NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY
[IJIERT] ISSN: 2394-3696 Website: ijiert.org
VOLUME 8, ISSUE 9, Sep. -2021

MACHINE LEARNING MODELING OF SEUCR TEST RESULTS FOR CHRONIC KIDNEY DISEASE (CKD) PROGNOSIS

Ernest-Okoye Ngozi
Computer Engineering Technology/Anambra State Polytechnic, Mgbakwu, Nigeria
ernestokoyengozi@gmail.com

Anigbogu Kenechukwu S.
Computer Science Department/Nnamdi Azikiwe University, Awka, Nigeria
ksy.anigbogu@unizik.edu.ng

ABSTRACT
CKD is a deadly disease that has been posing a challenge to mankind. Determining the timeline between treatment and hemodialysis has become crucial hence the intervention of bioengineering for precision and accuracy in CKD management. Machine Learning (ML) was utilized as AI algorithm in harmonizing the four established surrogates of chronic kidney failure diagnosis (SEUCR) by predicting the timeline for kidney failure. The machine sorted the SEUCR data of 1,129 susceptible CKD patients with 70% utilized in training and 30% for testing and predicted at 81% accuracy that with Creatinine (0 -14); Sodium [138 – Max limit (200)] & Urea insignificant, the kidneys are normal; Sodium [(138 – max limit (200)], Urea 0-20 & Creatinine insignificant, the kidneys are normal and Urea 0-20, Creatinine (0- 3.4) & Sodium at max limit (200), the kidneys are normal. Other arbitrary points in the engine outside the points in a, b, and c above yielded abnormal kidneys scenarios.

1. INTRODUCTION
The success and survival of the human body is dependent on the ability of separate body systems to work together. All these systems make use of the five important organs of the body which include; Brain, Heart, Kidneys, Liver and Lungs. According to Rettner, (2016), estimation of thirty trillion cells exist in the human body with at least 10 times as many bacteria as cells. The job of the kidneys as one of the organs of the body is to remove waste and extra fluid from the blood with the help of other organs and the twelve body systems. As such, jobs of the kidneys in the body are not limited to removal of waste products and extra fluid from the blood. The kidneys take urea out of the blood and combine it with water and other substances to make urine, clean blood, control chemicals and fluids in the body, help control the blood pressure and help make red blood cells.

Rettner opined that each day, the kidneys process about 200 quarts (50 gallons) of blood to filter out about 2 quarts of waste and water as urea out of the blood and combine it with water and other substances to make urine. When ones' kidneys are damaged, they may not work as well as they should. If the damage to the kidneys continues to get worse, there is tendency of the kidneys being less and less able to do their job. This condition according to Rettner leads to chronic kidney disease. Kidney failure is the last (most severe) stage of chronic kidney disease hence End-Stage-Renal-disease (ESRD). When the kidneys fail, it means they have stopped working well enough for one to survive without dialysis or a kidney transplant. This paper sorts to predict the timeline between diagnosis and actual failure to assist care givers in decision making towards CKD treatment. The dataset deployed in this paper for the determination of CKD consists of minimum, normal/average and maximum limits. Minimum limit refers to the barest obtainable limit and it transcends to the normal range; a level considered safe for the smooth operation of the kidneys in the renal system as opined by Papadakis, McPhee and Rabow (2018). The normal range leads into maximum limit which is triggered by either other ailments, excessive intake of toxic substances e.g drugs and / or hereditary factors.

2. GENERAL REVIEW
Kidney failure or end-stage renal disease is a medical condition in which the kidneys are functioning at less than 15% of normal. Ferri (2017) stated that Kidney failure can be divided into two categories: acute kidney failure or chronic kidney failure. He categorized renal failure with the difference between the trends in the

serum creatinine. Wikipedia (2019) defined Chronic kidney disease (CKD) as a health condition that develops slowly and, initially, show few symptoms with a long term consequence of irreversible acute disease/s or part of a disease progression. It follows that symptoms can vary from person to person with a patient in early stage kidney disease not feeling sick or noticing symptoms as they occur hence need for an intelligent system.

Artificial Intelligence (AI) according to encyclopedia refers to the induction of human intelligence in computers which are experts to think like humans and imitate their behaviour. This science makes use of high-level algorithms in the forms of Deep Learning, Machine Learning, Neural Networks, Bayesian networks and evolutionary algorithms. Bassem and Abu (2017) in their work titled Medical Expert System Survey described Expert systems as computer programs that are derived from a branch of computer science research called Artificial Intelligence (AI). The scientific goal of AI is to comprehend intelligence by building software that reveals intelligent comportment. The duo ascertained that AI programs which attain expert-level proficiency are capable of solving problems in any field by taking into account knowledge-based or expert systems in a precise field. Expert system therefore is earmarked for a program whose knowledge base comprises the knowledge used by a human expert, in comparison with the knowledge collected from books or non-human experts.

AI as an advanced science technology has been widely utilized in medical fields to promote medical development applied mainly to early diagnosis cum management and disease control. AI is capable of taking individual conditions into account, produce suitable decisions and promise to make great strides in kidney disease management. The above capabilities are evident in bioengineering endeavours that include Kuo, Chang and Chen (2019) developed Deep Learning approach for automatically determining the estimated GFR and CKD status. They exploited the transfer learning technique, integrating the PowerfulResNet model retrained on an ImageNet dataset in NN architecture to predict kidney function based on 4,505 kidney ultrasound images derived from creatinine concentrations. Length of the kidney was also utilized. Results were showcased with Pearson correlation coefficient of 0.741 indicated the strong relationship between Artificial Intelligence and creatinine estimations at model accuracy of 85.65. Their work according to the group is the first model for realizing the potential of transforming kidney ultrasound imaging into an effective, realtime, distant screening tool.

Segal, Kalife and Koren (2020) utilized ML algorithm for early detection of End-Stage Renal disease by predicting a model for progression to ESRD based on a large scale multidimensional database. 10,000,000 medical insurance claims from 550,000 patient records under commercial health insurance database for patients over 18 years were utilized for the prediction. Implementation was with Word2Vec algorithm and analysis with gradient boasting tree algorithm (XGBoost implementation). At 95% accuracy, the machine which was to send a warning signal to a nephrology at the NPV threshold predicted 0.715 sensitivity and 0.958 specificity with initial positive predictive value (PPV) as 0.517 and negative predictive value (NPV) as 0.981.

Charumathi (2020) developed a Deep Learning algorithm (DLA) to detect CKD from retinal images data sourced from three-population –based multiethnic, cross-sectional studies in Singapore and China. 5188 patients from 40years and above patients' records were used and 1297 to validate. Three models were trained; image (DLA), Risk factors: age, sex, ethnicity, diabetes and hypertension (RF) and Hybrid of DLA and RF. At 95% accuracy, the machine predicted that retinal image DLA shows good performance for estimating chronic kidney disease, underlying the feasibility of using retinal photography as a screening tool for community CKD populations.

This paper sorts to use Machine Learning (ML) as AI algorithm in harmonizing the four established surrogates of chronic kidney failure diagnosis by predicting the timeline for kidney failure. ML utilized in the work sorted the SEUCR data of 1,129 susceptible CKD patients with 70% utilized in training and 30% for testing.

AI involves a system that consists of software, hardware and data. From a software perspective, AI is particularly concerned with algorithms whereas hardware entails system specifications.

- **Hardware specifications:**

Computer (Minimum of: 64 bit OS, 2GB RAM (Recommended based on workload and number of batches for the model training), Windows 7).

- **Software used:**

Python 3, libraries and dependencies.
Other libraries deployed were;

- Pandas
- MatplotLib
- Seaborn
- Sklearn
- Numpy
- Ms Excel

3. METHODOLOGY AND RESULT

Seven steps of machine learning as iterated by Dovan et al (2020) was utilized in CKD prediction and they include:

i. Data collection:

The dataset for this work was generated from test samples of persons with renal disease and susceptible cum potential CKD patients gathered from local health facilities; Rock Foundation Hospital Awka, Runia Specialist Hospital, Awka, Glanson Laboratories Awka, Nnamdi Azikiwe Teaching Hospital, Nnewi and the online data collected from National Kidney Foundation archives.
These data involve various results of sodium in mmol/L, urea in mg/L, and creatinine in mg/dl as represented in the SEUCR test used for the detection and treatment cum management of CKD. Electrolytes as the fourth parameter are viewed as positively or negatively charged molecules. They refer to ions found within and between cells, in the bloodstream, and in other fluids throughout the body. In this paper, electrolyte tests will be concentrated on whole blood, plasma, or serum, usually collected from a vein or capillary in the renal system. Electrolytes are said to be of positive charge when there are presence of sodium, potassium, calcium, and magnesium; the negative ions are chloride, bicarbonate, and phosphate. The concentrations of these ions in the bloodstream remain fairly constant throughout the day in a healthy person.
Gale Encyclopaedia of Surgery (2019) states that changes in the concentration of one or more of these ions are triggered during various acute and chronic disease states and can lead to serious consequences with red alert on highly positive electrolyte. Electrolyte parameter is utilized in this work for the determination of kidney critical point failure as the slightest increase is a pointer to anomaly in all human irrespective of location or ethnic group as opined in the encyclopaedia.

ii. Data preparation:

The 1129 data samples were collated with Ms Excel, sorted horizontally to maintain line orderliness and minimize ambiguity due to similarity. The data were then set to 0/1 using Papadakis, McPhee and Rabow (2018) proposed ranges.

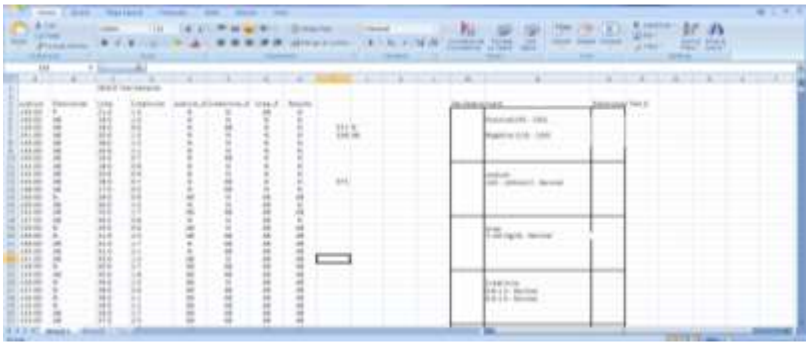


Figure 1: SEUCR dataset

iii. Choosing a model:

In the field of data science, machine learning can be described as the branch of artificial intelligence that concerns itself with extracting patterns from raw data with a set of features, using various iterative computational algorithms in order to build an effective knowledge base. This basically consists of learning a mapping function f(X) that takes the set of inputs (where X = SEUCR quantities; Creatinine, Sodium and Urea) containing features hitherto unknown to the researcher and maps them with minimal error to the outputs (Y) in the form of Normal (1) or Failed kidney (0). The general process of the mapping involves finding signals/patterns in the data after the algorithm has essentially reduced the dispersion in the data to an acceptable minimum.

iv. Training the model:

Actual training was done using python 3.7 versions in DOS prompt with matplot, SKlearn, Numpy, Pandas as toolkits as stated below:

```
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
import  matplotlib.pyplot as plt
import  numpy as np
import  pandas as pd
from sklearn import svm
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, mean_squared_error, r2_score
from sklearn import linear_model
from mpl_toolkits import mplot3d
# Load data
kidney_df = pd.read_csv("kidney_data.csv")

feature_df = kidney_df[[ 'sodium' , 'Urea' , 'Creatinine']]
X = np.asarray(feature_df)
y = np.asarray(kidney_df["Result"])

np.random.seed(42)

X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.2, random_state = 4)

classifier = svm.SVC(kernel='linear', gamma = 'auto', C = 10)
classifier.fit(X_train,y_train)
y_predict = classifier.predict(X_test)

print(classification_report(y_test,y_predict))
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

v. Evaluating the model:

The model was evaluated using .csv read to determine the minimum, maximum, standard deviation, quartile, median, upper quartile before applying the train, test, predict, classification, report algorithm engaged with Pandas, Numpy and Sklearn libraries as shown below

```
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split

# Load data
kidney_df = pd.read_csv("kidney_data.csv")
```

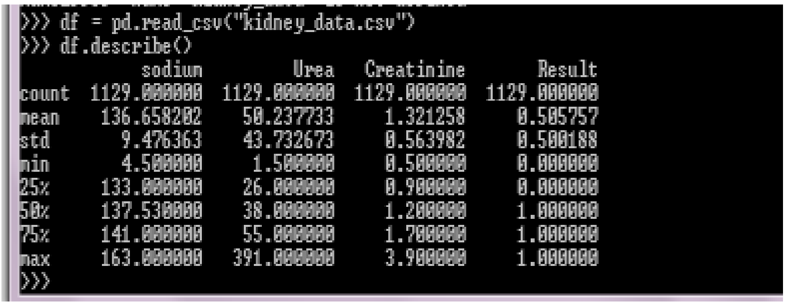
```
feature_df = kidney_df[[ 'sodium' , 'Urea' , 'Creatinine']]
X = np.asarray(feature_df)
y = np.asarray(kidney_df["Result"])

np.random.seed(42)

X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.2, random_state = 5)

print("X_train Shape: ", X_train.shape)
print("X_test Shape: ", X_test.shape)

print("y_train Shape: ", y_train.shape)
print("y_test Shape: ", y_test.shape)
```



```
>>> df = pd.read_csv("kidney_data.csv")
>>> df.describe()
            sodium         Urea   Creatinine       Result
count  1129.000000  1129.000000  1129.000000  1129.000000
mean    136.658202    50.237733     1.321258     0.505757
std       9.476363    43.732673     0.563982     0.500188
min       4.500000     1.500000     0.500000     0.000000
25%     133.000000    26.000000     0.900000     0.000000
50%     137.530000    38.000000     1.200000     1.000000
75%     141.000000    55.000000     1.700000     1.000000
max     163.000000   391.000000     3.900000     1.000000
>>>
```

Figure 2: The output from the train test yielded

X_train Shape: (903, 3)
X_test Shape: (226, 3)
y_train Shape: (903,)
y_test Shape: (226,)

vi. Parameter tuning:

The machine harmonized the three main surrogates of SEUCR criterion in order to ascertain levels of orderliness, disorderliness and relationship at various stages of the modeling. Parameters deduced from the harmonization were plotted with 3D scatter to portray relationships between the models. The plot shows no statistical bound as represented in figure below.

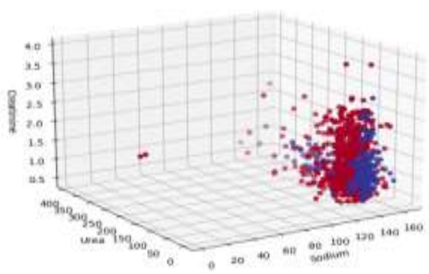


Figure 3: 3D scatter relationship of the model.

vii. Predictions/Results

Using the initial five CKD susceptible patients' data, the machine predicted at 81% accuracy thus:

```
===RESTART:\Users\DELL\Desktop\ml seucr\ng_ml1.py ============
                  precision    recall  f1-score   support
```

```
           0       0.84      0.79      0.81       117
           1       0.78      0.83      0.81       109

accuracy                               0.81       226
macro avg          0.81      0.81      0.81       226
weighted avg       0.81      0.81      0.81       226

The Kidney is: abnormal
```

Sample 2:

```
== RESTART: C:\Users\DELL\Desktop\ml seucr\ng_ml1.py ============


              precision    recall  f1-score   support

           0       0.84      0.79      0.81       117
           1       0.78      0.83      0.81       109

    accuracy                           0.81       226
   macro avg       0.81      0.81      0.81       226
weighted avg       0.81      0.81      0.81       226

The Kidney is: normal

>>>
```

Sample 3:

```
== RESTART: C:\Users\DELL\Desktop\ml seucr\ng_ml1.py ============
              precision    recall  f1-score   support

           0       0.84      0.79      0.81       117
           1       0.78      0.83      0.81       109

    accuracy                           0.81       226
   macro avg       0.81      0.81      0.81       226
weighted avg       0.81      0.81      0.81       226

The Kidney is: normal
>>>
```

Sample 4:

```
== RESTART: C:\Users\DELL\Desktop\ml seucr\ng_ml1.py ============
              precision    recall  f1-score   support

           0       0.84      0.79      0.81       117
           1       0.78      0.83      0.81       109

    accuracy                           0.81       226
   macro avg       0.81      0.81      0.81       226
weighted avg       0.81      0.81      0.81       226
```

The Kidney is: abnormal
>>>

Sample 5

```
== RESTART: C:\Users\DELL\Desktop\ml seucr\ng_ml1.py =============
              precision    recall  f1-score   support

           0       0.84      0.79      0.81       117
           1       0.78      0.83      0.81       109

    accuracy                           0.81       226
   macro avg       0.81      0.81      0.81       226
weighted avg       0.81      0.81      0.81       226
```

The Kidney is: abnormal
>>>

The five initial predictions are summarized in table 1 below:

Table 1: Summary of SEUCR data Predictions

S/N	Patients' Parameters	Reading	Parameters Used	Result
1	Blood Urea Creatinine Potassium Sodium Chloride Bicarbonate Calcium	61 3.5 4.4 140 105 25 2.2	61 3.5 140	Failed kidneys
2	Blood Urea Creatinine Potassium Sodium Chloride Bicarbonate Calcium	21 1.3 4.0 145 90 22 1.2	21 1.3 145	Normal kidneys
3	Blood Urea Creatinine Potassium Sodium Chloride Bicarbonate Calcium	17 0.5 4.4 138 106 15 2.2	17 0.5 138	Normal kidneys
4	Blood Urea Creatinine Potassium Sodium Chloride Bicarbonate Calcium	50 2.7 3.4 128 100 26 2.2	50 2.7 128	Failed kidneys
5	Blood Urea Creatinine Potassium Sodium Chloride Bicarbonate Calcium	27 1.2 3.4 125 144 23 1.8	27 1.2 125	Failed kidneys

4. CONCLUSION

a. Using ML algorithms at 81% accuracy, the predictor predicted that with Creatinine (0 -14); Sodium (138 – Max limit (200) & Urea insignificant, the kidneys are normal;

b. Sodium [(138 – max limit (200)], Urea 0-20 & Creatinine insignificant, the kidneys are normal;

c. Urea 0-20, Creatinine (0- 3.4) & Sodium at max limit (200), the kidneys are normal.
Other arbitrary points in the platform outside the points in a, b, and c above yielded abnormal kidneys scenarios.

REFERENCES

1) Bassem, S. & Abu, N.(2017). Medical Expert Systems Survey. International Journal of Engineering and Information Systems, Vol, 1 (7), pp.218-224. hal-01610722
2) Charumathi, S. (2020). DL algorithm to detect CKD from retinal photographs in community based populations. The Lancet Digital Health. Open Access May, 2020. DOI:https://doi.org/10.1016/S2589-7500(20)30063-7
3) Dovgan, E. et al. (2020). Using ML models to predict the initiation of renal replacement therapy among chronic kidney disease patients. European Union Research and Innovation Program.
4) Ferri, F. (2018). *Ferri's Clinical Advisor,* 2018 E-Book: 5 Books in 1. Elsevier Health Sciences. p. 294. ISBN 9780323529570
5) Gale Encyclopaedia of Surgery (2019). Electrolyte Tests. *A Guide for Patients and Caregivers.* Encyclopedia.com. <https://www.encyclopedia .com>.
6) Kuo, C., Chang, C. and Chen, K. (2019). Automation of the kidney function prediction and classification through ultrasound – based kidney imaging using deep learning. Digital Medicine. Open Access article 2019.
7) Papadakis, M.; McPhee, S.; Rabow, M. (2018). *Current Medical Diagnosis & Treatment* (eds.). New York, NY: McGraw-Hill Education
8) Segal, Z., Kalife, D. and Koren, G. (2020). Machine Learning Algorithm for Early Detection of End-stage Renal Disease. BMC Nephrology, article No 518
9) https://e.wikipedia.org/wiki/kidney_failure#cite_note_15. Retrieved August 2019.
10) https://www.livescience.com. The Human Body: Anatomy, Facts & functions/Live Science.