EARLY CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

HARIPRIYA C

SRIRANGANATH M

VINITHA V

VINOTHINI J

in partial fulfillment for the completion of the course

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

KGiSL INSTITUTE OF TECHNOLOGY, COIMBATORE

ANNA UNIVERSITY :: CHENNAI 600 025

MAY 2023

CERTIFICATE VALUATION

COLLEGE NAME	**:**	KGiSL INSTITUTE OF TECHNOLOGY
BRANCH	**:**	INFORMATION TECHNOLOGY
SEMESTER	**:**	VIII
TITLE	**:**	EARLY CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING

STUDENT NAME	REGISTER NO	SUPERVISOR NAME
HARIPRIYA C	711719205019	Mr. SURESHKUMAR C
SRIRANGANATH M	711719205054	
VINITHA V	711719205061	
VINOTHINI J	711719205062	

The report of this project is submitted by the above students in partial fulfillment for the completion of the course in eight semester Bachelor of Technology degree in Information Technology of Anna University and is evaluated and confirmed to the reports of work done by the above students.

SUPERVISOR

ANNA UNIVERSITY : CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"EARLY CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING"** is the bonafide work of **HARIPRIYA C, SRIRANGANATH M, VINITHA V and VINOTHINI J** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
DR. SANKAR RAM N	**MR. SURESHKUMAR C**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
PROFESSOR	**ASSISTANT PROFESSOR**
Department of **Information Technology**	Department of **Information Technology**
Kgisl Institute of Technology	**Kgisl Institute of Technology**
Coimbatore-641035	**Coimbatore-641035**

Submitted for the University project viva examination held on ______________

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

We wish to express our deep sense of gratitude to **Padmashri**. **Dr. G Bakthavathsalam**, Founder- Chairman of KG Hospital, for having provided the facilities during the course of our study in the college.

We express our heartfelt gratitude to our venerated Managing Director, **Dr.Ashok Bakthavathsalam,** KGiSL Educational Institutions, who gave the opportunity to frame the project to the full satisfaction.

We are grateful to **Dr. Selvam M**, M.E., Ph.D., Principal, KGiSL Institute of Technology for his valuable guidance and blessings.

We express our deep and sincere gratitude to **Mr. Aravind Kumar Rajendran**, CEO, Academic Initiatives, KGiSL Educational Institutions, for his motivation.

Our grateful thanks is also extended to Head of the Department **Dr. Sankar Ram N**, M.E, Ph.D., for his support throughout the project.

In this regard, we wish to express our gratitude and sincere thanks to our supervisor **Mr. Sureshkumar C**, M.E., and coordinator **Dr.Hemalatha B** for extending her inspiration and guidelines in implementing this project successfully.

Finally, we take this opportunity to extend our deep appreciation to our family and friends, for all that they meant to us during the crucial times of the completion of our project.

ABSTRACT

Chronic kidney disease (CKD) is a prevalent health condition worldwide, and early detection and management can significantly improve outcomes. Machine learning (ML) can help predict CKD risk and a combined food and exercise recommendation system can help patients manage the disease. The early CKD prediction system uses ML classifiers trained on clinical and demographic data to predict the likelihood of developing CKD. The food recommendation system analyzes a patient's dietary habits and preferences, as well as their health data, to suggest personalized meal plans and recipes. It considers the patient's CKD status and recommends foods that maintain a healthy weight and manage blood pressure. The exercise recommendation system can also be integrated, providing personalized exercise plans based on the patient's CKD status and individual needs. Exercise is a crucial component of managing CKD, and the system recommends exercises that improve overall health and well-being. The combined system provides personalized care that takes into account a patient's individual needs and health status, allowing clinicians to make informed decisions about patient care. The system helps patients manage their kidney disease and maintain a healthy lifestyle, reducing the risk of complications and improving outcomes. Overall, the ML-based early CKD prediction system with food and exercise recommendation can significantly improve patient outcomes and reduce healthcare costs.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	
	LIST OF FIGURES	
	LIST OF ABBREVIATIONS	
1.	**INTRODUCTION**	1
2.	**LITERATURE SURVEY**	3
3.	**SYSTEM ANALYSIS**	7
	3.1 EXISTING SYSTEM	7
	3.2 PROPOSED SYSTEM	8
	3.3 MOTIVATION	8
	3.4 PROBLEM STATEMENT	9
4.	**SYSTEM REQUIREMENTS**	10
	4.1 HARDWARE REQUIREMENTS	10
	4.2 SOFTWARE REQUIREMENTS	10
5.	**RESEARCH METHODOLOGY**	11
	5.1 DATASET	11

	5.2 PROCESS FLOW	12
	5.3 SYSTEM ARCHITECTURE	14
	5.4 MODEL TRAINING	15
	5.5 TESTING THE MODEL	17
	5.6 FOOD AND EXERCISE RECOMMENDATION	18
	5.7 RESULT AND DISCUSSION	19
6.	**CONCLUSION AND FUTURE ENHANCEMENT**	24
7.	**REFERENCES**	26
	APPENDICES	
A1	**SOURCE CODE**	28
A2	**CONFERENCE CERTIFICATE**	37

LIST OF FIGURES

FIGURE NO	TITTLE	PAGE NO
5.1	Data set	11
5.2	Process flow diagram	12
5.3	System architecture	14
5.4	Model Training	15
5.6	Model Testing	17
5.7	UI Interface	19
5.8	Classification about Disease	19
5.9	Patient records as per year	20
5.10	Expert suggestions	20
5.11	Recommended Foods	21
5.12	Recommended Exercises	21
5.13	Prediction	22
5.14	Negative results	22
5.15	First Stage of Disease	23
5.16	Positive Results	23

LIST OF ABBREVATION

RFA	Random forest algorithm
NB	Naïve Bayes
KNN	K Nearest Neighbor
SVM	Super vector machine
DT	Decision Tree

CHAPTER 1

INTRODUCTION

Chronic kidney disease (CKD) is a serious health condition that affects millions of people worldwide. Early detection and intervention can significantly improve outcomes, but traditional screening methods may not always be effective. Machine learning (ML) has emerged as a promising tool for early CKD detection, and a combination of ML algorithms and a food and exercise recommendation system can be a valuable asset in managing the disease. The goal of this system is to provide an early warning system for CKD, enabling patients to receive timely intervention and better manage their condition. The system is designed to be user-friendly and accessible to a wide range of patients, including those with limited technical skills. The system consists of several components: data collection, data processing, and prediction using ML algorithms. The data collection component involves gathering demographic and clinical data, such as age, sex, race, blood pressure, and laboratory tests. The data processing component involves cleaning and formatting the data, identifying missing values, and transforming the data into a format suitable for ML algorithms. The ML algorithms used in this system include decision tree, random forest, and logistic regression models. These algorithms are trained on large datasets of clinical and demographic information to identify patterns and risk factors associated with CKD. The models can predict the probability of developing CKD based on the patient's demographic and clinical information, allowing for early intervention and improved outcomes. In addition to the prediction component, the system includes a food and exercise recommendation system. Patients with CKD often require dietary modifications and regular exercise to manage their condition effectively. The recommendation system provides tailored dietary and exercise advice to patients based on their demographic and clinical information. The food recommendation system is based on the patient's dietary restrictions and preferences. The system can recommend foods that are high in nutrients and low in sodium, potassium, and phosphorus, which can help manage CKD symptoms. The system can also generate shopping lists and meal plans to help patients plan their meals more effectively. The exercise recommendation system is based on the patient's physical activity level and preferences. The system can recommend specific exercises that can help manage CKD symptoms, such as walking, cycling, and swimming. The system can also track their progress over time and provide feedback and motivation to help them stay on track. In summary, the combination of ML algorithms and a food and exercise recommendation system can provide an effective tool for managing CKD. The system can provide early warning of the disease, enabling timely intervention and improved outcomes. The food and exercise recommendation system can help patients manage their symptoms more effectively, leading to improved quality of life. The system is designed to be user-friendly and accessible to a wide range of patients.

CHAPTER 2

LITERATURE SURVEY

Aljaaf, A.J. 2018 Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC). Wellington. New Zealand. A total of 4 machine learning based classifiers have been evaluated within a supervised learning setting, achieving highest performance outcomes of AUC 0.995, sensitivity 0.9897, and specificity 1. The experimental procedure concludes that advances in machine learning, with assist of predictive analytics, represent a promising setting by which to recognize intelligent solutions, which in turn prove the ability of predication in the kidney disease domain and beyond.

A. Ogunleye, Q.-G. Wang, XGBoost model for chronic kidney disease diagnosis. IEEE/ACM Trans. Computer. Bio inform. 17, 2131–2140 (2020). They developed tha model to cover the widest range of people, the time and monetary costs of CKD diagnosis have to be minimized with fewest patient tests. Thus, the reduced model using fewer features is desirable while it should still maintain high performance. To this end, the set-theory based rule is presented which combines a few feature selection methods with their collective strengths. The reduced model using about a half of the original full features performs better than the models based on individual feature selection methods and achieves accuracy, sensitivity and specificity, of 1.000, 1.000, and 1.000, respectively.

F. Aqlan, R. Markle, A. Shamsan, "Data mining for chronic kidney disease prediction." in IIE Annual Conference. Proceedings, Institute of Industrial and Systems Engineers, (IISE 2017), pp. 1789–1794. Data mining and analytics techniques can be used for predicting CKD by utilizing historical patient's data and diagnosis records. In this research, predictive analytics techniques such as Decision Trees, Logistic Regression, Naive Bayes, and Artificial Neural Networks are used for predicting CKD. Preprocessing of the data is performed to impute any missing data and identify the variables that should be considered in the prediction models.

R.S. Walse, G.D. Kurundkar,etc…, Effective use of naïve bayes, decision tree, and random forest techniques for analysis of chronic kidney disease, in International Conference on Information and Communication Technology for Intelligent Systems. ed. by T. Senjyu, P.N. Mahalle, T. Perumal, A. Joshi (Springer, Singpore, 2020). The main objective of the research paper is comparative study of NB classifier, DT J48, and RF to analyze chronic kidney disease (CKD) patient's data and to predict how many patients are having CKD. When analyzing the same algorithm, the decision tree J48 shows that the tree variant can be 100% diagnosed with kidney disease in the future, and the random forest algorithm has analyzed it 100%.

A.Nithya, A. Appathurai, N. Venkatadri. Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images. Measurement (2020). In this work, they proposed a kidney stone detection using artificial neural network and segmentation using multi-kernel k-means clustering algorithm. Normally, the system comprises of four modules like (i) preprocessing, (ii) feature extraction, (iii) classification and (iv) segmentation. Primarily, we eliminate the noise present in the input image using median filter.

Abdullah Al Imran, Md Nur Amin, and Fatema Tuj Johora. Classification of chronic kidney disease using logistic regression, feedforward neural network and wide & deep learning. In 2018 International Conference on Innovation in Engineering and Technology (ICIET) In this research, they primary focus was to apply 3 modern machine learning techniques namely logistic regression, feedforward neural networks and wide & deep learning to diagnose CKD as well as finding the best performing technique by evaluating their diagnosis performance.

B. Navaneeth, M. Suchetha, A dynamic pooling based convolutional neural network approach to detect chronic kidney disease. Biomed. Signal Proce. Control 62, 102068 (2020). We have compared our proposed model with other existing algorithms, and it is observed that the performance achieved by this model is higher than other well-known data classification methods. Conclusion: Combining CNN with the SVM classifier enables the network to analyze the sensor data to make predictions more accurately. The use of dynamic pooling and feature pruning algorithm significantly improved the prediction accuracy of the network. The experimental results show that the proposed method provides acceptable classification accuracy and has the potential to be implemented in clinical practice. Significance: Our study result shows that the proposed methodology can be used for detecting CKD non-invasively. The proposed deep learning network provides accurate predictions compared to other data classification methods.

Imesh Udara Ekanayake; Damayanthi Herath. Dept. Computer Engineering, University of Peradeniya, Peradeniya, Sri Lanka. his work proposes a workflow to predict CKD status based on clinical data, incorporating data prepossessing, a missing value handling method with collaborative filtering and attributes selection. Out of the 11 machine learning methods considered, the extra tree classifier and random forest classifier are shown to result in the highest accuracy and minimal bias to the attributes. The research also considers the practical aspects of data collection and highlights the importance of incorporating domain knowledge when using machine learning for CKD status prediction.

Gazi Mohammed Ifraz,1Muhammad Hasnath Rashid, Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods. A number of physiological variables, as well as ML techniques such as logistic regression (LR), decision tree (DT) classification, and -nearest neighbor (KNN), were used in this work to train three distinct models for reliable prediction. The LR classification method was found to be the most accurate in this role, with an accuracy of about 97 percent in this study. The dataset that was used in the creation of the technique was the CKD dataset, which was made available to the public.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Chronic kidney disease is a general term for heterogeneous disorders affecting kidney structure and function. We emphasise clinical practice guidelines, clinical trials, and areas of uncertainty. It does not provide accurate prediction . Present a narrative review of the prevalence and incidence of diabetes-related kidney disease worldwide. Mortality among those with diabetes and kidney disease will also be explored. Chronic Kidney Disease also recognized as Chronic Renal Disease, is an uncharacteristic functioning of kidney or a failure of renal function expanding over a period of months or years. It requires more inputs to predict the results. To classify patients with chronic kidney disease by using support vector machines algorithm is investigated. The chronic kidney disease dataset is based on clinical history, physical examinations, and laboratory tests. The data set which has to be provide clearly otherwise it does not works. Machine learning can be hope in this problem it is best in prediction and analysis. Requires more fields to be filled for accurate results. If CKD is not detected and cured in early stage then patient can show following Symptoms: Blood Pressure, anaemia, weekboans, poor nutrition health and nerve damage, Decreased immune response because at advanced stages dangerous levels of fluids, electrolytes, and wastes can build up in your blood and body. Hence it is essential to detect CKD at its early stage but it is unpredictable as its Symptoms develop slowly and aren't specific to the disease. Some people have no symptoms at all so machine learning can be helpful in this problem to predict that the patient has CKD or not. Machine learning does it by using old CKD patient data to train predicting model.

3.2 PROPOSED SYSTEM

The proposed system "Early Chronic Kidney Disease Prediction using Machine Learning" will be a web-based platform that provides personalized food and exercise recommendations to individuals at risk of developing CKD. The system will analyze patient data, including medical history, lab results, and demographic information, to identify early warning signs of CKD. The machine learning algorithms will be trained on a large dataset of patient records to accurately predict the risk of developing CKD in individuals. In addition to the CKD prediction system and food recommendation system, the proposed system will also include an exercise recommendation system. The exercise recommendation system will provide personalized exercise plans based on the patient's health status and physical activity needs. The system will consider the patient's CKD risk score, body mass index, and physical activity level to suggest exercises that are beneficial for kidney health and overall physical fitness. Based on this information, the system will provide personalized meal plans, grocery lists, and exercise plans. The proposed system has the potential to significantly improve outcomes for patients with CKD by enabling early detection and intervention, providing personalized dietary recommendations to prevent disease progression, and promoting physical activity to improve overall health and fitness.

3.3 MOTIVATION

“Prevention is Better than Cure!”. This is based mostly on the use of machine learning process in the sector of chronic kidney disease prediction. This prediction is based on the detection using automated techniques is advantage, because it helps to prevent the people from the disease and also helps to recover by suggesting the healthy food and exercise effectively.

3.4 PROBLEM STATEMENT

Chronic kidney disease (CKD) is a serious and common health condition that affects millions of people worldwide, with significant morbidity, mortality, and economic burden on healthcare systems. Early detection and intervention are critical for managing CKD and improving patient outcomes, but traditional screening methods may not always be effective in identifying patients at risk for CKD. The goal of this research is to develop an ML-based system for early CKD detection that can accurately predict CKD risk based on clinical and demographic information, along with dietary and exercise habits. The current problem is the lack of an effective and efficient method for early CKD detection and personalized management. The proposed ML-based system with a food and exercise recommendation system has the potential to address this problem by enabling healthcare professionals to identify CKD risk factors early and provide tailored dietary and exercise advice to manage the disease. The proposed system also addresses the problem of patient non-compliance with lifestyle modifications. By providing personalized and relevant dietary and exercise recommendations based on the patient's individual needs and preferences, the system can potentially increase patient engagement and adherence to lifestyle modifications, leading to better outcome, the problem statement of this research is the need for an effective and efficient method for early CKD detection and personalized management, which the proposed ML-based system with a food and exercise recommendation system can potentially address by enabling early detection, personalized management, and increased patient engagement and adherence.

CHAPTER 4

SYSTEM REQUIREMENTS

4.1 HARDWARE REQUIREMENTS:

System : 32-bit or 64-bit operating system (x-64 based processor)

Ram : 4 GB

Hard drive : 500GB

4.2 SOFTWARE REQUIREMENTS:

Operating System: Windows 7,8,10

Language : Python, HTML, CSS, Javascript, bootstrap, flask.

Version : Python 3.8.16

IDE : Visual Studio Code

CHAPTER 5

RESEARCH METHODOLOGY

Chronic kidney disease are leads to kidney failure etc…It provides the prediction model to predict the disease and take subsequent measures using machine learning. Machine learning techniques will be used to identify whether the person is affected or not. It will analyze the performance of our trained model after it has been trained. The result will be provided by predicting the disease using machine learning algorithms and recommends healthy food and exercise.

5.1 DATASET

There are different types of test results are taken in dataset. The following tests are included in the dataset: age, blood pressure, specific gravity, albumin, sugar, red blood cell, pus cell, bacteria, blood glucose level, blood urea level, serum creatinine, sodium, potassium etc….

1	Age
2	Blood pressure
3	Sugar
4	Albumin
5	Red blood cell
6	Pus cell
7	Pus cell clumps
8	Specific gravity
9	bacteria (present , not present)
10	Bgr - blood glucose random in mgs/dl
11	Bu - blood urea in mgs/dl
12	Sc - serum creatinine mgs/dl
13	sod - sodium in mEq/L
14	pot - potassium in mEq/L
15	Hemo - hemoglobin in gms
16	Pcv - packed cell volume
17	wc - white blood cell count in cells/cumm
18	rc - red blood cell count in millions/cmm
19	htn - hypertension (yes or no)
20	dm - diabetes mellitus
21	cad - coronary artery disease (yes or no)
22	appet - appetite (yes or no)
23	pe - pedal edema (yes or no)
24	ane - anemia (yes or no)
25	class - classification (ckd , not ckd)

FIG 5.1 : Dataset

5.2 PROCESS FLOW

To implement an early chronic kidney disease (CKD) prediction system using machine learning (ML) with a food recommendation system, the following steps can be taken:

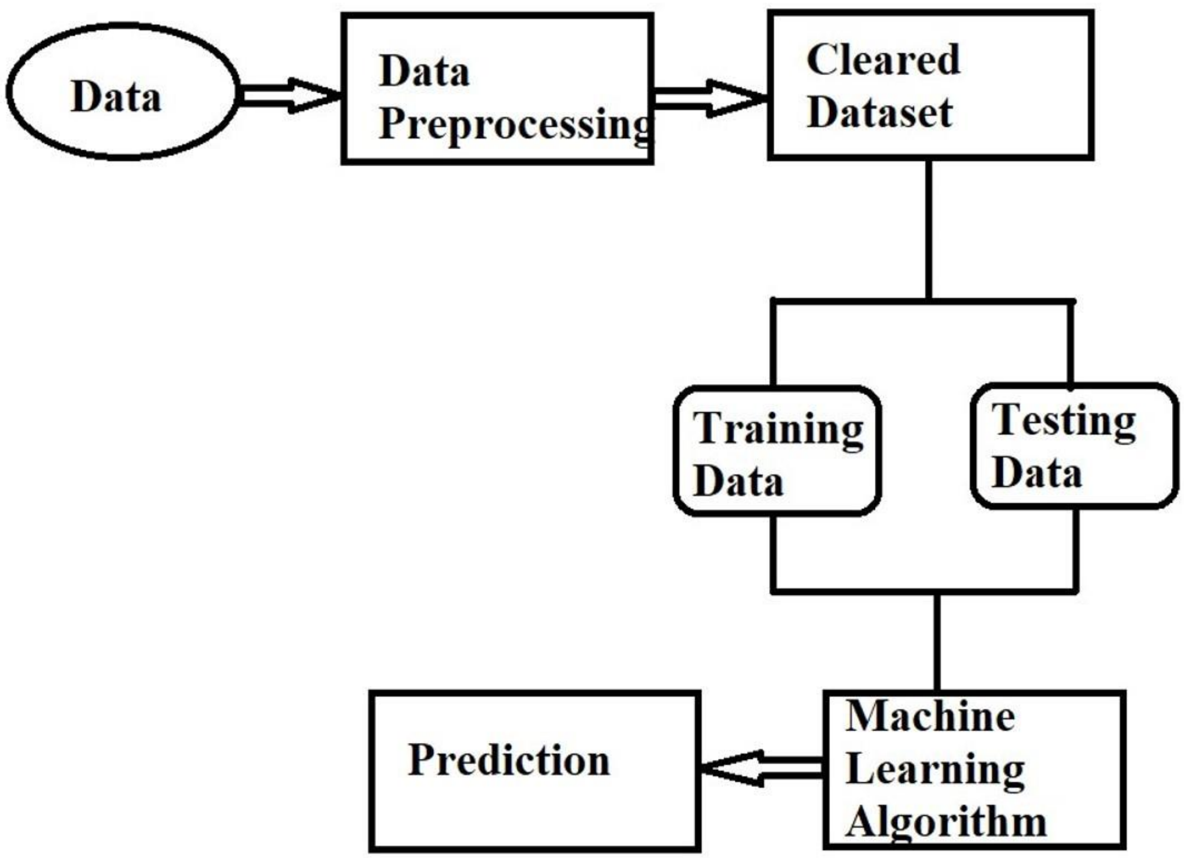

FIG 5.2 : Process flow

Data Collection: Gather clinical and demographic data from electronic health records, medical claims, and patient surveys. This data will include information such as age, sex, race, blood pressure, laboratory tests, and dietary habits.

Data Preprocessing: Clean and preprocess the data to remove any inconsistencies, errors, or missing values.

Feature Engineering: Extract relevant features from the preprocessed data, such as blood pressure and laboratory test values, and create new features if necessary.

ML Model Development: Train ML classifiers, such as decision tree and logistic regression models, on the preprocessed data to predict the likelihood of developing CKD. Also, train a food recommendation system using ML techniques to provide personalized dietary advice.

Model Evaluation: Evaluate the performance of the ML models using metrics such as accuracy, precision, recall, and F1 score.

Deployment: Integrate the CKD prediction and food recommendation systems into the healthcare system, allowing clinicians to access the system and make informed decisions about patient care.

Overall, an ML-based CKD prediction system with a food recommendation system can provide personalized care to patients, helping them to manage their kidney disease and maintain a healthy lifestyle.

5.3 SYSTEM ARCHITECTURE

System architecture diagrams provide a visual illustration of a system's various components and show how they communicate and interact with each other.

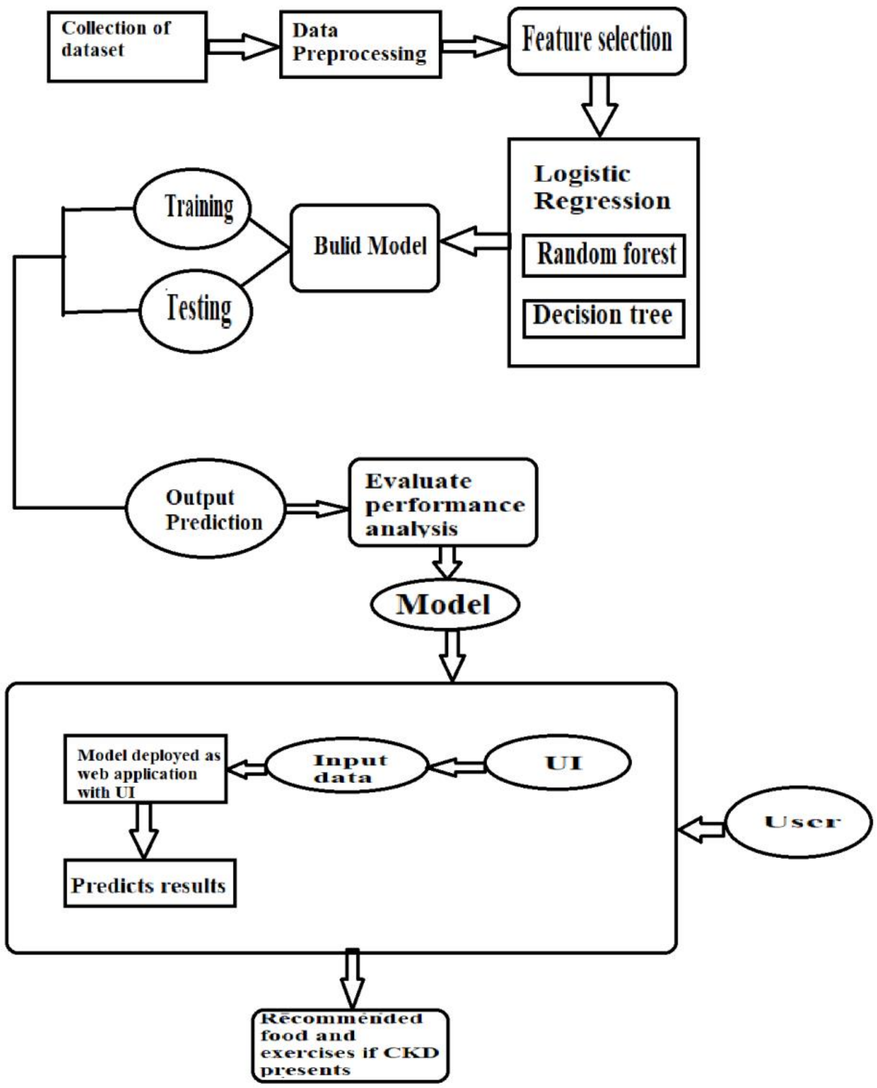

FIG 5.3 : Architecture of system

5.4 MODEL TRAINING

Model training is the phase in the data science development lifecycle where practitioners try to fit the best combination of weights and bias to a machine learning algorithm to minimize a loss function over the prediction range.

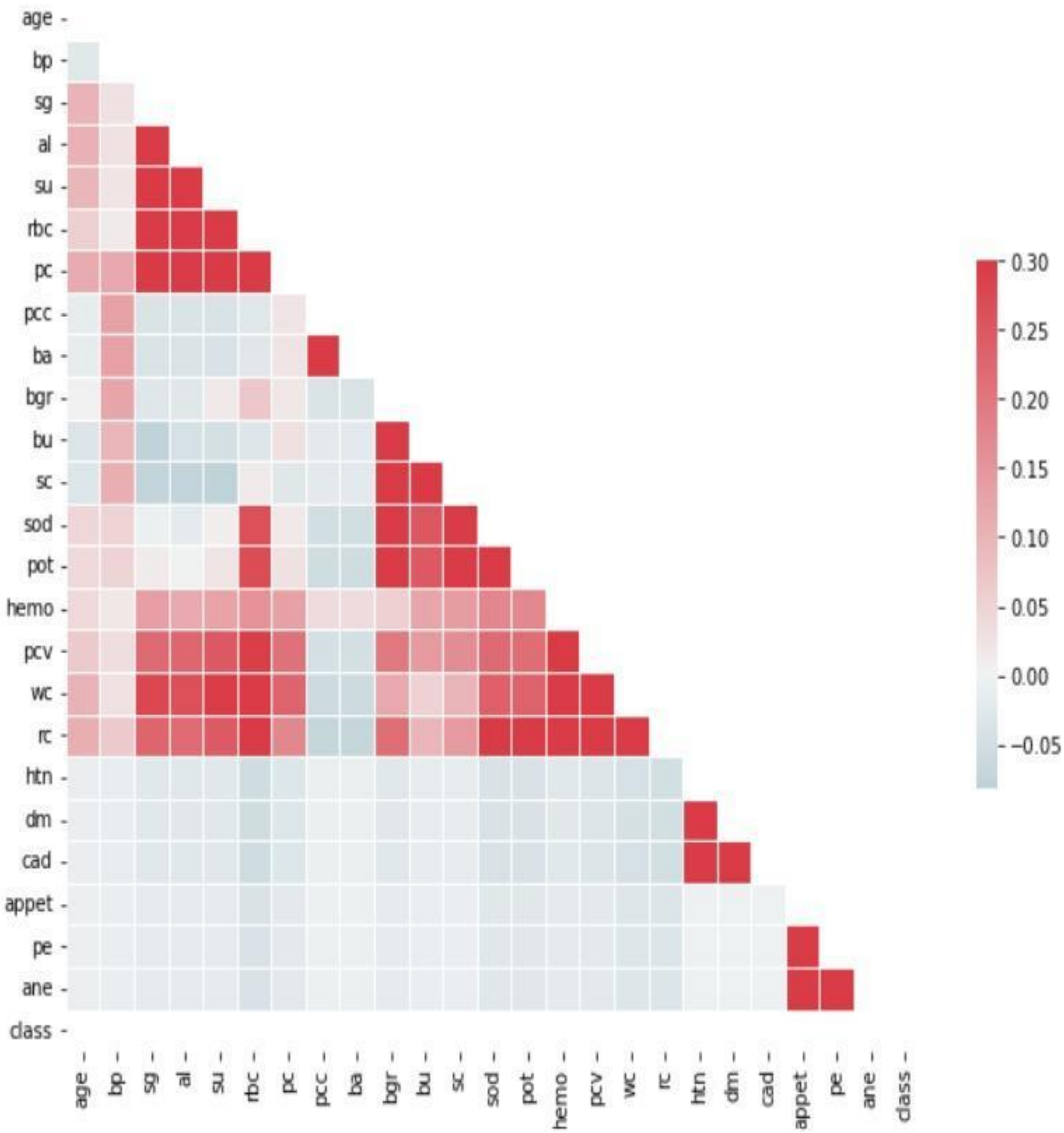

FIG 5.4 : Correlation between dataset

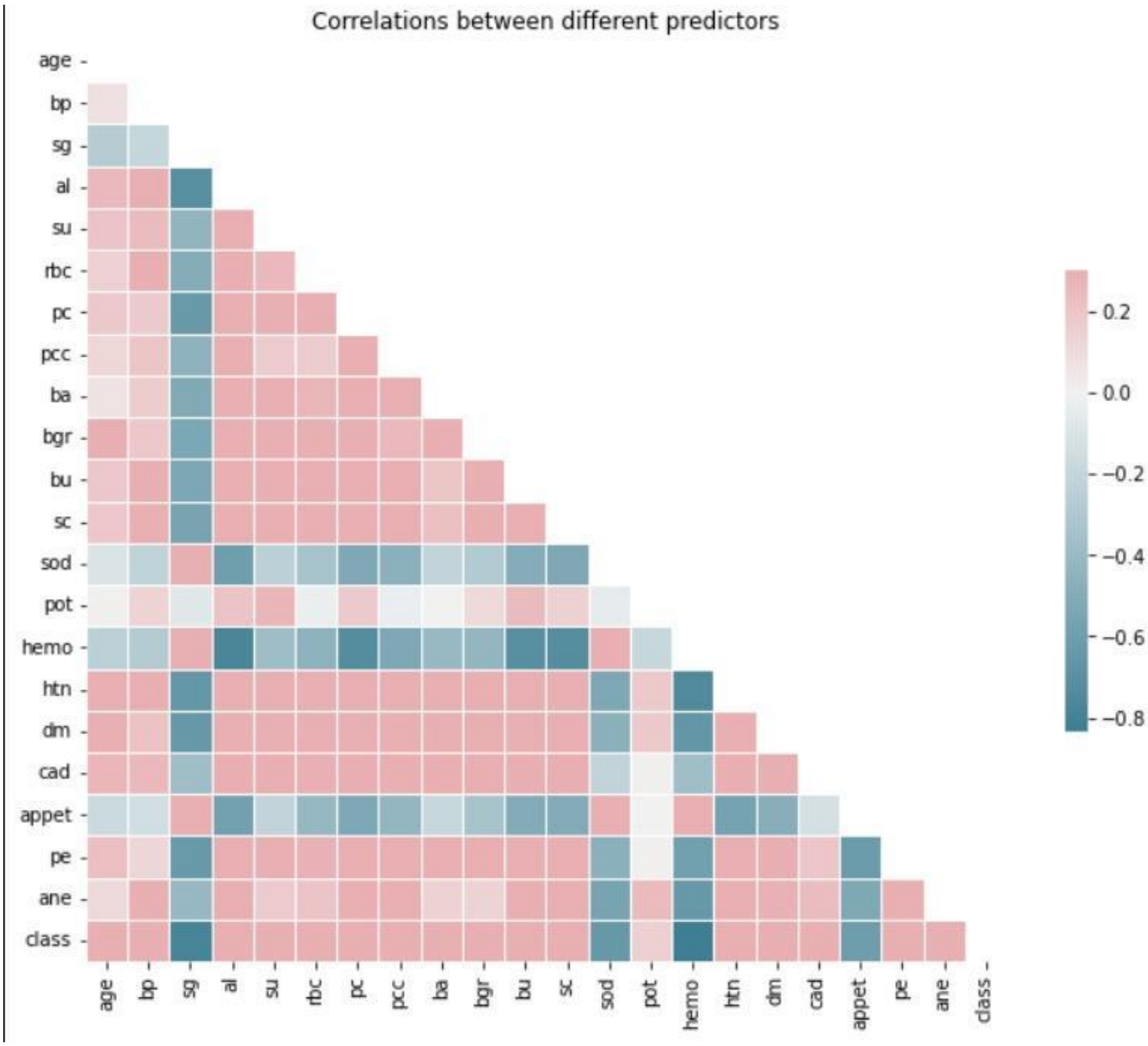

FIG 5.5 : Correlation between predictors

5.5 MODEL TESTING

This visually represents the accuracy rate of the model predicted for testing the dataset.

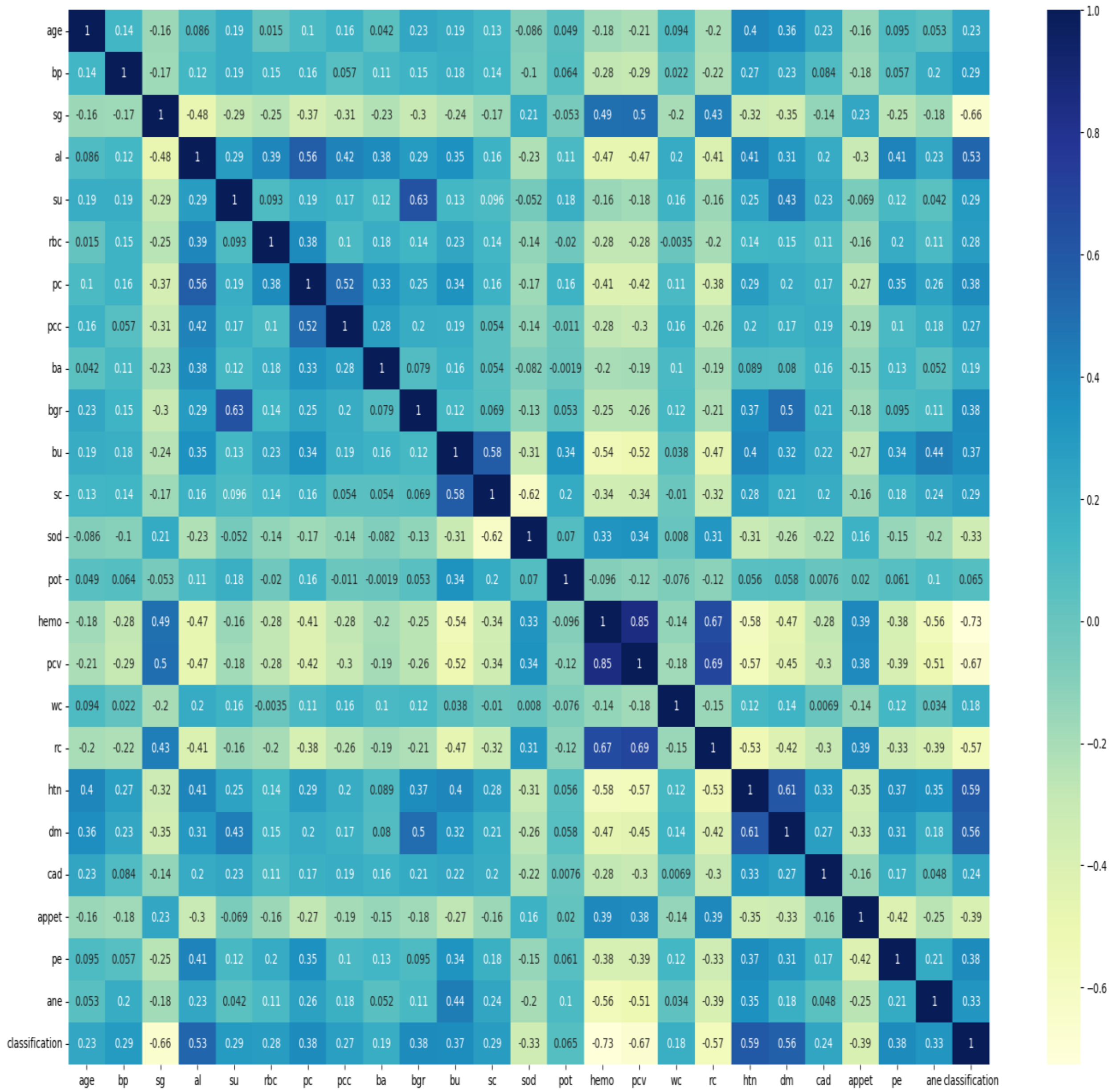

FIG 5.6 : Testing the Model

5.6 FOOD AND EXERCISE RECOMMENDATION

The Early Chronic Kidney Disease Prediction System is an ML-based system that uses various algorithms to predict the likelihood of developing CKD. This system combines ML classifiers, a food recommendation system, and an exercise recommendation system to provide personalized care to patients. The system collects and preprocesses clinical and demographic data, extracts relevant features, trains ML models, evaluates their performance, and deploys them into the healthcare system. The food recommendation system provides tailored dietary advice to patients, taking into account their health condition and dietary habits. Similarly, the exercise recommendation system suggests physical activities based on the patient's health condition and personal preferences. By providing personalized care, this system can improve patient outcomes and reduce the burden of CKD on patients and healthcare providers. The system is based on the principle that early detection and intervention are crucial in managing CKD. Traditional screening methods may not always be effective, and the ML-based approach can improve the accuracy of CKD prediction. The food recommendation system and exercise recommendation system complement the CKD prediction system by providing patients with actionable advice to manage their condition and maintain a healthy lifestyle. Overall, the Early Chronic Kidney Disease Prediction System with a food recommendation system and exercise recommendation is a valuable tool in managing CKD. The system's ability to provide personalized care can improve patient outcomes and reduce the burden on healthcare providers. By integrating ML technology with personalized care, this system can revolutionize the management of CKD and improve the lives of millions of patients worldwide.

5.7 RESULT AND DISCUSSION

5.7.1 OUTPUT

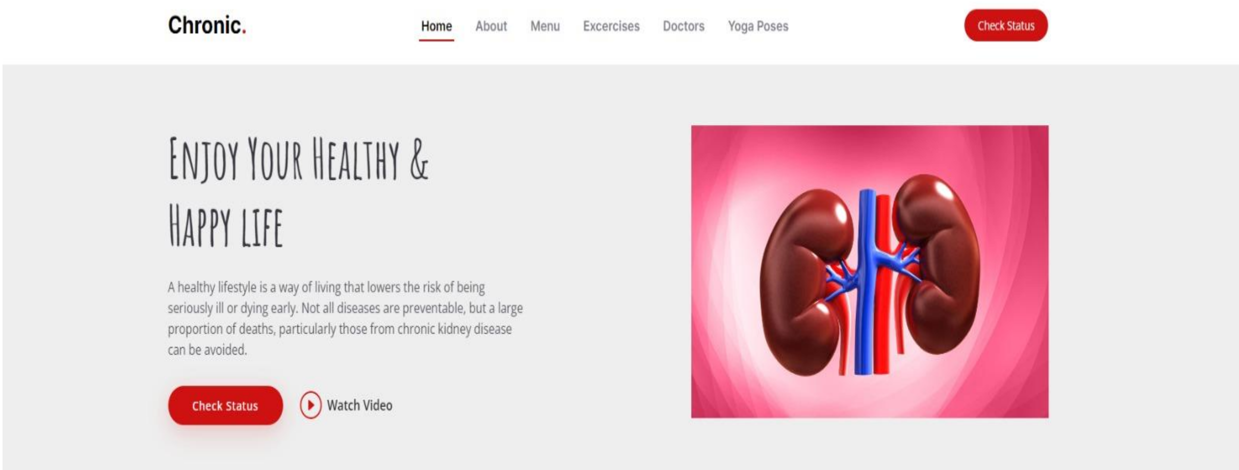

FIG 5.7 UI Interface

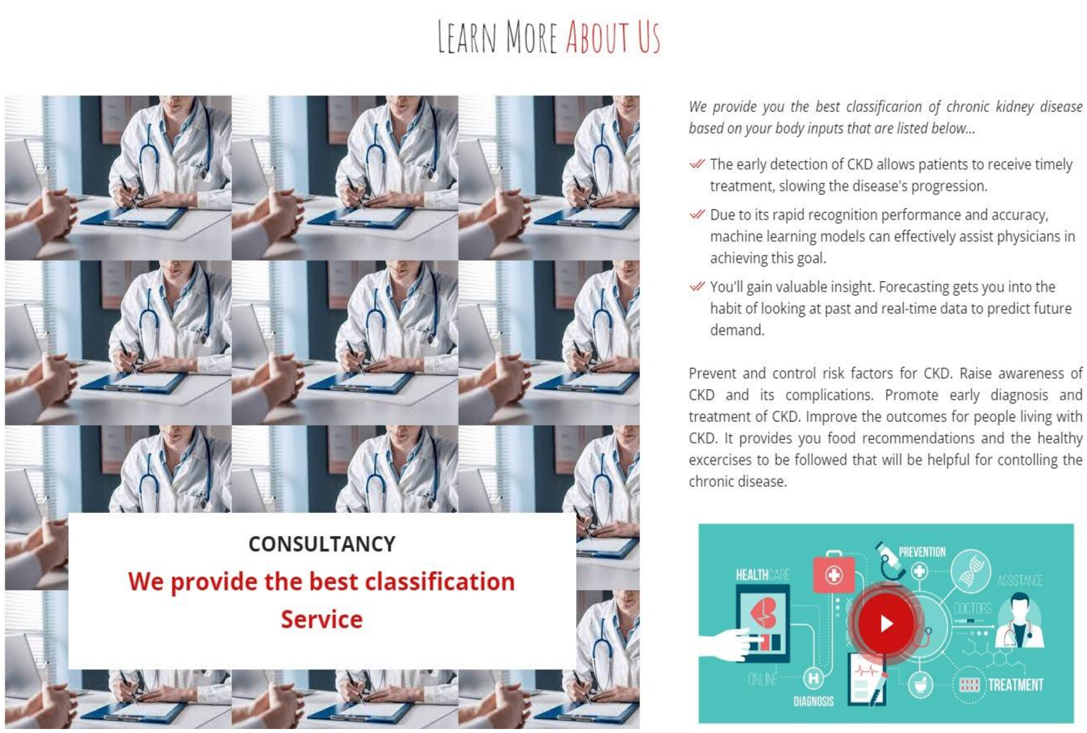

FIG 5.8 Classification about Disease

FIG 5.9 Patient details as per year

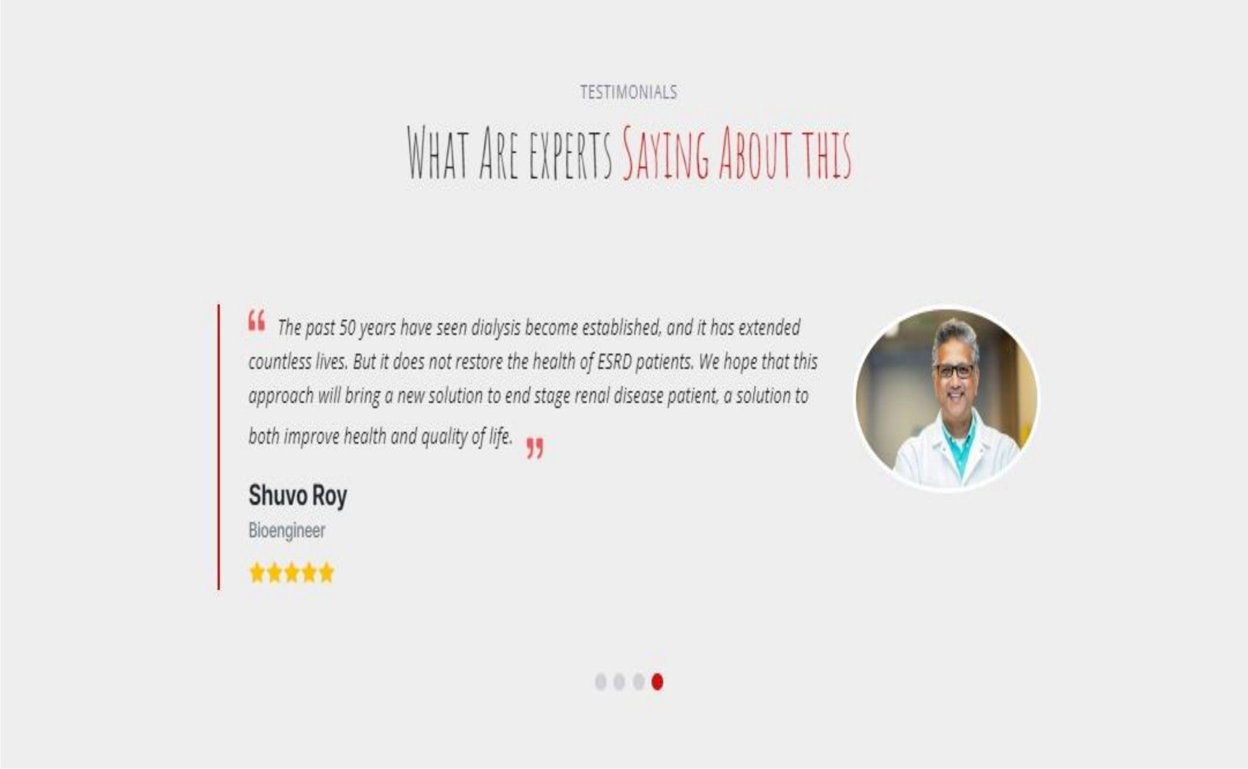

FIG 5.10 Expert suggestions

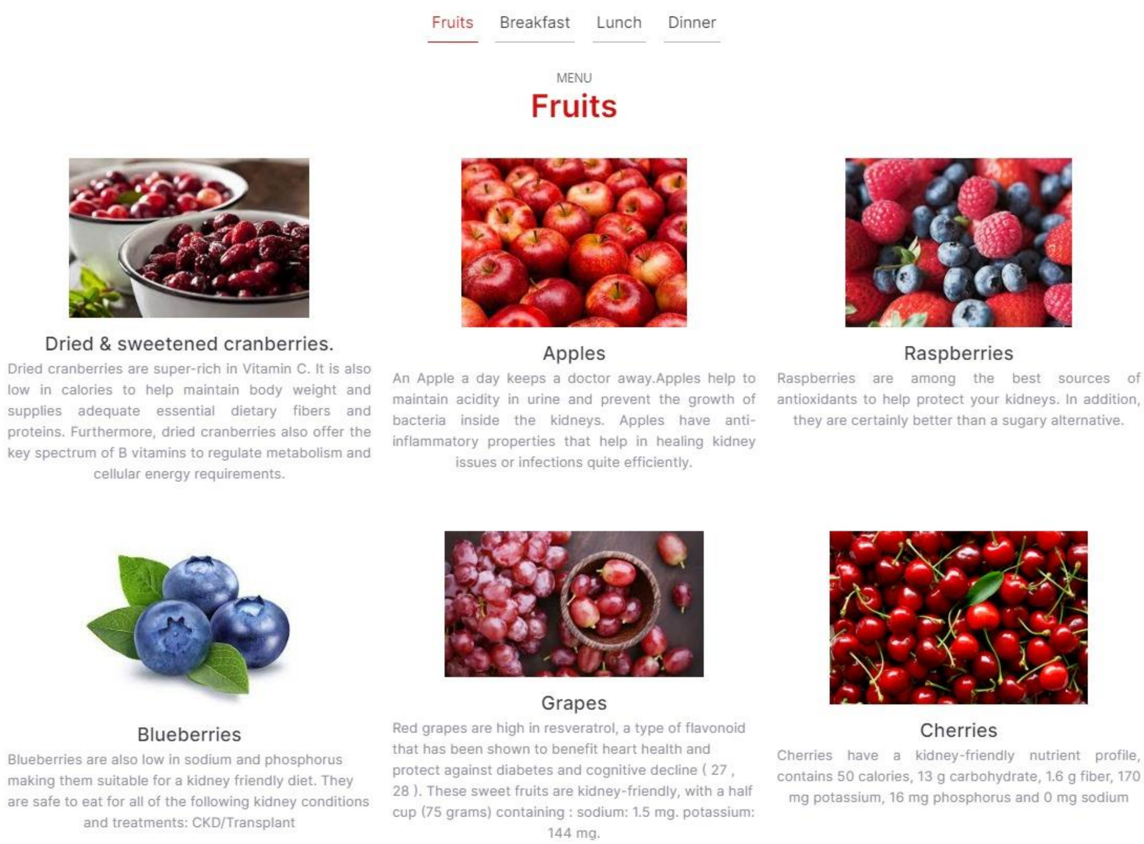

FIG 5.11 Recommended Foods

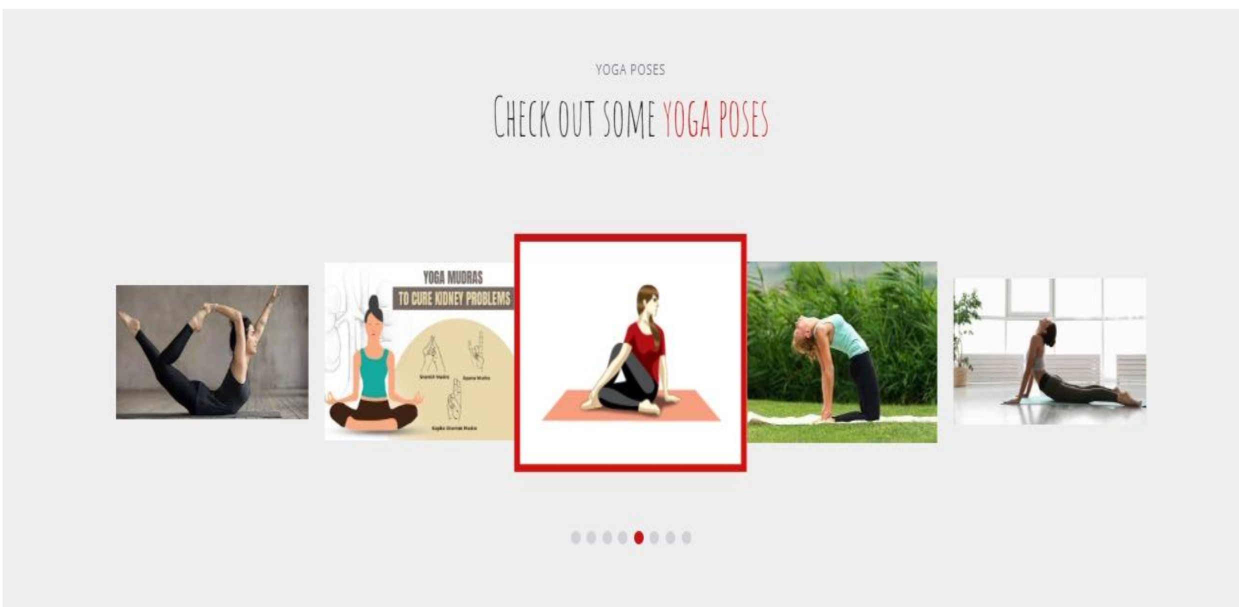

FIG 5.12 Recommended Exercises

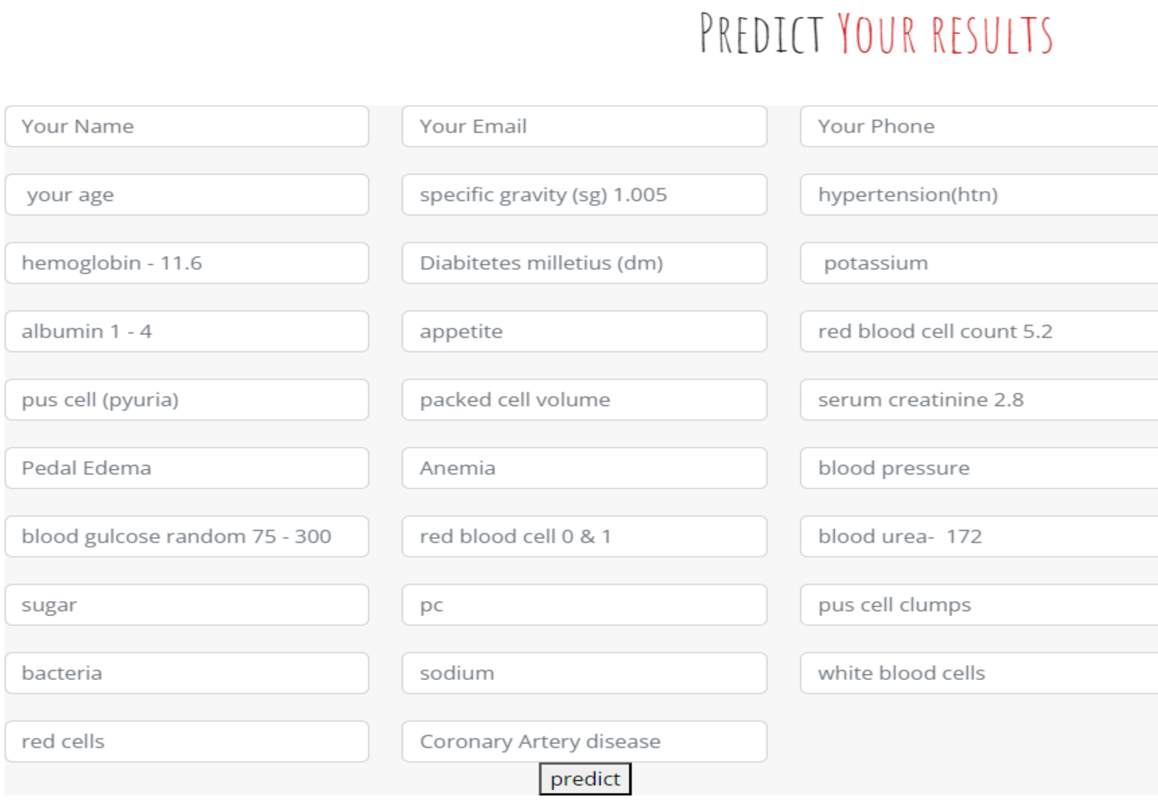

FIG 5.12 Prediction

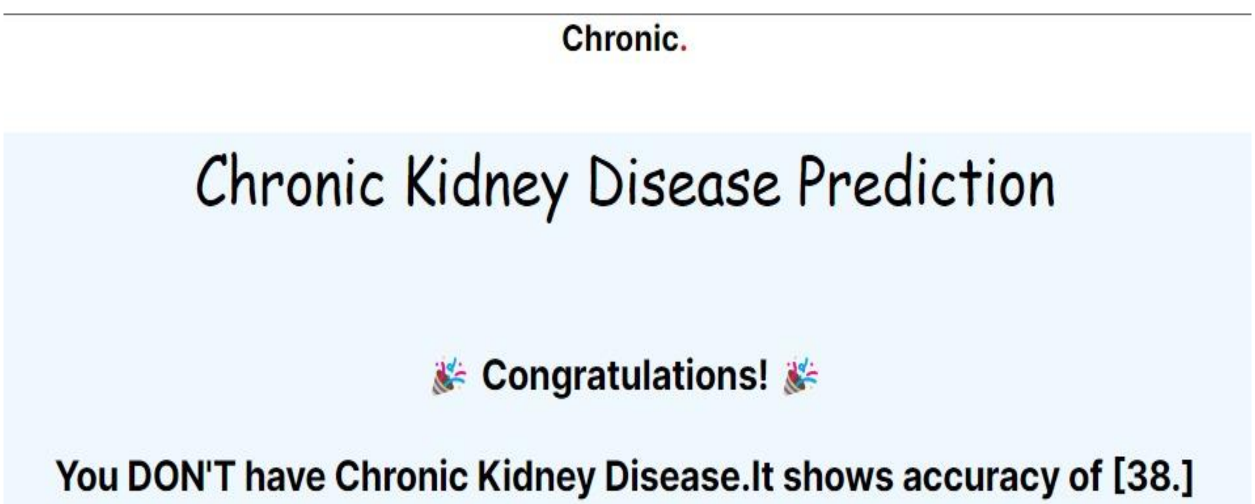

FIG 5.13 Negative results

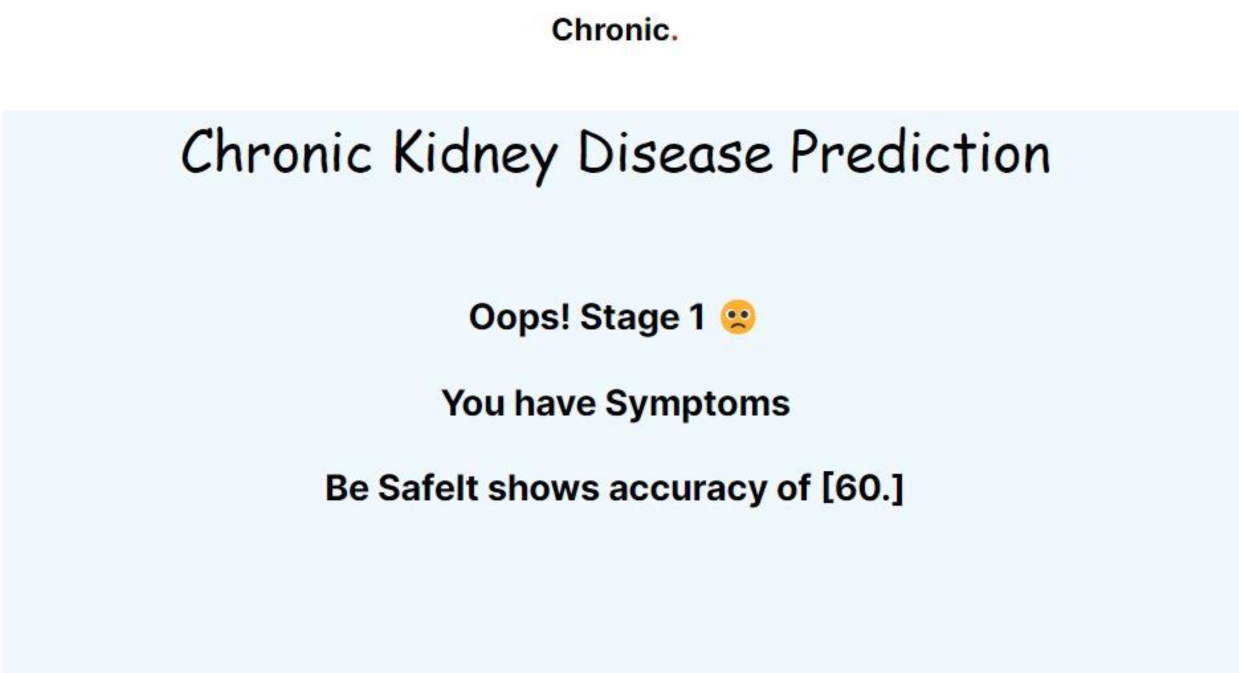

FIG 5.14 First Stage of Disease

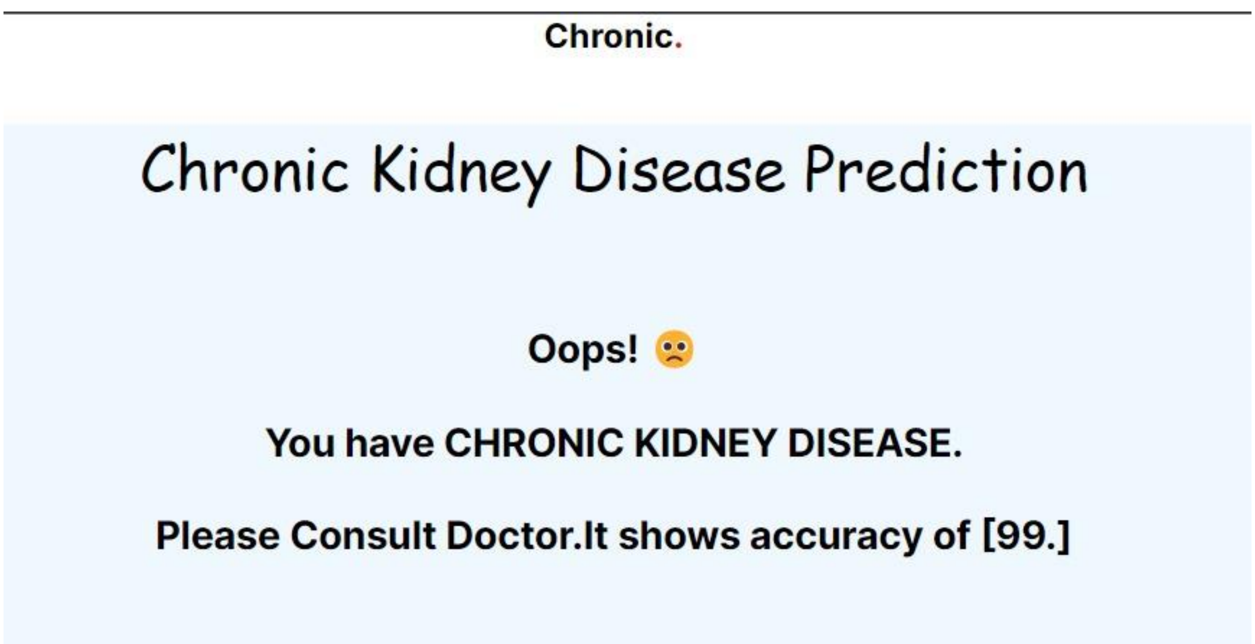

FIG 5.15 Positive Results

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

In conclusion, chronic kidney disease (CKD) is a serious health condition that requires proper management and prevention. A balanced and healthy diet is a key component of preventing and managing CKD, as it can help control blood pressure, blood sugar, and cholesterol levels, and reduce the risk of complications. A food recommendation system that takes into account the individual's health status, dietary preferences, and cultural background can be an effective tool for improving dietary adherence and outcomes in CKD patients. Such a system can provide personalized meal plans, recipes, and nutritional guidance to help CKD patients make informed and healthy food choices. However, it is important to note that a food recommendation system should be used as a supplement to, not a replacement for, medical advice and care from qualified healthcare professionals. Together with medical management and lifestyle modifications, a food recommendation system can contribute to better CKD outcomes and overall health.

The following are the future scope of the project,

One potential enhancement to the proposed ML-based system for early CKD detection is the integration of image processing techniques to analyze kidney images and identify early signs of kidney damage. This would involve incorporating machine learning algorithms that can accurately classify kidney images into normal or damaged kidney tissue. The integration of image processing would allow for a more comprehensive and accurate assessment of CKD risk. Furthermore, incorporating genetic data analysis into the system could provide a more complete understanding of a patient's risk for developing CKD. Genetic markers have been identified that are associated with an increased risk of developing CKD, and incorporating this information into the ML models could enhance the accuracy of the system's predictions. Another potential future enhancement is the incorporation of wearable technology to track physical activity and other health metrics. Wearable devices can provide continuous monitoring of patient activity levels, sleep quality, heart rate, and other vital signs. Integrating this data with the ML models could provide a more comprehensive assessment of a patient's risk for CKD and enable personalized recommendations for physical activity and exercise.Finally, integrating the system with electronic health records (EHRs) could further enhance its effectiveness. By analyzing data from EHRs, the system could potentially identify high-risk patients who have not yet been diagnosed with CKD, enabling early intervention and management. Additionally, the integration of EHRs could improve the accuracy of the system's predictions by incorporating a broader range of patient data.

CHAPTER 7

REFERENCES

[1] Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods Yedilkhan Amirgaliyev;Shahriar Shamiluulu;Azamat Serek 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT).

[2] Chronic Kidney Disease Detection using AdaBoosting Ensemble Method and K-Fold Cross Validation N. Mohana Suganthi;Jemin V.M;P. Rama;E. Chandralekha 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS).

[3] Chronic Kidney Disease Detection using AdaBoosting Ensemble Method and K-Fold Cross Validation N. Mohana Suganthi;Jemin V.M;P. Rama;E. Chandralekha 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS).

[4] Early Discovery of Chronic Kidney Disease by Attributing Missing Values S. Madhavi;Pathmanaban. J;Sangeetha. S;Riya K S 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI).

[5] Intelligent Systems for Diagnosis of Chronic Kidney Disease – A Review Arvind Sharma;Dalwinder Singh 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).

[6] Intelligent Systems for Diagnosis of Chronic Kidney Disease – A Review Arvind Sharma;Dalwinder Singh 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).

[7] Multilevel Ensemble Method to Identify Risks in Chronic Kidney Disease Using Hybrid Synthetic Data,Karamsetty Shouryadhar;P Kiran Rao;Subarna Chatterjee, 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT).

[8] Optimal Feature Selection for Chronic Kidney Disease Classification using Deep Learning Classifier K. Shankar;P. Manickam;G. Devika;M. Ilayaraja 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC).

[9] Predicting the Chronic Kidney Disease using Various Classifiers Pramila Arulanthu;Eswaran Perumal 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT).

[10] Predicting Chronic Kidney Disease by Applying Feature Engineering & Performance Analysis of Machine Learning Classifiers R. Praveen Kumar;Sarath Erive;Ganta Jayasri;Annapalli Srujana;Vallala Niharika 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA).

[11] Survey on Diagnosis of Chronic Kidney Disease UsingMachine Learning Algorithms A. Vijayalakshmi;V. Sumalatha 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS).

[12] Texture analysis of ultrasound images of chronic kidney disease Fadil Iqbal;Aruna S. Pallewatte;Janaka P. Wansapura 2017 Seventeenth International Conference on Advances in ICT for Emerging Regions (ICTer).

[13] Texture analysis of ultrasound images of chronic kidney disease Fadil Iqbal;Aruna S. Pallewatte;Janaka P. Wansapura 2017 Seventeenth International Conference on Advances in ICT for Emerging Regions (ICTer)

APPENDIX

A1 : SOURCE CODE

SAMPLE CODE

```
import numpy as np
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd
pd.pandas.set_option('display.max_columns', None)
dataset = pd.read_csv("kidney_disease.csv")
dataset.head()
dataset = dataset.drop('id', axis=1)
dataset.shape
dataset.isnull().sum()
dataset.describe()
dataset.columns
dataset.dtypes
dataset['rbc'].value_counts()
dataset['rbc'] = dataset['rbc'].replace(to_replace =
{'normal' : 0, 'abnormal' : 1})
dataset['pc'].value_counts()
dataset['pc'] = dataset['pc'].replace(to_replace =
{'normal' : 0, 'abnormal' : 1})
dataset['pcc'].value_counts()
```

```
dataset['pcc'] = dataset['pcc'].replace(to_replace =
{'notpresent':0,'present':1})
dataset['ba'].value_counts()
dataset['ba'] = dataset['ba'].replace(to_replace =
{'notpresent':0,'present':1})
dataset['htn'].value_counts()
dataset['htn'] = dataset['htn'].replace(to_replace =
{'yes' : 1, 'no' : 0})
dataset['dm'].value_counts()
dataset['dm'] = dataset['dm'].replace(to_replace =
{'\tyes':'yes', ' yes':'yes', '\tno':'no'})
dataset['dm'] = dataset['dm'].replace(to_replace =
{'yes' : 1, 'no' : 0})dataset['cad'].value_counts()
dataset['cad'] = dataset['cad'].replace(to_replace =
{'\tno':'no'})
dataset['cad'] = dataset['cad'].replace(to_replace =
{'yes' : 1, 'no' : 0})
dataset['appet'].unique()
dataset['appet'] =
dataset['appet'].replace(to_replace={'good':1,'poor':0,
'no':np.nan})
dataset['pe'].value_counts()
dataset['pe'] = dataset['pe'].replace(to_replace =
{'yes' : 1, 'no' : 0})
dataset['ane'].value_counts()
dataset['ane'] = dataset['ane'].replace(to_replace =
{'yes' : 1, 'no' : 0})
```

```
dataset['classification'].value_counts()
dataset['classification']                                     =
dataset['classification'].replace(to_replace={'ckd\t':'
ckd'})
dataset["classification"] = [1 if i == "ckd" else 0 for
i in dataset["classification"]]
dataset.dtypes
dataset['pcv']      =      pd.to_numeric(dataset['pcv'],
errors='coerce')
dataset['wc']       =       pd.to_numeric(dataset['wc'],
errors='coerce')
dataset['rc']       =       pd.to_numeric(dataset['rc'],
errors='coerce')
dataset.dtypes
dataset.describe()
dataset.isnull().sum().sort_values(ascending=False)
dataset.isnull().any().sum()
plt.figure(figsize=(24,14))
sns.heatmap(dataset.corr(), annot=True, cmap='YlGnBu')
plt.show()
dataset.drop('pcv', axis=1, inplace=True)
sns.countplot(dataset['classification'])
X = dataset.iloc[:, :-1]
y = dataset.iloc[:, -1]
from sklearn.ensemble import ExtraTreesClassifier
import matplotlib.pyplot as plt489+
model=ExtraTreesClassifier()
```

```
model.fit(X,y)
plt.figure(figsize=(8,6))
ranked_features=pd.Series(model.feature_importances_,in
dex=X.columns)
ranked_features.nlargest(24).plot(kind='barh')
plt.show()
X = dataset[['sg', 'htn', 'hemo', 'dm', 'al', 'appet',
'rc', 'pc']]
X.head()
X.tail()
y.head()
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test  =  train_test_split(X,y,
test_size=0.3, random_state=33)
print(X_train.shape)
print(X_test.shape)
from     sklearn.metrics     import     accuracy_score,
confusion_matrix, classification_report
import seaborn as sns
from sklearn.metrics import classification_report
from sklearn import metrics
from sklearn.ensemble import RandomForestClassifier
RandomForest = RandomForestClassifier()
RandomForest = RandomForest.fit(X_train,y_train)
y_pred = RandomForest.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
```

```
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('RF')
from sklearn.ensemble import AdaBoostClassifier
AdaBoost = AdaBoostClassifier()
AdaBoost = AdaBoost.fit(X_train,y_train)
y_pred = AdaBoost.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('ADB')
from sklearn import tree
tree=tree.DecisionTreeClassifier()
dtree=tree.fit(X_train,y_train)
y_pred = dtree.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('DB')
from sklearn.ensemble import GradientBoostingClassifier
GradientBoost = GradientBoostingClassifier()
GradientBoost = GradientBoost.fit(X_train,y_train)
y_pred = GradientBoost.predict(X_test)
```

```
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('GB')
from sklearn.linear_model import LogisticRegression
LogisticRegression=LogisticRegression()
LogisticRegression=LogisticRegression.fit(X_train,y_tra
in)
y_pred = LogisticRegression.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('LR')
from sklearn.naive_bayes import GaussianNB
GaussianNB=GaussianNB()
GaussianNB=GaussianNB.fit(X_train,y_train)
y_pred = GaussianNB.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('NB')
```

```
from sklearn.neighbors import KNeighborsClassifier
knn= KNeighborsClassifier(n_neighbors = 5, metric =
'minkowski', p = 2)
knn=knn.fit(X_train, y_train)
y_pred = knn.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('KNN')
from sklearn.svm import SVC
svm= SVC(kernel = 'linear', random_state = 0)
svm=svm.fit(X_train, y_train)
y_pred = svm.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('SVM')
from sklearn.svm import SVC
ksvm = SVC(kernel = 'rbf', random_state = 0)
ksvm=ksvm.fit(X_train, y_train)
y_pred = ksvm.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
```

```
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('kSVM')
from sklearn import tree
tree=tree.DecisionTreeClassifier()
dtree=tree.fit(X_train,y_train)
y_pred = dtree.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
x = metrics.accuracy_score(y_test, y_pred)
acc.append(x)
model.append('DB')
plt.figure(figsize=[10,5],dpi = 100)
plt.title('Accuracy Comparison')
plt.xlabel('Accuracy')
plt.ylabel('Algorithm')
sns.barplot(x = acc,y = model,palette='dark')
import pickle
with open("Kidney_model.pkl","wb") as file:
pickle.dump(dtree,file)
n_estimators = [int(x) for x in np.linspace(start=100,
stop=2000, num=18)]
max_features = ['auto', 'sqrt']
max_depth = [int(x) for x in np.linspace(5, 30, num=6)]
min_samples_split = [2, 5, 10, 15, 20,25]
```

```
min_samples_leaf = [1, 2, 5, 10,12,15]
params = {
    'n_estimators': n_estimators,
    'max_features': max_features,
    'max_depth': max_depth,
    'min_samples_leaf': min_samples_split,
    'min_samples_leaf': min_samples_leaf
}
from  sklearn.model_selection import RandomizedSearchCV
rf = RandomForestClassifier()
tuned_model = RandomizedSearchCV(rf, params,
cv=5, n_iter=20, random_state=43, n_jobs=-1)
tuned_model.fit(X_train, y_train)
print(tuned_model.best_estimator_)
RF         =         RandomForestClassifier(max_depth=15,
max_features='auto', min_samples_leaf=2,
n_estimators=435)
RF = RF.fit(X_train,y_train)
y_pred = RandomForest.predict(X_test)
print('Accuracy:', accuracy_score(y_test,y_pred)*100)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
```

A2: CONFERENCE CERTIFICATE

KGiSL INSTITUTE OF TECHNOLOGY

Affiliated to Anna University, Chennai, Approved by AICTE, New Delhi, Reaccredited by NBA
Coimbatore, Tamil Nadu.

NCEIICT'23

CERTIFICATE

OF APPRECIATION

HARI PRIYA C

has presented a paper titled Early Chronic kidney disease prediction using ML and actively participated in the 2nd National Conference on Emerging Innovations in Information and Communication Technology (NCEIICT'23) organized by the Department of Information Technology,. KGiSL Institute of Technology on May 5th 2023.

Dr. Sankar Ram N
NCEIICT'23 - Chair

Dr Shankar P
Director, Academics.

KGiSL INSTITUTE OF TECHNOLOGY

Affiliated to Anna University, Chennai, Approved by AICTE, New Delhi, Reaccredited by NBA
Coimbatore, Tamil Nadu.

NCEIICT'23

CERTIFICATE

OF APPRECIATION

SRIRANGANATH M

has presented a paper titled Early chronic kidney disease prediction using ML and actively participated in the 2nd National Conference on Emerging Innovations in Information and Communication Technology (NCEIICT'23) organized by the Department of Information Technology,. KGiSL Institute of Technology on May 5th 2023.

Dr. Sankar Ram N
NCEIICT'23 - Chair

Dr Shankar P
Director, Academics.

KGiSL INSTITUTE OF TECHNOLOGY

Affiliated to Anna University, Chennai, Approved by AICTE, New Delhi, Reaccredited by NBA
Coimbatore, Tamil Nadu.

NCEIICT'23

CERTIFICATE

OF APPRECIATION

VINITHA V

has presented a paper titled Early chronic kidney disease prediction using ML and actively participated in the 2nd National Conference on Emerging Innovations in Information and Communication Technology (NCEIICT'23) organized by the Department of Information Technology,. KGiSL Institute of Technology on May 5th 2023.

Dr. Sankar Ram N
NCEIICT'23 - Chair

Dr Shankar P
Director, Academics.

KGiSL INSTITUTE OF TECHNOLOGY

Affiliated to Anna University, Chennai, Approved by AICTE, New Delhi, Reaccredited by NBA
Coimbatore, Tamil Nadu.

NCEIICT'23

CERTIFICATE

OF APPRECIATION

VINOTHINI J

has presented a paper titled Early chronic kidney disease prediction using ML and actively participated in the 2nd National Conference on Emerging Innovations in Information and Communication Technology (NCEIICT'23) organized by the Department of Information Technology,. KGiSL Institute of Technology on May 5th 2023.

Dr. Sankar Ram N
NCEIICT'23 - Chair

Dr Shankar P
Director, Academics.