

Problem Definition, Data Set, and Research Question

Analyze the effect of drugs and lifestyles on the survival of patients with Chronic Kidney Disease (CKD) using Machine Learning Approaches

Data Analytics: Major Research Project

**Sayed Ahmed
Graduate Student
MSc in Data Science and Analytics
500869723**

Research Questions to Answer:

Primary Research Question:

What is the effect of life-styles such as diet and exercises on the survival of patients with chronic kidney disease?

Optional Research Questions to Answer:

Analyze: what is causing/leading to kidney diseases such as CKD, and Renal Failure

- Try to understand from lifestyle data
- Try to understand from diagnosis results

Analyze and find out what treatment options provided the best outcome

- **Extended Primary:** Lifestyle such as diet, and exercises to prevent, to slow progress, to improve survival
- Effect of drugs: how drugs help and to what extent? how lifestyle and drug can work together
- Do preventing and treating hypertension and diabetes can prevent kidney disease, and slow the progress

Related Datasets

I will utilize multiple related datasets

Primary Dataset:

- Patient Characteristics i.e diagnostic results for Kidney/CKD/Renal patients
https://www.usrds.org/2018/ref/ESRD_Ref_C_PatientChars_2018.xlsx
From: <https://www.usrds.org/reference.aspx>

Other Closely Related Datasets:

- Chronic Kidney Disease dataset
<https://www.kaggle.com/mansoordaku/ckdisease>
Data has 25 features which may predict a patient with chronic kidney disease
- Chronic_Kidney_Disease Data Set
https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease
Abstract: This dataset can be used to predict chronic kidney disease and it can be collected from the hospital nearly 2 months of period.

Partially Related Datasets:

- Heart Disease dataset. Might check if this can be used in relation to other datasets
<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- Diabetes Dataset: Might check if this can be used in relation to other datasets
<https://archive.ics.uci.edu/ml/datasets/Diabetes>

Remotely Related Datasets:

1. Dialysis Facility Compare
<https://catalog.data.gov/dataset/dialysis-facility-compare-aa0fa>
2. Disease Indicators
<https://data.world/datasets/chronic-kidney-disease>

Abstract:

Chronic kidney disease (CKD) is very prevalent in today's world and continually increasing such as 30 millions of Americans now have CKD [2]. CKD and other interrelated diseases such as Hypertension, Heart Diseases, and Diabetes cause a majority of the early deaths [3]. In addition to kidney failure, CKD is also a major cause of death from stroke, and heart diseases. On the other hand, hypertension and diabetes also cause CKD. Studies show that drugs as well as lifestyle choices can prevent CKD, slow the progression of CKD [1], delay dialysis and kidney transplantation; consequently can prevent early deaths. Though there are many studies on the effect of drugs to control CKD and related complications, there are few studies on the effect of diets and lifestyles [1]. In our research, we will use the public data-set provided by USRDS [4] and other related datasets [6, 7] to study the effect of drugs, diets, and exercises on the survival of patients with CKD. We will provide emphasis on lifestyles such as exercise and diet focused treatments and study how these can improve survival and reduce mortality. We will utilize Machine Learning approaches such as Decision Trees, Random Forests, SVM, Ensemble Methods, and/or others to find relations from treatment options to survival/mortality. Optionally, we will also apply similar ML approaches to classify the causes of CKD to find the cause-effect path (i.e. cause of diabetes, effect of treatment) to draw the relation between cause and treatment option performance.

Primary Dataset:

- Patient Characteristics i.e diagnostic results for Kidney/CKD/Renal patients
https://www.usrds.org/2018/ref/ESRD_Ref_C_PatientChars_2018.xlsx
From: <https://www.usrds.org/reference.aspx>

Example Mortality/Survivability Indicators To use:

Ref: ESRD_Ref_I_Survival_2018.xlsx (From the above link)

10-year survival probabilities: incident ESRD patients

10-year survival probabilities: incident dialysis patients

10-year survival probabilities: incident CAPD/CCPD patients

10-year patient survival probabilities: incident living-donor transplant recipients

Potential Treatment Options to Consider:

- Drugs to treat Hypertension, and Diabetes
- Drugs to balance calcium, phosphorous, uric acid, and sodium
- Dialysis
- Kidney Transplantation

Lifestyle Considerations with combinations of Treatment Options above:

Dataset: ESRD_Ref_C_PatientChars_2018.xlsx

- Diets

- Low Sodium Diet
- Low Phosphorus Diet
- Low Protein Diet

Potential Future Work:

Optionally develop a complex adaptive system to analyze the cause, prevention, treatment, and progress of kidney and renal diseases

References:

1. Jacek R., Beata F., Aleksandra C., Anna G. The Effect of Diet on the Survival of Patients with Chronic Kidney Disease. *Nutrients* 2017, 9(5), 495; <https://doi.org/10.3390/nu9050495>
2. <https://www.kidney.org/news/one-seven-american-adults-estimated-to-have-chronic-kidney-disease>
3. <https://www.medicalnewstoday.com/articles/282929.php>
4. <https://www.usrds.org/reference.aspx>
5. <https://www.kaggle.com/mansoordaku/ckdisease>
6. https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease