for prediction of CKD in patients using health care data. Feature selection was performed to decrease the number of health data variables used in the model from 24 to 5. In addition, 2-component PCA and t-SNE were performed on the dataset before and after feature selection. The table below shows a summary of the train and test accuracy for several different classification models performed with each dataset. The green outline denotes the train and test accuracies for the chosen model.

Model Summary: Scaled Data, PCA Data, t-SNE Data with Top 5 Selected Features

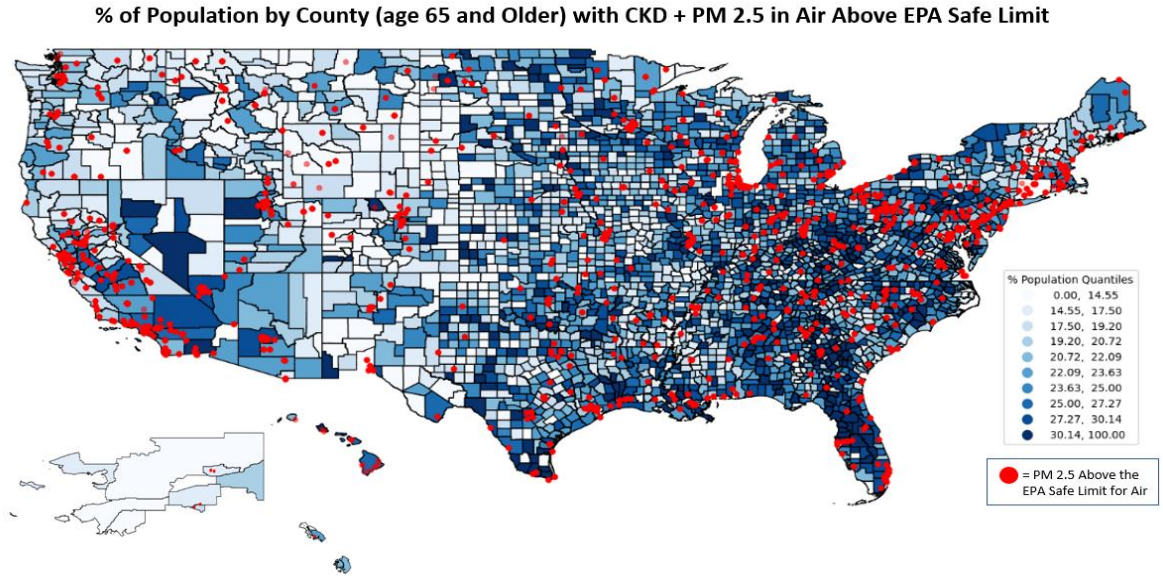
Model Type	Scaled Data		PCA		t-SNE	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
Logistic Regression	99%	99%	99%	98%	100%	99%
Decision Tree	100%	100%	100%	99%	100%	99%
Random Forest	100%	99%	100%	99%	100%	99%
SVC	99%	99%	100%	98%	99%	99%
KNN	99%	99%	100%	99%	99%	99%
Naive Bayes	85%	83%	99%	99%	98%	98%

Overall, the models performed best using the PCA of the scaled data, and the decision tree and random forest models performed the best on all three datasets. Train accuracy and test accuracy did reach 100% in some cases. This is usually unrealistic and impractical in ML models, but since our dataset is relatively small, with 400 rows, it is not uncommon. The training and test accuracies will likely change once additional patient data is collected and added to the dataset in the future. Adding more patient data in the future will be required to develop a more robust model. The PCA, t-SNE, and machine learning modeling was performed using Python sklearn in Google Colab.

Other Factors Affecting CKD Prevalence

Based on the Water Quality dataset, there is no clear relationship with the identified water contaminants that are linked to kidney disease and prevalence of kidney disease in US counties. There are, however, some areas of contaminated water that need to be improved for public health and safety purposes, in Florida, South Carolina, and Montana.

Based on the Air Quality dataset, PM 2.5 is highly linked to counties in the US with increased prevalence of diagnosed CKD, as shown in the figure below.



PM 2.5 is defined as particles in the air that are 2.5 microns in diameter or smaller. When people breathe in PM 2.5, these particles can cross the blood gas barrier and enter the kidneys where they accumulate. With enough accumulation of PM 2.5, the kidneys become damaged, causing renal injury, and eventually leading to CKD. The main current sources of PM 2.5 are due to industry and include cooking smoke, welding smoke, cigarette smoke, smoke from heating emissions, and traffic exhaust[4]. Governments should investigate ways to reduce PM 2.5 to ultimately reduce the harmful effects on public health, especially as it relates to CKD.

Lastly, the relationships between food access and fast-food restaurants with CKD rates were studied. Based on the data, there is a significant percentage of people living in the US that have low access to fresh food at grocery stores (i.e., food deserts), many states with 20% of the

[4] Xu, Wenqi, Shaopeng Wang, Liping Jiang, Xiance Sun, Ningning Wang, Xiaofang Liu, Xiaofeng Yao, et al. "The Influence of PM2.5 Exposure on Kidney Diseases." *Human & Experimental Toxicology* 41 (February 17, 2022). https://doi.org/10.1177/09603271211069982.

population or more. For fast food establishments per 100,000 people, only a few states have notably more restaurants (about 200 or more per 100,000 people): California, Texas, Florida, and New York.

Correlations between both food access and fast-food restaurants versus CKD rate were performed. There is a positive correlation between the percentage of a state's population in a food desert and the state's CKD rates. This makes sense because food deserts restrict people's access to healthy food. There is no clear correlation between the number of fast-food establishments per state and CKD rate per state. This also makes sense because most states have about the same number of fast-food restaurants per 100,000 people. Lastly, an analysis on income and its correlation to CKD was performed, but there was no correlation between lower income and rates of CKD in the US.

Web/Mobile Applications

To support our client's needs, a web-based application was developed that includes a predictive component, as well as a dashboard for data analysis and retrieval. End users can leverage the predictive endpoint to input certain health-related data and receive output that indicates their level of risk regarding developing chronic kidney disease. The predictive component uses a simple form-based approach, making it very easy and quick for the users to input their information and receive a prediction within moments. Behind the scenes, the predictive component is using Principal Component Analysis to reduce the dimensionality of the dataset and then passing these values to the Decision Tree Classifier model that was developed for this project. The dashboard portion of the application provides a range of functionality that allows healthcare providers with an appropriate level of access to view healthcare related information for patients that are under their care; these providers can filter, sort, add/remove/edit data, and export data to an excel file format for their own use. Data visualizations have been included that allow for insights to be gleaned from a glance, and charts can be added or removed by the development team as needed to support our client.

The application hosted by Google Cloud Platform can be accessed over the internet with a browser at the following link: https://techstat-app-u63grmon7a-uc.a.run.app/.

To log into the the 'Provider Overview' portion of the application, use the following username and password:

user: dr_jones
password: techstat_dashboard

In addition to the web-based application, the current trend in our industry is to have a mobile application that also supports the needs of our clients; to that end, an Android-based mobile application was also developed that provides a means for our clients to offer an 'on-the-go' experience, with the ability to access the features and functionality of the web-based application anywhere from their mobile device. An iOS version of the application will also be developed in the next phase after funds are released, to ensure the accessibility of the mobile application by as wide a user base as possible.

Recommendations

For the CKD predictive model, the TechStat team recommends the decision tree model using the two-component PCA dataset with only the top 5 features selected. This model performed very well with the CKD dataset and is also optimized to use fewer computing resources. While the two-component t-SNE dataset also performed well with this model, there are several challenges when feeding t-SNE reduced features into an ML model. T-SNE is non-deterministic, meaning that each time it runs, the results can be different even with the same hyperparameters. In addition, t-SNE is more difficult to use when new data is added to the dataset. Once additional data is added to the dataset, the t-SNE embeddings will be recalculated. Because of this, PCA was chosen since new data can be added to the current projection.

Environmental factors have potential to influence the development of CKD. The effects that air pollution can have on the body can be reduced with the use of air purifiers or filters in common indoor areas[5]. It is recommended to inform citizens that N95 Masks be worn when filters are not available and while spending time outdoors while exposure to the PM2.5 are at a high level[6]. While individuals take precautionary measures, it is recommended that regulations for people and businesses that contribute to higher levels of PM 2.5 are enforced by some financial and/or legal penalty[7]. Despite the evidence that water contaminants likely have low impact on kidney health in the US, it is still recommended to inform citizens to filter drinking water within areas

[5] 1. "Extremely High Levels of PM2.5: Steps to Reduce Your Exposure," Extremely High Levels of PM2.5: Steps to Reduce Your Exposure | AirNow.gov, accessed May 18, 2023, https://www.airnow.gov/aqi/aqi-basics/extremely-high-levels-of-pm25/.

[6] 1. James Parsons, "What Is a PM2.5 Mask, and Is It Different than an N95?," Ellessco, July 4, 2022, https://ellessco.com/blog/2022/07/pm25-mask-different-n95.

[7] 1. Yevgen Nazarenko, Devendra Pal, and Parisa A Ariya, "Air Quality Standards for the Concentration of Particulate Matter 2.5, Global Descriptive Analysis," *Bulletin of the World Health Organization* 99, no. 2 (2020), https://doi.org/10.2471/blt.19.245704.

that have tested for levels of mercury and lead beyond the EPA safety limit. These have been noted to have potential cause for kidney damage over time from the EPA[8].

TechStat recommends deploying mobile and web applications to as many healthcare providers and patients as possible. This will allow for quicker detection and treatment of CKD in patients and will help to inform patients of environmental and lifestyle factors that affect their risk of developing CKD.

The client should create a database to house information input to the mobile and web applications for the purpose of improving the robustness of the CKD predictive model in the future. The current dataset is somewhat small, so additional data will improve the accuracy of the model for the future.

[8] 1. "National Primary Drinking Water Regulations," EPA, accessed May 19, 2023, https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations.