Research Paper
NephronAI

Early detection and etiology of chronic kidney disease using deep learning

Ayush Anand
14th April 2020

Table of Contents

Abstract

The kidneys are complex and resilient organs, each roughly the size of a fist. But many medical conditions can strain the kidneys, including diabetes, and high blood pressure causing chronic kidney disease. Chronic Kidney Disease is a major public health problem that affects 1.6 billion people globally each year. Around 10% of the world population suffers from the disease and can cause complications including kidney failures, an increased risk of heart disease, high blood pressure, bone disease, and anemia. Chronic Kidney disease is a five-stage disease, often has no symptoms in its early stages and can go undetected until it is very advanced (For this reason, Chronic Kidney Disease is also referred to as a "silent disease"). Kidney disease causes gradual loss of kidney function over a period of months to years.

Each year, chronic kidney disease kills more people than breast or prostate cancer. And patient awareness is less than 10% in stages 1 to 3. Unfortunately, there has not been much research done on fighting the disease and there is an apparent lack of seriousness among public health officials against this silent healthcare crisis.

Chronic kidney disease can be treated. With early diagnosis and treatment, it's possible to slow or stop the progression of kidney disease.

For long, existing research works involving artificial intelligence have focused solely on the detection of a disease. Understanding the cause of the disease is equally and probably more important. However, to tackle a disease on such a large scale a more comprehensive work must be done. Therefore, I came up with the idea to tackle chronic kidney disease using a three-step approach – early detection, knowing the underlying cause, and recommending a lifestyle approach to minimize risks – using one of the most advanced technology in computational intelligence, 'deep learning'. The work has been implemented to make a web-based solution – NephronAI, which is accessible to anybody for free with the ease of a web browser.

A typical prediction using NephronAI (diagnosis, etiology, and creation of a recommended approach) takes less than 10 minutes for a new user, without costing any money. Therefore, it proves to be a zero-cost, accessible, and comprehensive, tool to tackle Chronic Kidney Disease.

Question

How can we fight Chronic Kidney Disease using deep learning?

How to create an AI system that not just only detects diseases but also understand the underlying cause?

Why should AI just predict diseases? It must also understand the underlying cause.

Hypothesis

I was successful in making a web-based solution – **NephronAI** for **Early detection** and **understanding of** underlying **cause** of a **Chronic Kidney Disease** using **deep learning**.

Background Research

Kidneys are essentials organs that filter some 140 liters of blood each day, leaving behind a liter or two of water and waste in the form of urine. Each kidney features a latticework of roughly one million tiny filtering units, called nephrons. Blood entering a nephron passes through a cluster of tiny vessels called the glomerulus. The thin walls of the glomerulus enable waste, water, and other small molecules to pass through while blocking larger ones such as proteins and blood cells. From there, the filtered fluid flows into kidney tubules, where the balance of minerals, water, salts, and glucose is calibrated and molecules necessary for bodily functions are reabsorbed into the bloodstream. But many medical conditions are depleting the function of kidneys. Diabetic people and those suffering from hypertension are more likely to develop chronic kidney disease. In the most advanced stages of Chronic Kidney Disease, the survival rate is less than 5 years-shorter than for many cancers.

The diagnosis of chronic kidney disease typically involves a blood test to measure the glomerular filtration rate (to check how well your kidneys are filtering your blood) and a urine test to check for albumin. If the glomerular filtration rate is lower than normal, it indicates chronic kidney disease. However, these tests might not be affordable to the lower strata of the population.

The treatment is also very costly, as much as $91,000 annually per patient in the United States. And as well as using a lot of water, the current approach consumes vast quantities of power and

materials such as plastics. However, if we increase patient awareness, make self-testing of kidney disease easy for vulnerable groups than we can save millions of lives and billions of dollars each year.

While access to physicians and high-grade healthcare facilities is still not accessible to many poor people. Plus, some physicians will turn away patients not because they do not need care but to reduce lab costs or because their insurers pay relatively less. A computational approach can effectively address all the above issues - cost, accessibility, and accuracy.

Procedure

The project entailed using deep learning for fighting chronic kidney disease through a three-phased approach – detecting the disease, finding the underlying cause, and then recommending some changes to lifestyle to reduce the effects of the disease. This section describes the proposed methodology for designing and building the application and has been divided into 3 sections.

A | Diagnosis of the disease

For the diagnosis of the disease a deep learning neural network using a dataset on chronic kidney disease has been used. The obtained dataset was converted to a machine-readable (CSV) format. This dataset contains 24 features + 1 class ('CKD' for chronic kidney disease and 'NOTCKD' for no disease) = 25 columns with around 400 rows of data, most of which are clinical in nature. Using such a high number of parameters for a very small set of data can result in underfitting and lesser accuracy in the real world. Therefore, to fit for the data only significant features must be kept and the insignificant ones removed. This is known as feature selection and is a commonly used data preprocessing technique in data mining. It not only improves the performance of the model but also reduces training time and results in better correlations.

Features were selected using a correlation-based algorithm with eight different searching techniques. The chi-square test helps you to solve the problem in feature selection by testing the relationship between the features. Given the data of any two variables, we can get observed count O and expected count E. Chi-Square measures how expected count E and observed count O deviates each other.

The Formula for Chi Square Is

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where:

c = degrees of freedom

O = observed value(s)

E = expected value(s)

When two features are independent, the observed count is close to the expected count, thus we will have smaller Chi-Square value. In simple words, higher the Chi-Square value the feature is more dependent on the response and it can be selected for model training. Implementing this algorithm, 9 independent variables were selected and used further for training. These, however, contain many missing values. And, to prevent under-fitting the missing values must be eliminated using a suitable algorithm (MICE in this case).

Benefits of using MICE: For each missing value, this method assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values. Using the seaborn module in Python3, this new data was visualized to identify trends in the features compared to the label. The trends have been discussed in the "Data Analysis" section. The cleaned data is then split into a training set (70%) and validation or testing set (30%). A Two class decision Tree algorithm was applied on the cleaned dataset (with feature selection). Performance measures such as accuracy, false positives, false negatives, specificity, AUC (Area Under Curve) for Receiver Operating Characteristics (ROC) curve are used to evaluate the trained model.

This trained model is then designed to function as a REST API. A new web application is then created in a way to take inputs from the user on the 8 attributes and integrated with the API to return a predicted value (chronic kidney disease or no chronic kidney disease).

B | Knowing the cause of the disease

Doctors often ask their patients a series of questions to narrow down on the set of plausible conditions matching the observed symptoms. This helps them know the possible cause of the medical condition. This section of the application was designed to investigate the cause for the kidney disease based on the answers provided by the user to some pre-defined questions.

A QA system such as a symptom checker, that enables the emulation of this conventional approach by asking the relevant questions to refine the differential diagnosis is the closest approach to look for this experiment. The dataset used for this section contains a list of simple and complex questions with their answers and the associated label (being the cause). An

The basic principle of working of the algorithm is very simple. Let there be N causes and M questions, here there are evaluated weights for each question which are multiplied with the value of answer it receives from the user for a question. Users answer a particular question, in the form of yes/no/don't know/, and their value is some float in the interval [0,1] (a form of fuzzy logic; 0 being less likely and 1 being very likely). This value is multiplied by the weights for the particular question and this score is recorded for each cause.

At the end, when the user has answered all questions the result set containing evaluated scores for all causes is obtained and the highest-ranked amongst them, which needs to be greater than a set minimum value is returned as the potential cause. If the highest-ranked score does not pass the above test, then the algorithm returns 'other causes' as the result.

C | Recommending the lifestyle change

The application does not stop here, it also recommends a change in lifestyle to ease the ill-effects of the disease. It lays out the biological principle of how the prime cause, as supplied by the previous layer of the application, causes chronic kidney disease in a simplified manner.

For this part of the experiment, a set of steps (lifestyle/diet/health habit changes) to fight the disease was formulated. As per the response of questions by the user from the previous step, a feedback will be provided instantly with the set of precautions/steps that the user is not adopting.

All three layers of the application are integrated to create a comprehensive and immersive solution to help fight chronic kidney disease. The UI of the application is also intuitive and easy to use with the instructions in basic English.

Data Analysis & Result

Results for deep learning model: Chronic Kidney Disease detection

In this section, the results of the experiment done for chronic kidney disease detection are discussed. The table below displays the results of the performance measures for the model such as accuracy, false positives, false negatives, specificity, AUC (Area Under Curve) for the Receiver Operating Characteristics (ROC) curve.
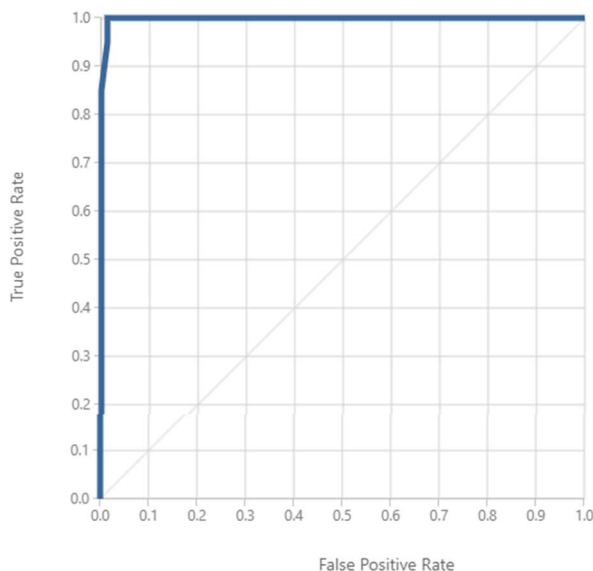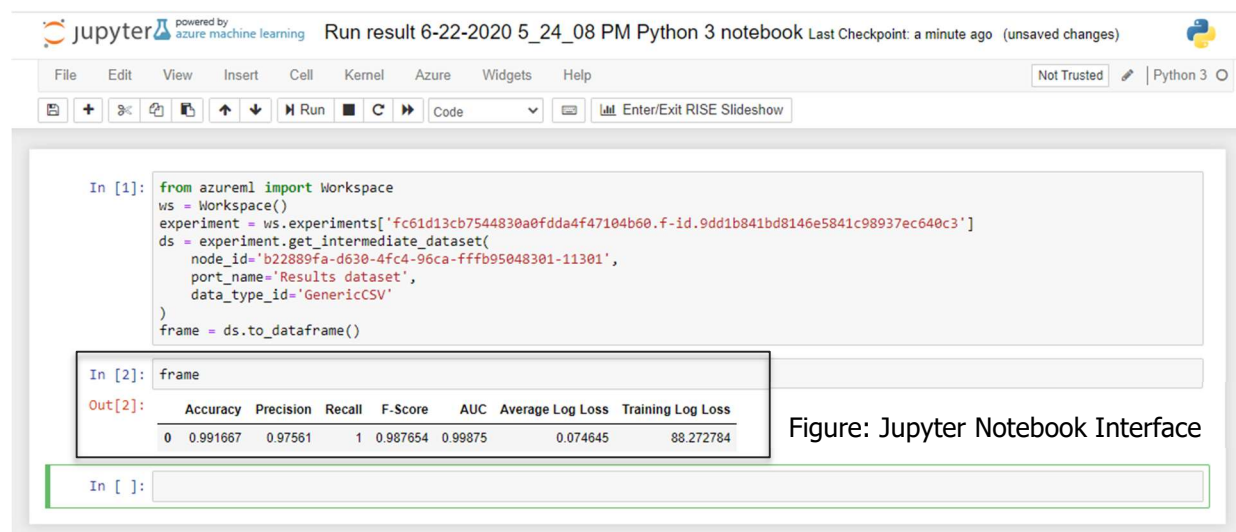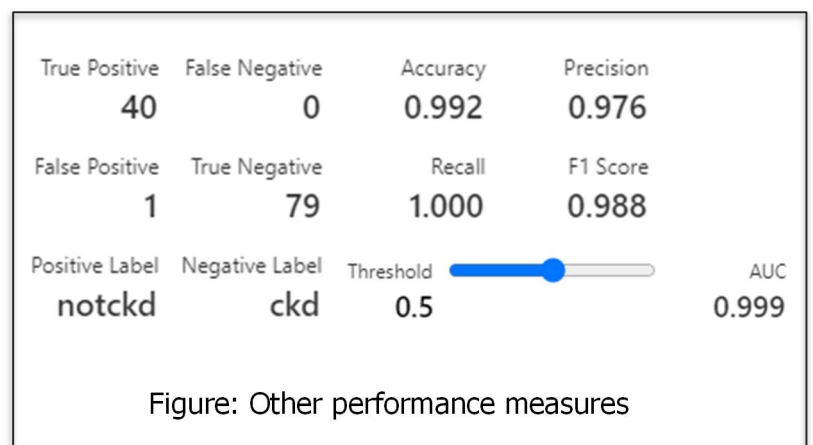

Figure: Jupyter Notebook Interface

Figure: ROC Curve


Figure: Other performance measures

Trends in data: Chronic Kidney Disease detection

Data for the detection of chronic kidney disease clearly show the correlation between parameters used to assess and the probability of being affected. For the diagnosis of chronic kidney disease a comprehensive data containing specific gravity[1], albumin[2] (in urine), serum creatinine[3], hemoglobin[4], packed cell volume[5], sodium count[6], red blood count[7], diabetes mellitus[8], hypertension[9] from the lab reports of patients were analyzed. The trend between these values (pair plotted i.e. each feature compared against each other) and the diagnosis is outlined below.
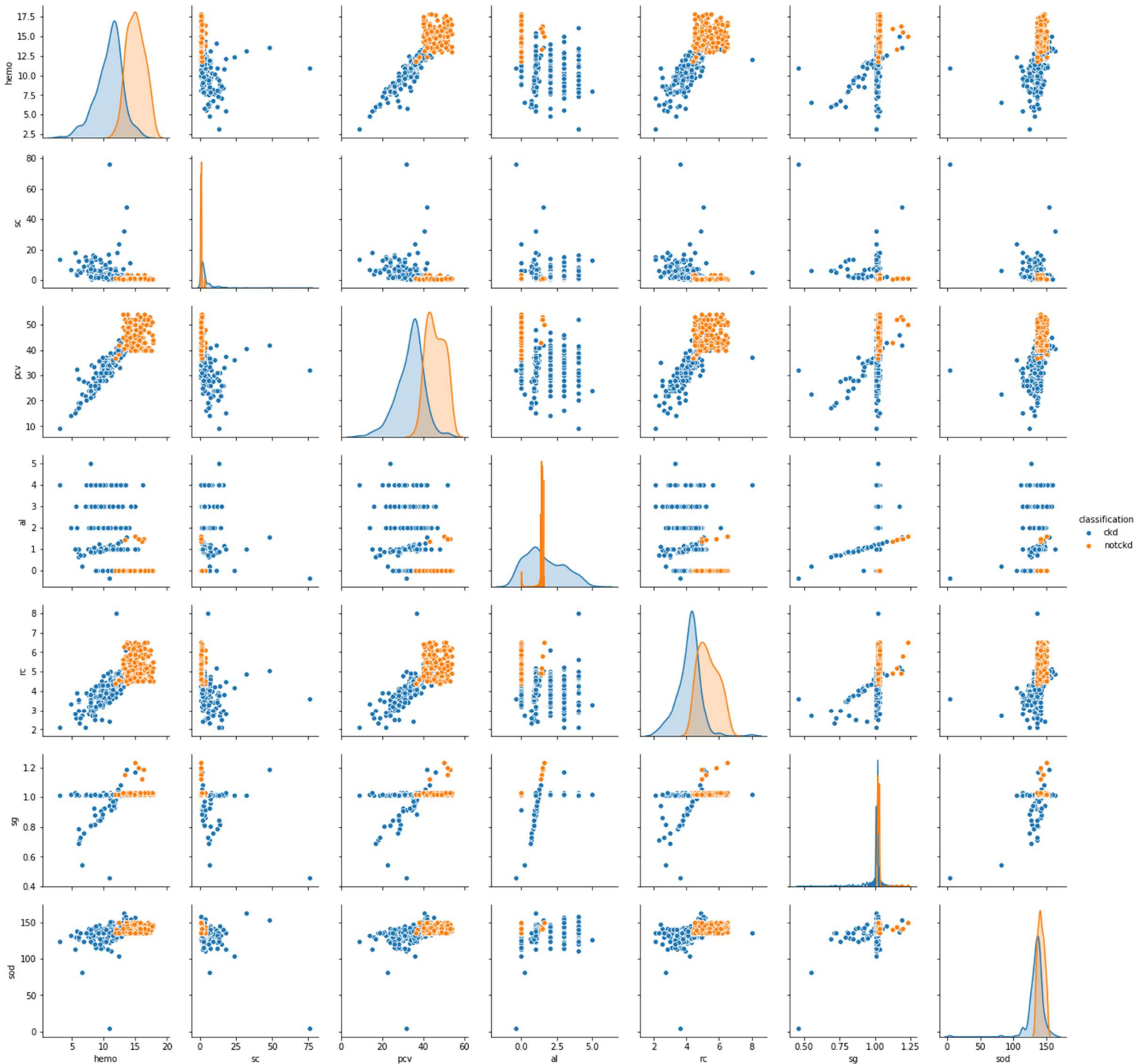
Trends in data: Causes behind Chronic Kidney Disease

From data available at the US NIH, the fact that some medical conditions account for the development of chronic kidney disease in many patients is evident.
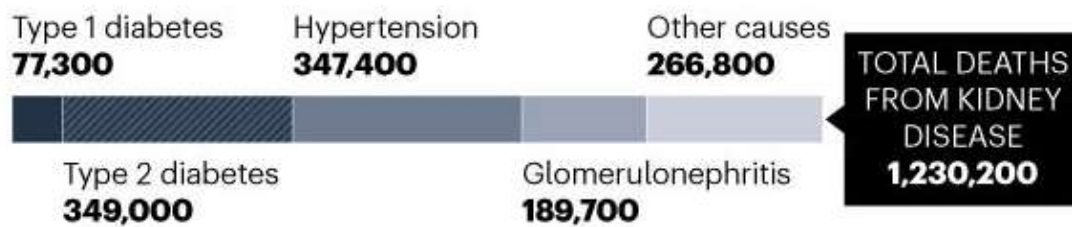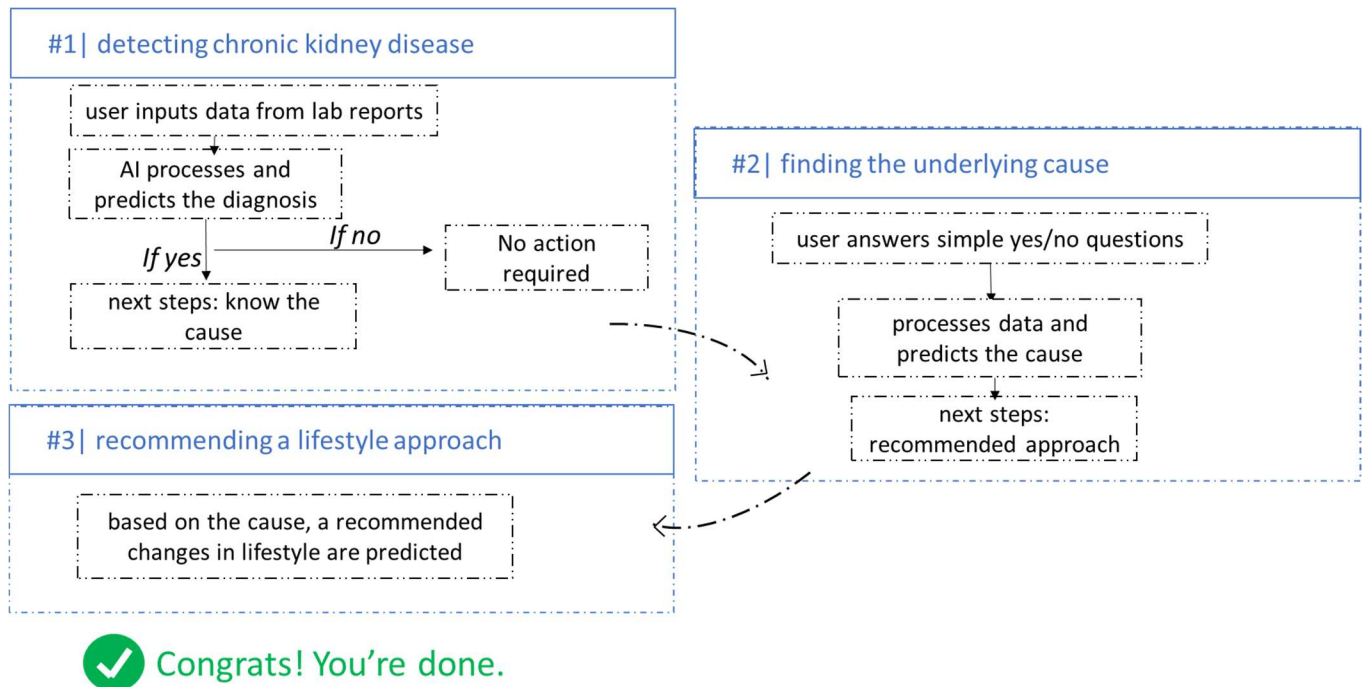

Individuals who are diabetic (both type I and II) the risk of developing chronic kidney disease is very high. The following data shows the rate per million population of those requiring hemodialysis (treatment in advanced stages of chronic kidney disease).
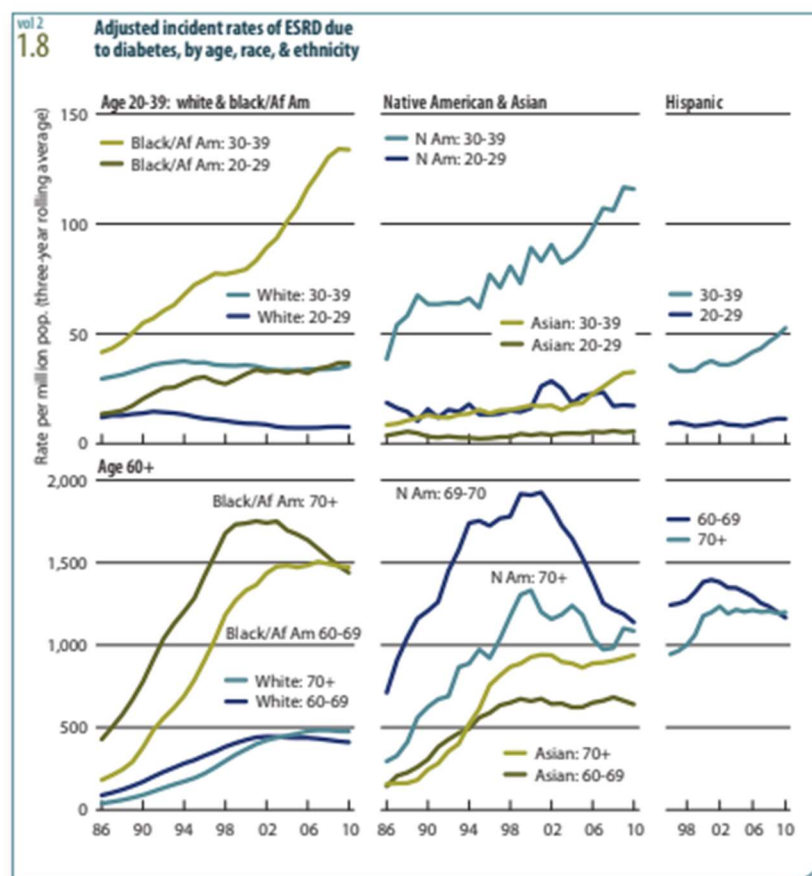
Functioning

The application works via a three-phase approach – diagnosis or detection of disease, etiology, or finding the underlying cause and the last step recommending a lifestyle approach to reduce complications based on the cause detected.

The following graphic shows the way the application works.

#1| detecting chronic kidney disease

user inputs data from lab reports

AI processes and predicts the diagnosis

If yes   If no

No action required

next steps: know the cause

#2| finding the underlying cause

user answers simple yes/no questions

processes data and predicts the cause

next steps: recommended approach

#3| recommending a lifestyle approach

based on the cause, a recommended changes in lifestyle are predicted

Congrats! You're done.

Requirements to use the service:

- A web browser (mobile, or desktop, etc.), and
- Internet Connectivity.

Conclusion

The result of the experimental project demonstrates that the proposed methodology can not only effectively learn how to detect chronic kidney disease but also investigate its cause. For instance, the neural network to diagnose patients clocked an accuracy of 99% over test data. Upon large-scale implementation, the result will prove to be beneficial because,

a. It does not require any hefty scientific or technical investment and can be used from the comfort of any home around the world "without" investing on any third-party software or integrating the main network.

b. This project aims at targeting a huge set of affected individuals, who are either unaware of their disease or lie in the more vulnerable group.

c. Using lesser parameters than a physician would require, this application produced a remarkable accuracy of 99% in detecting chronic kidney disease over the test data.

d. It does not require any kind of specialized training or technical knowledge to use and is user friendly.

e. It is completely free making it available to people all around the world without any regard for their financial status.

f. The project is embodying a more comprehensive approach to tackle chronic kidney disease. No other research in the field has focused on knowing the cause or even recommending the patient a lifestyle change.

g. This project not only addresses those patients whose primary diagnosis come out to be positive but also those who do not seem to be affected yet. "Individuals who have the disease, get to know the underlying cause and a lifestyle recommendation, but those who do not have the disease, too get a checklist to remain healthy so that they stay alert and do not develop the disease in the future."

Future Research

I would wish to pursue further research on the project with the following objectives:

a. How artificial intelligence can be applied to benefit the process of 'hemodialysis' (the treatment to chronic kidney disease).

b. Investigate into other parameters which may help us in more accurate and feasible diagnosis of chronic kidney disease.

c. How to fully integrate this project into real-world healthcare systems to increase adoption and accessibility.

Commercial prospects: There are no plans for monetizing this service anytime in the future. Healthcare institutions, public bodies, and/or other concerned organizations are welcome to pair up with this project to deliver the benefits to the end-user on a mass scale.

Who am I?

Hi, I'm Ayush Anand, a high schooler from India.

I'm an AI enthusiast and a Gen Zer who loves leveraging big data, and web technologies to build apps that address daily-life problems. I am passionate about how technology like deep learning can be put to overcome social challenges and have been dedicated to creating diverse, casual, and innovative tech solutions for social good. Right now, I am working on how AI can fight chronic kidney disease.

I'm deeply interested in knowing how Artificial Intelligence and Deep Learning can be put together into use to transform the Healthcare industry. I have a deep interest in STEM and computer technologies. Outside of STEM, I am interested in law, politics, writing, and learning new languages. My inspiration is Dr. APJ Abdul Kalam (former president of India). Also known as the "Missile Man of India", he has worked extensively for the development of Science and Technology in the country.

I'd love to continue exploring STEM to see where it takes me.

Acknowledgement

I would like to show my gratitude to the (Name Surname, title, institution) for sharing their pearls of wisdom with us during the course of this research, and I thank 3 "anonymous" reviewers for their insights in order to improve user experience with the project interface.

References

1. Rajesh Misir, Malay Mitra, and Ranjit Kumar Samanta, 2017 June 19 [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5497482/]

2. [https://www.niddk.nih.gov/health-information/kidney-disease/chronic-kidney-disease-ckd/causes]

4. [https://rarediseases.info.nih.gov/diseases/6577/heavy-metal-poisoning]

5. [https://www.webmd.com/hypertension-high-blood-pressure/guide/hypertension-symptoms-high-blood-pressure]

6. [https://www.niddk.nih.gov/health-information/diabetes/overview/symptoms-causes]

7. [https://www.webmd.com/a-to-z-guides/what-is-heavy-metal-poisoning#1]

8. [https://www.webmd.com/diabetes/tips-diabetes-lifestyle#1]

9. [https://www.healthline.com/nutrition/15-ways-to-lower-blood-sugar#section15]

10. [https://www.healthline.com/health/high-blood-pressure-hypertension/lower-it-fast#8]

12. [https://www.healthline.com/health/heavy-metal-poisoning]

13. [https://www.healthline.com/health/kidney-health/how-to-prevent-kidney-failure#11-tips]

14. [https://www.kidney.org/kidneydisease/global-facts-about-kidney-disease]

15. [https://www.niddk.nih.gov/health-information/health-statistics/kidney-disease]

16. [https://www.nature.com/articles/d41586-020-00671-8]

17. [https://www.usrds.org/2012/pdf/v2_ch1_12.pdf]

18. Sampath Kumar Gajawada 2019 October 4 [https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223]

▪ ▪ ▪f