CHRONIC KIDNEY DISEASE AND KIDNEY STONE PREDICTION USING SML TECNIQUE

A PROJECT REPORT

Submitted by

DEEPA . S	**211419205035**
GRACE NESAM . R	**211419205058**
PAVITHRAA.S	**211419205124**

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

PANIMALAR ENGINEERING COLLEGE,PONAMALLE

ANNA UNIVERSITY : CHENNAI 600 025

APRIL 2023

BONAFIDE CERTIFICATE

Certified that this project report "CHRONIC KIDNEY DISEASE AND KIDNEY STONE PREDICTION USING SML TECHNIQUE" is the bonafide work of **"DEEPA.S (211419205035),GRACE NESAM.R (211419205058),PAVITHRAA.S (211419205124)"** who carried out the project under my supervision.

SIGNATURE

Dr. M. Helda Mercy M.E,Ph.D.,
HEAD OF DEPARTMENT
Department of Information Technology
Panimalar Engineering College
Poonamalle,Chennai-600123

SIGNATURE

Dr. M. Sumithra M.E,Ph.D.,
ASSOCIATE PROFESSOR
Department of Information Technology
Panimalar Engineerimg College
Poonamalle,Chennai – 600123

Submitted for the project and viva-voce examination held on _____

SIGNATURE

INTERNAL EXAMINER

SIGNATURE

EXTERNAL EXAMINER

DECLARATION

I hereby declare that the project report entitled "**CHRONIC KIDNEY DISEASE AND STONE PREDICTION USING SML TECHNIQUE"** which is being submitted in partial fulfilment of the requirement of the course leading to the award of the Bachelor Of Technology in Information Technology in **Panimalar Engineering College, Autonomous institution Affiliated to Anna university-Chennai** is the result of the project carried out by me under the guidance of **DR.M.SUMITHRA M.E,Ph.D., Associate Professor in the Department of Information Technology**. I further declared that I or any other person has not previously submitted this project report to any other institution/university for any other degree/ diploma or any other person.

Date**:** **DEEPA.S**

Place**:** Chennai **GRACE NESAM.R**

PAVITHRAA.S

It is certified that this project has been prepared and submitted under my guidance.

DATE: DR. M . SUMITHRA

PLACE: (Associate Professor/IT)

ACKNOWLEDGEMENT

A project of this magnitude and nature requires kind co-operation and supportfrom many, for successful completion. We wish to express our sincere thanks to allthose who were involved in the completion of this project.

Our sincere thanks to **Our Honorable Secretary and Correspondent, Dr. P. CHINNADURAI, M.A., Ph.D.,** for his sincere endeavor in educating us in his premier institution.

We would like to express our deep gratitude to **Our Dynamic Directors, Mrs. C. VIJAYA RAJESHWARI and Dr. C. SAKTHI KUMAR, M.E., Ph.D., Dr. SARANYA SREE SAKTHIKUMAR, B.E., MBA., Ph.D.,** for providing us with the necessary facilities for completion of this project.

We also express our appreciation and gratefulness to **Our Principal Dr. K. MANI, M.E., Ph.D.,** who helped us in the completion of the project. We wish to convey our thanks and gratitude to our **Head of the Department**, **Dr.M. HELDA MERCY, M.E., Ph.D.,** Department of Information Technology, for her support and by providing us ample time to complete our project.

We express our in debtedness and gratitude to our Project co-ordinator **Mr.M.DILLI BABU**, **M.E.,(Ph.D.,)** Associate Professor, Department of Information Technology for his guidance throughout the course of our project. We also express sincere thanks to our supervisor **Dr.M. SUMITHRA, M.E,Ph.D.,** Department of Information Technology for providing the support to carry out the project successfully. Last, we thank our parents and friends for providing their extensive moral support and encouragement during the course of the project.

ABSTRACT

The phrase Chronic kidney disease refers to kidney damage that lasts for a long time and may worsen over time. The kidneys may stop functioning if the damage is severe. This is referred to a end-stage renal disease or kidney failure (ESRD). Patients with kidney disease may enter the chronic phase. This CKD is characterized by a gradual decline in kidney function. To determine whether kidney disease is severe or not, a variety of algorithms are used in this project. By predicting the kidney stages the severity is determined. Additionally the presence of kidney stones are determined. In this project, the modules are also deployed which are developed in order to determine the appropriate accuracy of the disease. In this project, Supervised machine learning techniques were used to find the results. The algorithms that are used in this project are Naïve Bayes, Logistic regression, KNN (K-nearest neighbors) and Decision tree. In this project, the first step is data processing then data visualization is carried out and the prediction of various algorithms are carried in order to provide with the best accuracy result. The algorithm which provides the best accuracy will be considered forthe further process in providing the final output to the user. The prediction doesn't depends on each and every attributes separately, it just takes the average of the values for each and every attribute. The average values determine the stages of chronic kidney disease such as 0,1,2 and along with the detection of the presence of kidney stones are also provided to the user as a final result.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

CKD	Chronic Kidney Disease
NCKD	No Chronic Kidney Disease
SOD	Sodium
POT	Potassium
HEMO	Hemoglobin
RBC	Red Blood Cell
WBC	White Blood Cell
BP	Blood Pressure
SC	Serum Creatinine

CHAPTER – 1

INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

Using machine learning, we choose the subject of knowledge set to improve diagnosis of chronic kidney disease (CKD) and kidney stones. We are all aware that CKD is one of the most common diseases among individuals nowadays. By gathering a large amount of CKD and stone data. We are working hard to improve detection in people. It usually affects the Adults above the age of 30 and continues indefinitely. Males between the ages of 30 and 35 have stage 5 CKD. without a transplant, one could expect to live for 14 years. For women of the same age, 13 years are predicted to be their lifespan. Everybody has a four-year life expectancy between the ages of 70 and 75, regardless of gender. GFR Rate is a popular tool for identifying the majority of renal disorders (Glomerular filtration rate). the following three stages of chronic kidney disease: Stage 1 with normal or high GFR (GFR > 90 mL/min) Stage 2 Mild CKD (GFR= 60-89 mL/min) Stage 3A Moderate CKD (GFR = 45-59 mL/min) Stage 3B Moderate CKD (GFR = 30-44mL/min). We should not smoke. Restrict the alcohol consumption. Maintain a blood pressure of less than 140/90 mm Hg (or the target set by your doctor). If we have diabetes, we should try to maintain our blood sugar levels within the ideal range. We should get vaccinated for the flu during flu season. They are crucial in the fight against kidney stones and CKD.

1.2 NEED OF THE SYSTEM

As the kidney disease cases are increasing and need for the system also increases. When the patient goes for diagnosis first, He/she has to take an blood test then they gets the blood test report . The report generation can take days due to the demand and as the number of patients are also increasing. After the report is generated sometimes it takes time to visit the doctor so it may lead to delay in knowing the result whether your kidney is healthy or not , so to avoid that problem this project helps the patients or common people to get the idea whetherthey have kidney disease or not by entering their values in the website accordingto the report with that we will be able to find whether kidney disease is present ornot and can also able to find the stages of their disease without consulting the doctor . After knowing the result if they come know that they have severe kidneydisease they will be to know to consult doctor without delaying any further. The main problem is that the report generation can delay as it is a tedious process anddue the increase in patients and this can be prevented with our system as it detectsthe stages of kidney disease and whether

kidney stone is present or not. It also states the required treatment 3 type depending upon the stage. It reduces the waiting time, till the patient gets officialconfirmation from the doctor it can be used a reference which helps them get prepared mentally. By using this system, a patient can avoid anxiety during a duration for wait of result.

1.3 OBJECTIVE OF THE PROJECT

The main objective of the system is to reduce the time period for the patients to know the result whether they have chronic kidney or not and if there what is theirstage and can also know whether kidney stone is present or not. Blood test will given within 1-2 days. Our objective is to detect stages of chronic kidney diseasein an early stage by just giving values that they get after blood test report like albumin, sodium, potassium, creatinine value as input. In output the stages are specified for the input breast values given. By this we can confirm whether the kidney is healthy or sick. This causes the patient to suffer as treatment should bestarted as fast as they can. This gap can be avoided using our system. This reducesthe anxiety of the patients as they get to know what they have been diagnosed. It helps the patient to get mentally prepared before the official report comes.

1.4 SCOPE OF THE PROJECT

As the cases for chronic liver disease increases, the duration of diagnosis also increases. The delay in diagnosis may result in the loss of patient life. As there are many patients are diagnosed with kidney disease the report generation gets delayed. This creates panic among the patients. The scope of this project is to investigate a dataset of hospital records for the medical sector using machine learning techniques. To identify a patient is affected with CKD or not. We can also make diagnoses as quickly as possible.

CHAPTER-2

LITERATURE SURVEY

2.1 TITLE: Chronic Kidney Disease And its Complications. (2008)

AUTHOR: Robert Thomas, MD, Abbas Kans, John R, Sedor

DESCRIPTION: When aberrant albumin excretion or impaired kidney function last for longer than three months, as determined by a measured or estimated glomerular filtration rate (GFR), CKD is present. According to the stage of the disease, treatments are suggested for CKD and dialysis patients. These treatments could lower these patients' morbidity and fatality rates. But it has disadvantage that is it may not been determined for sure, it is prudent to abideby FDA instructions.

2.2 TITLE: Attributable causes of chronic kidney disease in adults: a five-year retrospective study in a tertiary-care hospital in the northeast of the Malaysian Peninsula (2015)

AUTHOR: Muhammad SalmanI, Amer Hayat Khan, Azreen Syazril Adnan, Syed Azhar Syed Sulaiman, Khalid Hussain, Naureen Shehzadi, Fauziah Jumaat.

DESCRIPTION: They set out to describe the adult patients' demographics, clinical profile, and potentially causal factors for CKD at a tertiary-care hospital in Malaysia. There advantage is to study alerts the general public to the likelihoodthat putting more emphasis on diabetes and hypertension primary prevention will significantly reduce hospital admissions due to CKD in Malaysia. And the disadvantage is that the data might not paint a complete picture of the CKD- related causes and Certain data in this study were not completely available due tothe back data.

2.3 TITLE: Detection of Chronic Kidney Disease Using Machine Learning Algorithms with Least Number of Predictors. (2013)

AUTHOR: Marwa Almasoud , Tomas E Ward.

DESCRIPTION: In this study, we use the lowest subset of features to investigate how well machine learning algorithms predict chronic kidney disease. The ANOVA test, Pearson's correlation, and Cramer's V test are only a few examples of the statistical tests that have been performed to eliminate redundant features.

The pros of this paper are the classifiers have been trained, tested, and verified using 10-fold cross-validation. The gradient boosting approach improved performance in terms of F1-measure (99.1%), sensitivity (98.8%), and specificity. (99.3%).And the cons are they want to compare the results with another datasetthat has the same attributes or validate our results using a large dataset. Moreover,to aid in lowering the occurrence of CKD**.**

2.4 TITLE: Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network (2016)

AUTHOR: Fuzhe Ma ,Tao Sun , Lingyun Liu, Hongyu Jing

DESCRIPTION: Computer vision and machine learning are the fields that develop methods to extract meaningful meanings from digital images. To create the reference standard for segmentation from the validation and training datasets, an experienced radiologist was evaluated to divide the abdomen MR and CT scanpictures into the left and right half. Pros are the proposed HMANN approach helps to segment the kidney picture and lowers noise for precise placement of the kidney stone diagnosis. To successfully solve this issue, tested. CONS: Early kidney stone detection is essential because renal damage might endanger life. Toundertake surgery to remove a kidney stone, the location of the kidney must be determined.

2.5 TITLE: Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques (2021)

AUTHOR: Ebrahime Mohammed Senan

DESCRIPTION: The originality of this study is in creating a system for diagnosing chronic renal illnesses. This study supports specialists in studying preventive methods for CKD through early diagnosis utilizing machine learning approaches. This study's main objective was to assess a dataset made up of 400 patients and 24 attributes. The missing nominal and numerical data were replacedusing the mean and mode statistical analysis methods. Recursive Feature Elimination (RFE) was used to select the most crucial characteristics. PROS: Our systems' accuracy ranged from 100% with random forest to 97.3% with SVM. CONS: These papers don't accurately depict chronic renal disease.

2.6 TITLE: Prediction of Chronic Kidney Disease – A Machine Learning Perspective (2021)

AUTHOR: Pankaj Chittora; Sandeep Chaurasia; Prasun Chakrabarti

DESCRIPTION: This article has examined chronic kidney disease prediction from this angle. In this study, seven classifier methods were used. The results have been calculated for each classifier based on the following features: (i) full features; (ii) correlation-based feature selection; (iii) Wrapper method feature selection; (iv) least absolute shrinkage and selection operator regression; (v) synthetic minority over-sampling technique with least absolute shrinkage and selection operator regression selected features; and (vi) synthetic features. PROS: Again, the linear support vector machine provided the maximum accuracy of 98.46% in the synthetic minority over-sampling technique with the least absolute shrinkage and selection operator selected features. CONS: Logistic and KNN were not employed in SMOTE since they did not produce the desired results.

2.7 TITLE: Neural network and support vector machine for the prediction of chronic kidney disease (2019)

AUTHOR: Njoud Abdullah Almansour, Hajra Fahim Syed, Nuha RadwanKhayat

DESCRIPTION: Using a dataset of 400 patients and 24 variables linked to thediagnosis of chronic kidney disease, we concentrate on using several machine learning classification methods in this article. In this study, artificial neural networks (ANN) and support vector machines were employed as classification methods (SVM). PROS: One of the most well-known machine learning methods is the support vector machine (SVM), while another is the artificial neural network (ANN). Both methods are beneficial and have a history of producing outstanding results across a range of industries. In comparison to SVM, which has shown the highest accuracy in prior studies, ANN has been presented as a novel model to more accurately predict CKD. The missing values were initially replaced after pre processing the dataset. CONS: The results in this paper were not accurate.

CHAPTER-3

SYSTEM DESIGN

3.1 PROPOSED SYSTEM

The proposed strategy is to create a machine learning model for categorizing renaldisease stages and stones. The procedure begins with data gathering, in which previous information about kidney disease stages and stones is gathered. In the healthcare arena, data mining is a widely utilized technique for processing massive amounts of data. The stages of kidney illness and stones, if detected early enough, can save lives. Machine learning is increasingly widely employed in health care, where it minimizes manual labor and, with a better model, errors are reduced, potentially saving lives. The dataset is analyzed and accurate variable identification is performed, which means that both dependent variables and independent variables are discovered. Then appropriate machine learning methods are used which is applied to the dataset where the data pattern is discovered. Following the use of many algorithm, a better algorithm is utilized topredict the outcome.

ADVANTAGES:

1. To forecast the stages of kidney illness and stone formation, we areemploying machine learning techniques.
2. For improved prediction, algorithms are compared, and the best model isassessed.
3. Several algorithms' performance indicators are evaluated in order toprovide a more accurate prediction.
4. Excellent performance and accuracy.
5. We completed the deployment procedure and compared variousalgorithms to improve accuracy.

3.2 ARCHITECTURE DIAGRAM

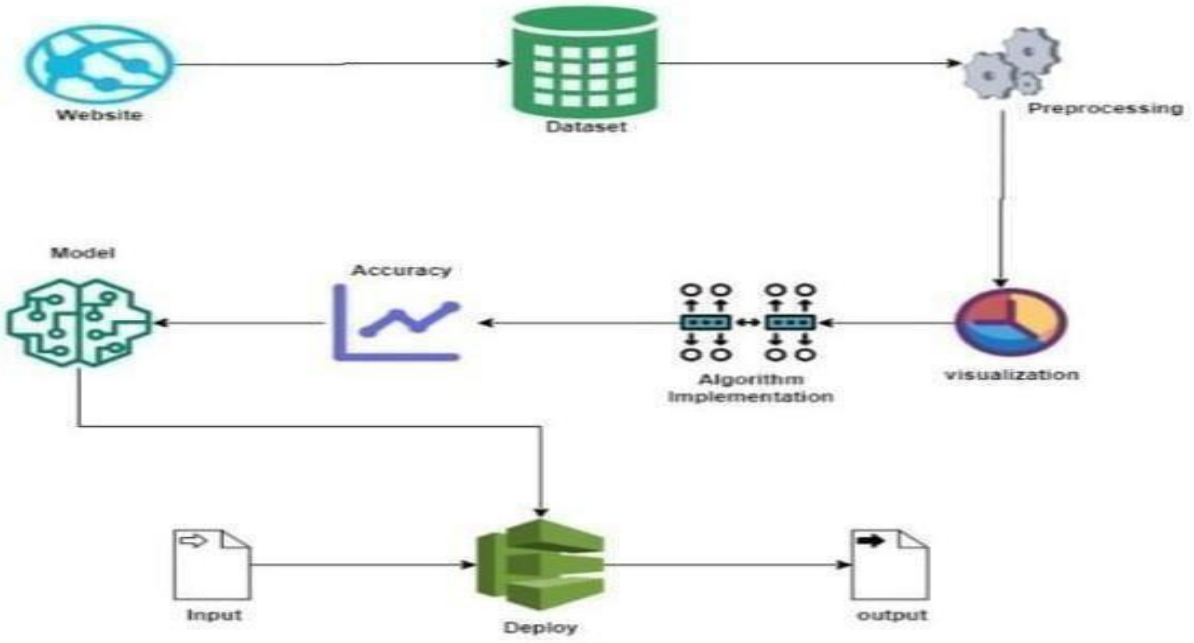


Fig 3.2.1 Architecture Diagram

DESCRIPTION: The basic Architecture diagram of the ‑CHRONIC KIDNEY DISEASE AND KIDNEY STONE PREDICTION USING SML TECNIQUE‖ is
shown in the above figure. Mainly this diagram consists of the following essentialblocks:

Pre processing
Visualization
Algorithm Implementation
Deploy

1.Data Preprocessing: Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the important process in machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data so to verify that andto get a clean dataset we do data pre-processing so after that we will be

able to obtain clean and formatted dataset and then we can proceed further with our project implementation.

- In this the input values will be in the form of raw dataset and after pre processing the output will be clean and formatted dataset.
- To get pre processed data first we will import libraries like pandas and numpy where pandas is used to analyze big data and to make conclusions and numpy is used for working with arrays. After importing the libraries we will link the datasetwith the notebook for processing with the help of df and read function. Then we use shape , size and column function to know more about the dataset for processing . isnull function is used to return a dataframe object where values are replaced with true for null values and false for non-null values. We use dropna function to remove rows which contain null values. And also we use groupby and duplicate function for aggregation , analysis and remove duplicate values.

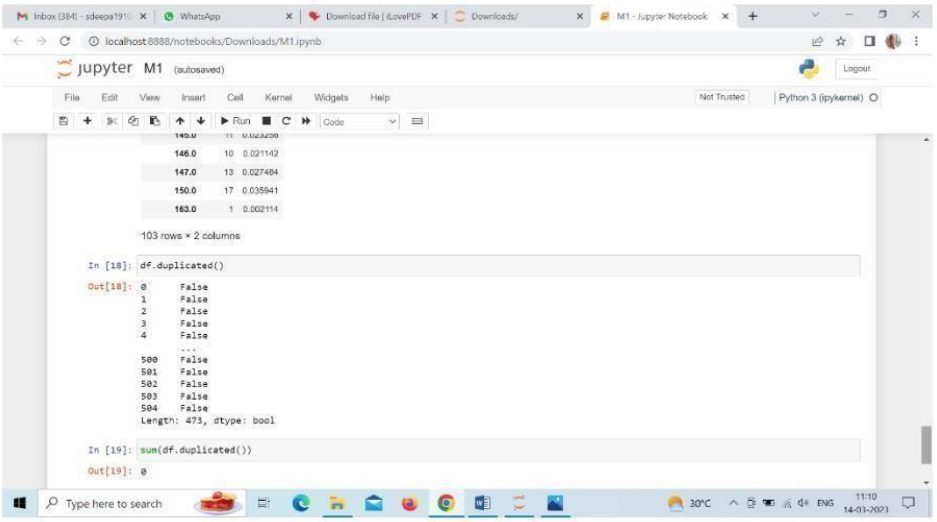


Fig 3.2.2 Sample screenshot for Data Pre-processing

2. **Data Visualization:** In this data visualization module the data's are displayed has visualized data using graphs and charts like histogram , pie-chart, bar graph etc.Data visualization is a valuable tool for gaining a qualitative understanding. This can help with identifying patterns, corrupt data, outliers, and other issues when exploring and getting to know a dataset.

- In this module the input data will be pre-processed data and output will be visualized data which means the data will be represented in the form of graph or plot.

- In this module also we will import libraries like NumPy and pandas but along with that we will also import libraries like seaborn and matplotlib where Seaborn uses Matplotlib underneath to plot graphs and It will be used tovisualize random distributions. And Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. In this modulewe use functionslike Pl. Subplot, plt.hist, plt.bar, plt.boxplot, sns.displot, sns.heatmap etc. to obtain the visualized data in the form of graph or plot.

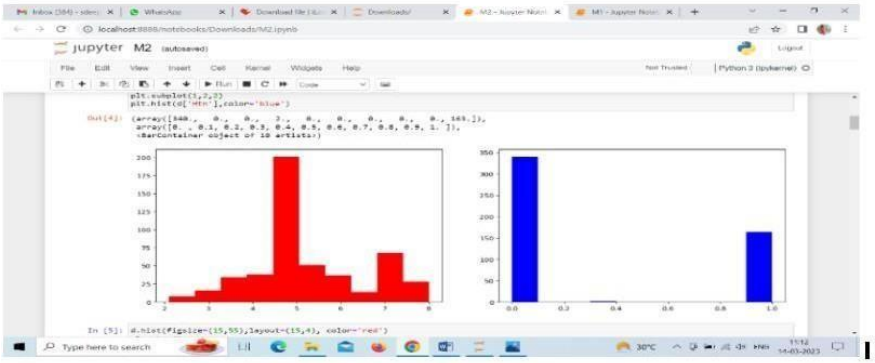


Fig 3.2.3 Sample screenshot for Data Visualization

3. ALGORITHM IMPLEMENTATION: - In this project we totally use 4 algorithms to predict the output they are:

1. DECISION TREE CLASSIFIER: It only poses a question and dividesthe tree into subtrees according to the response (Yes/No). There are 2 nodes they are decision node and leaf node. Decision node is used to make decision and leafnode is the outcome of such decisions.

2. LOGISTIC REGRESSION: It is used for predicting the categorical dependent variable using a given set of independent variables. Outcome will be either true or false, 1 or 0, yes or no. But instead of giving exact values it providesprobabilistic value.

3. KNN : It makes the assumption that new cases and existing cases are comparable and it places the new instance in the category that is most likely the existing system.

4. Naïve Bayes : It makes predictions using the probabilities of each attributebelonging each class.

- In this the input will be data and the output will be accuracy.
- Each and every algorithm is implemented separately to find accuracy then after finding , the best accuracy will be considered for predicting the output that is thestages of CKD and whether kidney stones is present or not.
- To find the accuracy for each and every algorithm first libraries are imported then dataset are linked, dropna function is performed. Then sklearn module is used to assess the quality of our prediction.
- In this data's are trained and tested they are x_train, x_test, y_train and y_test. Then confusion matrix is performed to find where errors is made in the module. We also perform cros_val_score function on dataset to test whether model can generalize over the whole dataset and cros_val_score returns the accuracy value. To display the confusion matrix in visual manner show and image interpolation functions are used. Finally the output will be displayed displayed in plot manner.
- Same procedure is followed for all the 4 algorithms to find the accuracy .

3. **DEPLOYMENT**: In this module the trained Machine learning model is convertedinto hierarchical data format file (.h5 file) which is then deployed in our Django framework for providing better user interface and predicting the stages of chronic kidney disease and whether kidney stone is present or not according to the data.

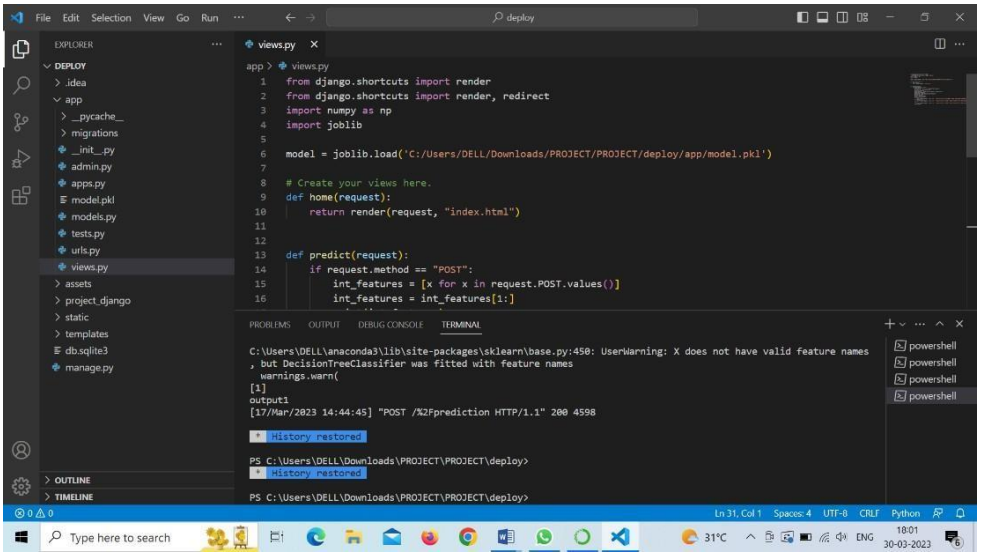


Fig 3.2.4 Sample screenshot for Deployment

CHAPTER - 4

REQUIREMENT SPECIFICATION

4.1 SOFTWARE REQUIREMENTS

4.1.1 Windows 10

4.1.2 Anaconda

4.1.3 Jupyter Notebook

4.1.4 Visual Studio

4.1.1 WINDOWS 10

Fig 4.1.1 Windows 10

- Windows is a graphical operating system developed by Microsoft.
- It allows users to view and store files, run the software, play games, watchvideos, and provides a way to connect to the internet.
- It was released for both home computing and professional works.
- Microsoft Windows 10 integrated support for multifactor authentication technologies, such as smartcards and tokens. In addition, Windows Hello brought biometric authentication to Windows 10, allowing users to log in with a fingerprint scan, iris scan or facial recognition technology.

4.1.2 ANACONDA

Fig 4.1.2 Anaconda

Anaconda is a distribution of the Python and R programming languages for scientific computing(data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012.As an Anaconda,Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition andAnaconda Enterprise Edition, both of which are not free.Package versions in Anaconda are managed by the package management system *conda*.This package manager was spun out as a separate open-source package as it ended up being useful on its own and for things other than Python.There is also a small, bootstrapversion of Anaconda called Miniconda, which includes only conda, Python, the packages they depend on, and a small number of other packages.

4.1.3 JUPYTER NOTEBOOK

Fig 4.1.3 Jupyter Notebook

Jupyter notebook is an open-source, interactive web application that allows users to create and share documents that contain interactive calculations, code, images, etc. Users can combine data, coe, and visualizations into a single notebook, and create interactive stories that they can edit and share. Notebooks are documents which contain both computer code (such as Python) and other text elements such as paragraph, markdown, figures, links, etc.The Jupyter notebook is widely used and well documented and offers an easy to use interface for creating, editing, andrunning notebooks. The notebook runs as a web application called the Dashboard or control panel that shows local files and allows users to open notebook documents and run snippets of code. The outputs are neatly formatted and displayed on the browser. The other component of the notebook is the kernel. The kernel is a –computational engine that executes the code written in the Notebook. It is similar to the back-end of the application. The IPython kernel (Jupyter was previously called IPython notebook) is used to execute Pythoncode in the Jupyter notebook. There are kernels for other languages as well, but in this article, we will explore running Python code in the notebook.

4.1.4 VISUAL STUDIO

Fig 4.1.4 Visual Studio

The Visual Studio IDE is a creative launching pad that you can use to edit, debug, and build code, and then publish an app. Over and above the standard editor and debugger that most IDEs provide, Visual Studio includes compilers, code completion tools, graphical designers, and many more features to enhance the software development process.

4.2 HARDWARE REQUIREMENTS

4.2.1 PROCESSOR: PENTIUM IV/III

4.2.2 HARD DISK: 80GB

4.2.1 PROCESSOR: PENTIUM IV/III

Fig 4.2.1 Processor: Pentium IV/III

- Pentium 4 was a series of single-core central processing units (CPU) for desktop PCs and laptops. The series was designed by Intel and launched in November 2000. Pentium 4 clock speeds were over 2.0 GHz.
- Intel shipped Pentium 4 processors until August 2008. Pentium 4 variants included code named Willamette, Northwood, Prescott and Cedar Mill with clock speeds that varied from 1.3-3.8 GHz.
- The Pentium 4 processor replaced the Pentium III via an embedded seventh- generation x86 microarchitecture, known as Net burst Microarchitecture, which was the first new chip architecture launched after the P6 microarchitecture in the1995 Pentium Pro CPU model.

4.2.2 HARD DISK: 80GB

Fig 4.2.2 Hard Disk:80 GB

Hard disk, also called hard disk drive or hard drive, magnetic storage medium for a computer. Hard disks are flat circular plates made of aluminum or glass and coated with a magnetic material. Hard disks for personal computers can store terabytes (trillions of bytes) of information. Data are stored on their surfaces in concentric tracks. A small electromagnet, called a magnetic head, writes a binary digit (1 or 0) by magnetizing tiny spots on the spinning disk in different directionsand reads digits by detecting the magnetization direction of the spots. A computer's hard drive is a device consisting of several hard disks, read/write heads, a drive motor to spin the disks, and a small amount of circuitry, all sealedin a metal case to protect the disks from dust. In addition to referring to the disksthemselves, the term hard disk is also used to refer to the whole of a computer's internal data storage. Beginning in the early 21st century, some personal computers and laptops were produced that used solid-state drives (SSDs) that relied on flash memory chips instead of hard disks to store information.

CHAPTER - 5

IMPLEMENTATION

5.1 FLOW DIAGRAM

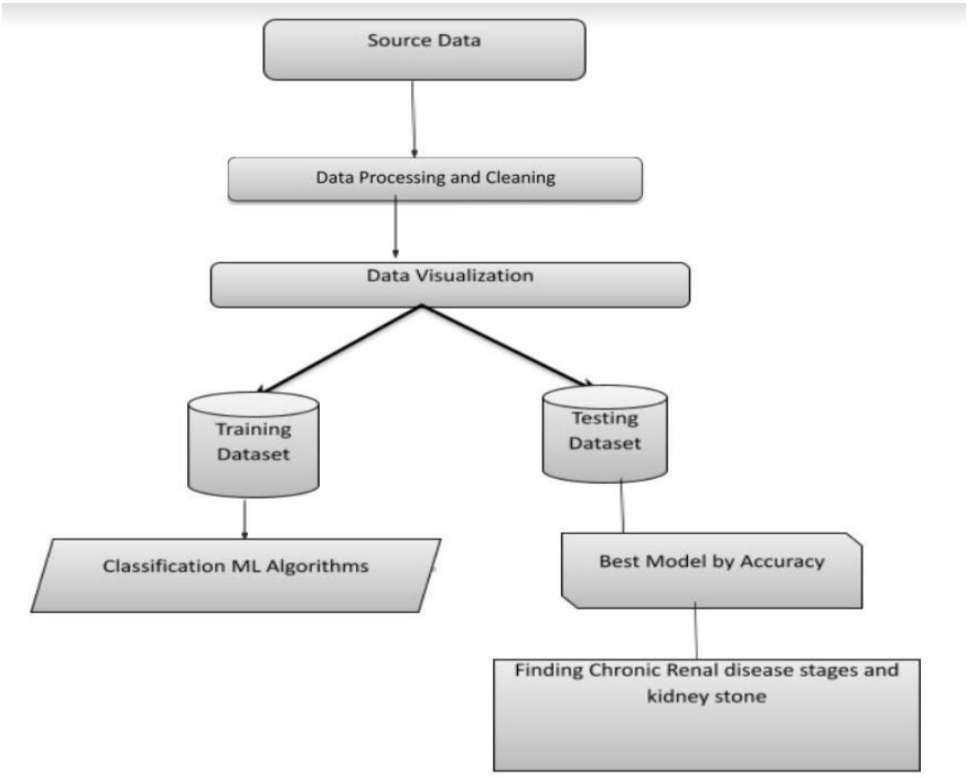


Fig 5.1 Flow Diagram

5.2 WORKING

According to the blood test report, the values for each attributes willbe entered in the website which we have given. Then the User will click the "PREDICT" button . After then the process will be carried out in the backend to produce the results according to the data which the user entered. In that the first process will be "DATA PREPROCESSING". After Data Preprocessing, Data visualization will be performed for better understanding of the data. After Data visualization, We are separating the dataset into training dataset and testing dataset. The output of the training dataset will be classification of ML algorithms.So with MLalgorithms, we would test the dataset for the accuracies of different algorithms. After the accuracy report of each algorithms, we consider the best accuracy algorithm to find the stages and kidney stone according to the values entered by the user. For finding the stages and kidney stones, the average of attributes values are considered to predict the result as such like Class 0, Class 1, Class 2.

5.3CODE

MODULE 1

```
import pandas as p
import numpy as n
import warnings
warnings.filterwarnings('ignore')
df=p.read_csv('A.csv')
df.head()
df.tail()
df.shape
df.size
df.columns
df.isnull()
df['Class'].unique()
df = df.dropna()
df.describe()
df.corr()
df.info()
p.crosstab(df["Wbcc"], df["Rbcc"])
df.groupby(["Hemo","Htn"]).groups
df["Class"].value_counts()
p.Categorical(df["Sod"]).describe()
df.duplicated()
sum(df.duplicated())
```

MODULE 2

```
import pandas as pd
import numpy as n
import matplotlib.pyplot as plt
import seaborn as sns
d = pd.read_csv('A.csv')
d.head()
d.columns
plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
plt.hist(d['Rbcc'],color='red')
plt.subplot(1,2,2)
plt.hist(d['Htn'],color='blue')
d.hist(figsize=(15,55),layout=(15,4), color='red')
plt.show()
d['Hemo'].hist(figsize=(10,5),color='red')
plt.bar(d['Su'],d['Sc'], color='red') # scatter, plot, triplot, stackplot
plt.boxplot(d['Rbc'])
d['Bu'].plot(kind='density')
sns.displot(d['Class'], color='red')
sns.residplot(d['Sc'],d['Sod'], color='red') # residplot, scatterplot
sns.pairplot(d)
fig, ax = plt.subplots(figsize=(20,15))
sns.heatmap(d.corr(),annot = True, fmt='0.2%',cmap = 'autumn',ax=ax)
def plot(d, variable):
    dataframe_pie = d[variable].value_counts()
    ax = dataframe_pie.plot.pie(figsize=(9,9), autopct='%1.2f%%', fontsize = 10)
    ax.set_title(variable + ' \n', fontsize = 10)
    return n.round(dataframe_pie/d.shape[0]*100,2)
plot(d, 'Class')
```

MODULE 3

```
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv('A.csv')
data.head()
df=data.dropna()
df
df.columns
x = df.drop(labels='Class', axis=1)
y = df.loc[:,'Class']
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=1, stratify=y)
print("Number of training dataset : ", len(x_train))
print("Number of test dataset     : ", len(x_test))
print("Total number of dataset    : ", len(x_train)+len(x_test))
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
from sklearn.model_selection import cross_val_score
from sklearn.tree import DecisionTreeClassifier
DTC = DecisionTreeClassifier()
DTC.fit(x_train,y_train)
predicted = DTC.predict(x_test)
cr = classification_report(y_test,predicted)
print('Classification report of DecisionTreeClassifier Result is:\n',cr)
print("\n")
cm = confusion_matrix(y_test,predicted)
print('Confusion Matrix result of DecisionTreeClassifier is:\n',cm)
print("\n")
accuracy = cross_val_score(DTC, x, y, scoring='accuracy')
print('Cross validation test results of accuracy:', accuracy)
print("\n")
a = accuracy.mean() * 100
print("Accuracy Result of DecisionTreeClassifier is:",a)
ef plot_confusion_matrix(cm, title='Confusion matrix-DecisionTreeClassifier',
```

```
cmap=plt.cm.cool):
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
cm1=confusion_matrix(y_test, predicted)
print('Confusion matrix-DecisionTreeClassifier:')
print(cm)
plot_confusion_matrix(cm)
import matplotlib.pyplot as plt
df2 = pd.DataFrame()
df2["y_test"] = y_test
df2["predicted"] = predicted
df2.reset_index(inplace=True)
plt.figure(figsize=(20, 5))
plt.plot(df2["predicted"][:100], marker='x', linestyle='dashed', color='red')
plt.plot(df2["y_test"][:100], marker='o', linestyle='dashed', color='green')
plt.show()
import joblib
joblib.dump(DTC, 'model.pkl')
```

MODULE 4

```
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv('A.csv')
data.head()
df=data.dropna()
df
df.columns
x = df.drop(labels='Class', axis=1)
y = df.loc[:,'Class']
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=1,
stratify=y)
print("Number of training dataset : ", len(x_train))
print("Number of test dataset     : ", len(x_test))
print("Total number of dataset    : ", len(x_train)+len(x_test))
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
```

```
from sklearn.model_selection import cross_val_score
from sklearn.neighbors import KNeighborsClassifier
KNN = KNeighborsClassifier()
KNN.fit(x_train,y_train)
predicted = KNN.predict(x_test)
cr = classification_report(y_test,predicted)
print('Classification report of KNeighborsClassifier Result is:\n',cr)
print("\n")
cm = confusion_matrix(y_test,predicted)
print('Confusion Matrix result of KNeighborsClassifier is:',cm)
print("\n")
accuracy = cross_val_score(KNN, x, y, scoring='accuracy')
print('Cross validation test results of accuracy:', accuracy)
print("\n")
a = accuracy.mean() * 100
print("Accuracy Result of KNeighborsClassifier is:",a)
def plot_confusion_matrix(cm, title='Confusion matrix-KNeighborsClassifier',
cmap=plt.cm.cool):
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
cm1=confusion_matrix(y_test, predicted)
print('Confusion matrix-KNeighborsClassifier:')
print(cm)
plot_confusion_matrix(cm)
import matplotlib.pyplot as plt
df2 = pd.DataFrame()
df2["y_test"] = y_test
df2["predicted"] = predicted
df2.reset_index(inplace=True)
plt.figure(figsize=(20, 5))
plt.plot(df2["predicted"][:100], marker='x', linestyle='dashed', color='red')
plt.plot(df2["y_test"][:100], marker='o', linestyle='dashed', color='green')
plt.show()
```

MODULE 5

```
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv('A.csv')
data.head()
df=data.dropna()
df
df.columns
x = df.drop(labels='Class', axis=1)
y = df.loc[:,'Class']
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=1, stratify=y)
print("Number of training dataset : ", len(x_train))
print("Number of test dataset     : ", len(x_test))
print("Total number of dataset    : ", len(x_train)+len(x_test))
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
from sklearn.model_selection import cross_val_score
from sklearn.linear_model import LogisticRegression
LR = LogisticRegression()
LR.fit(x_train,y_train)
predicted = LR.predict(x_test)
cr = classification_report(y_test,predicted)
print('Classification report of LogisticRegression Result is:\n',cr)
print("\n")
cm = confusion_matrix(y_test,predicted)
print('Confusion Matrix result of LogisticRegression is:',cm)
print("\n")
accuracy = cross_val_score(LR, x, y, scoring='accuracy')
print('Cross validation test results of accuracy:', accuracy)
print("\n")
a = accuracy.mean() * 100
print("Accuracy Result of LogisticRegression is:",a)
def plot_confusion_matrix(cm, title='Confusion matrix-LogisticRegression', cmap=plt.cm.cool):
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
cm1=confusion_matrix(y_test, predicted)
print('Confusion matrix-LogisticRegression:')
print(cm)
```

```
plot_confusion_matrix(cm)
import matplotlib.pyplot as plt
df2 = pd.DataFrame()
df2["y_test"] = y_test
df2["predicted"] = predicted
df2.reset_index(inplace=True)
plt.figure(figsize=(20, 5))
plt.plot(df2["predicted"][:100], marker='x', linestyle='dashed', color='red')
plt.plot(df2["y_test"][:100], marker='o', linestyle='dashed', color='green')
plt.show()
```

MODULE 6

```
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv('A.csv')
data.head()
df=data.dropna()
df
df.columns
x = df.drop(labels='Class', axis=1)
y = df.loc[:,'Class']
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=1,
stratify=y)
print("Number of training dataset : ", len(x_train))
print("Number of test dataset     : ", len(x_test))
print("Total number of dataset    : ", len(x_train)+len(x_test))
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
from sklearn.model_selection import cross_val_score
from sklearn.naive_bayes import GaussianNB
GN = GaussianNB()
GN.fit(x_train,y_train)
predicted = GN.predict(x_test)
cr = classification_report(y_test,predicted)
print('Classification report of GaussianNB Result is:\n',cr)
print("\n")
cm = confusion_matrix(y_test,predicted)
print('Confusion Matrix result of GaussianNB is:',cm)
print("\n")
accuracy = cross_val_score(GN, x, y, scoring='accuracy')
print('Cross validation test results of accuracy:', accuracy)
```

```
print("\n")
a = accuracy.mean() * 100
print("Accuracy Result of GaussianNB is:",a)
def plot_confusion_matrix(cm, title='Confusion matrix-GaussianNB',
cmap=plt.cm.cool):
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
cm1=confusion_matrix(y_test, predicted)
print('Confusion matrix-GaussianNB:')
print(cm)
plot_confusion_matrix(cm)
import matplotlib.pyplot as plt
df2 = pd.DataFrame()
df2["y_test"] = y_test
df2["predicted"] = predicted
df2.reset_index(inplace=True)
plt.figure(figsize=(20, 5))
plt.plot(df2["predicted"][:100], marker='x', linestyle='dashed', color='red')
plt.plot(df2["y_test"][:100], marker='o', linestyle='dashed', color='green')
plt.show()
```

MODULE 7

```
from django.shortcuts import render
from django.shortcuts import render, redirect
import numpy as np
import joblib
model =
joblib.load('C:/Users/DELL/Downloads/PROJECT/PROJECT/deploy/app/model.pkl')
# Create your views here.
def home(request):
    return render(request, "index.html")
def predict(request):
    if request.method == "POST":
        int_features = [x for x in request.POST.values()]
        int_features  =  int_features[1:]
        print(int_features)
        final_features = [np.array(int_features, dtype=object)]
        print(final_features)
        prediction = model.predict(final_features)
        print(prediction)
        output = prediction[0]
        print(f'output{output}')
        if output == 0:
```

```
        return render(request, 'index.html', {"prediction_text":"NO CHRONIC
KIDNEY DISEASE AND NO KIDNEY STONE"})
    elif output == 1:
        return render(request, 'index.html', {"prediction_text":"CHRONIC KIDNEY
DISEASE STAGE 1 AND NO KIDNEY STONE"})
    else:
        return render(request, 'index.html', {"prediction_text":"CHRONIC KIDNEY
DISEASE STAGE 2 AND KIDNEY STONE"})
    print(output)
```

5.4 OUTPUTS

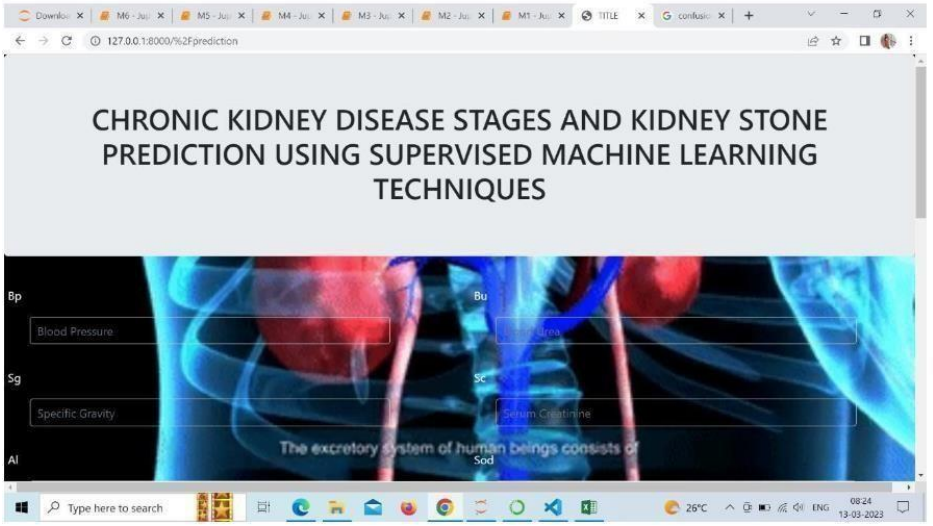
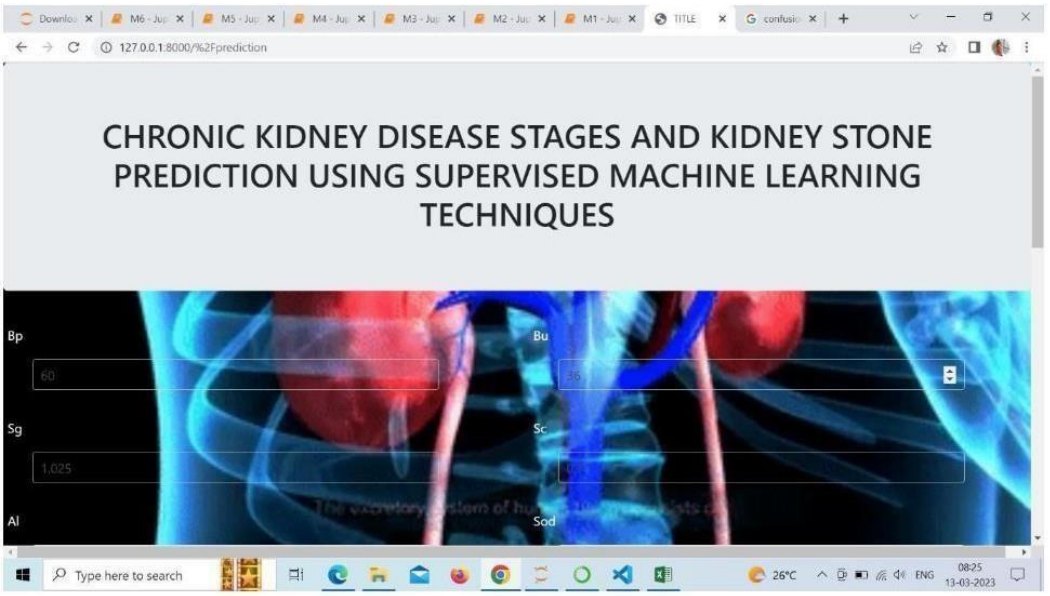


Fig 5.4.1 Prediction Page

STAGE 0



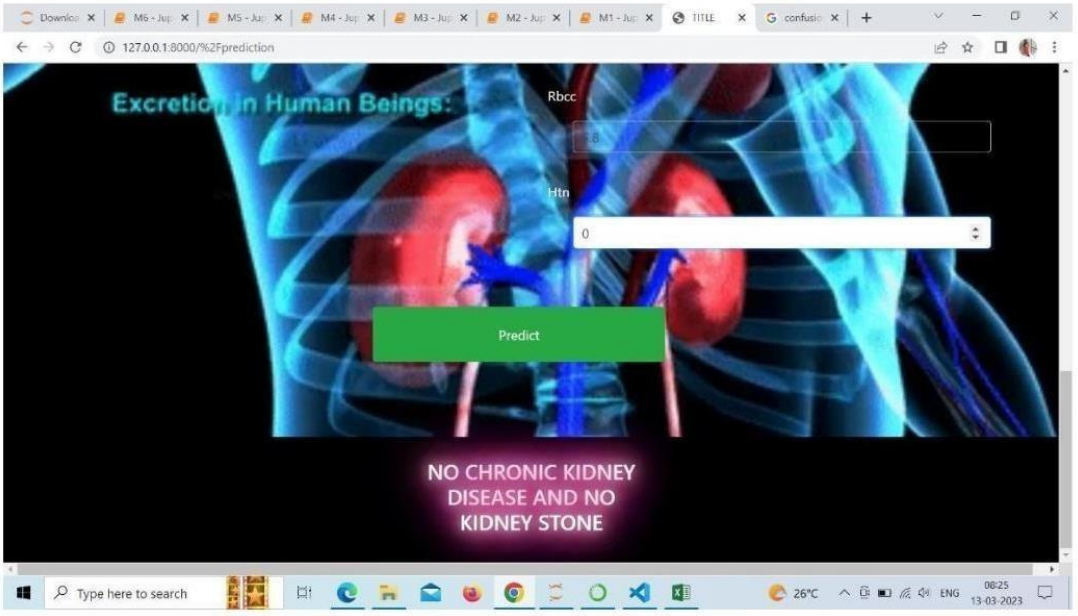


Fig 5.4.2 Sample output for stage 0

STAGE 1

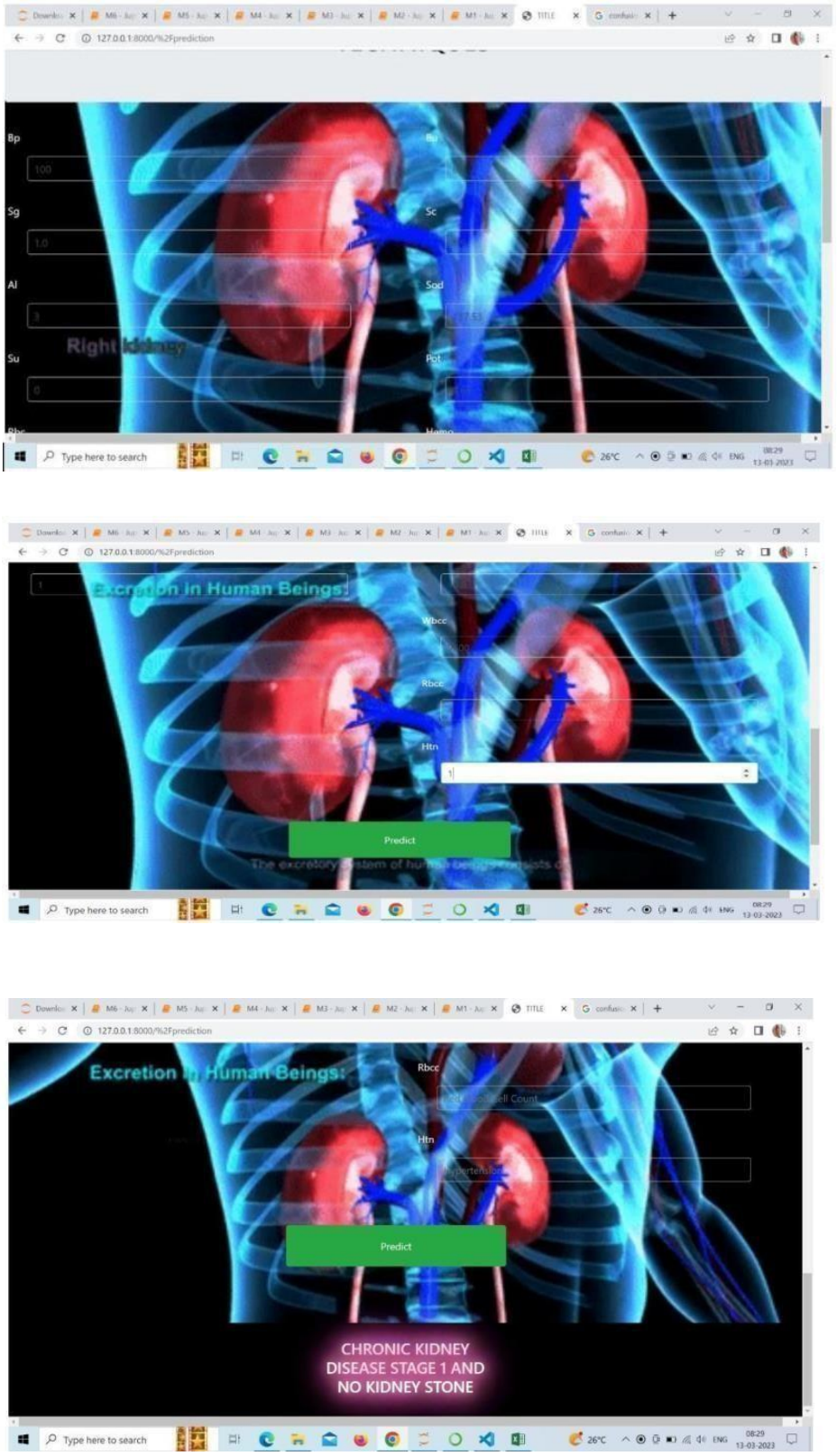


Fig 5.4.3 Sample output stage 1

STAGE 2

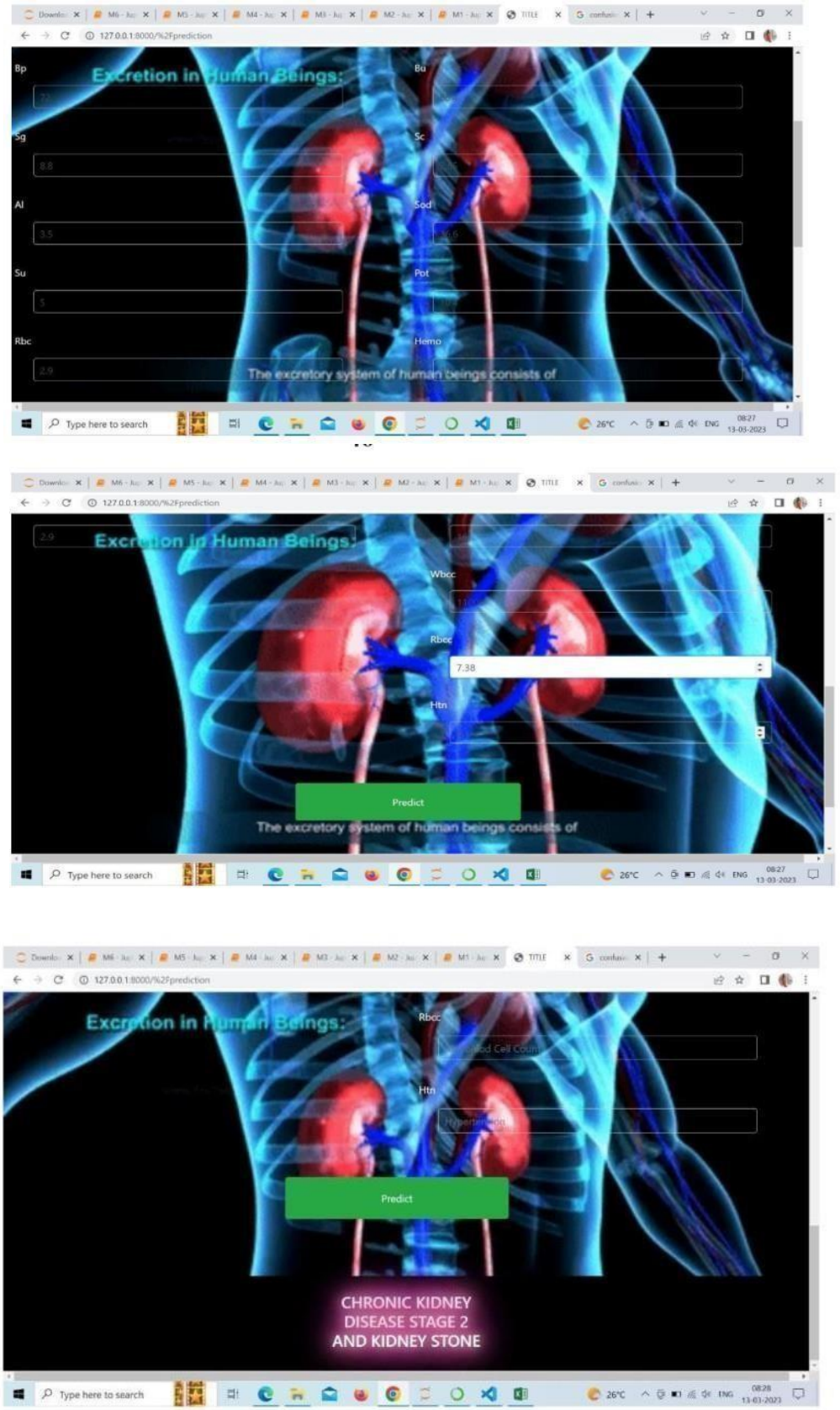


Fig 5.4.4 Sample output stage 2

CHAPTER 6

UML DIAGRAMS AND TESTING

6.1 UML DIAGRAMS

6.1.1 USE CASE DIAGRAM

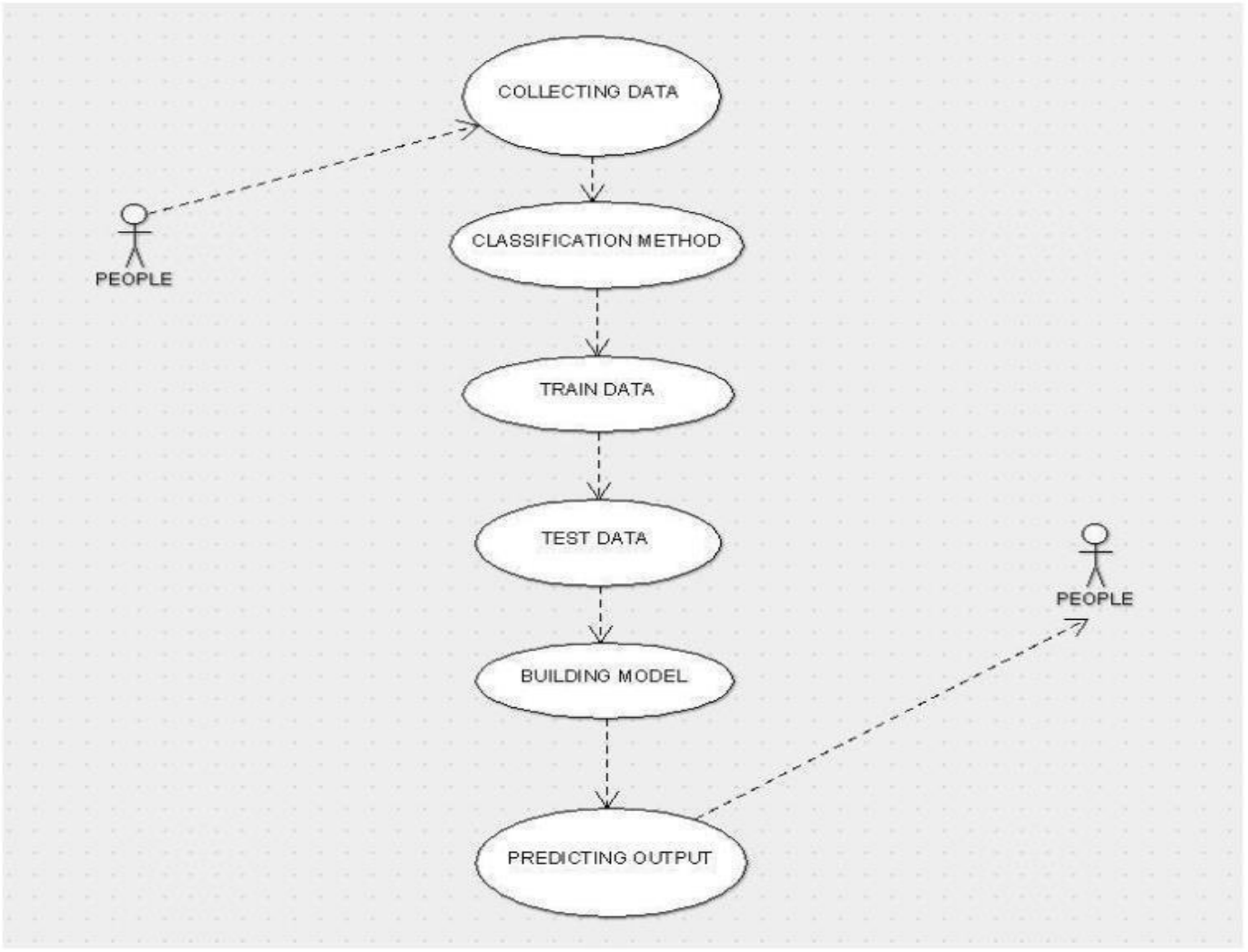


Fig 6.1.1 Use case Diagram

EXPLANATION

Use case diagrams are considered for high level requirement analysis of a system. So when the requirements of a system are analyzed the functionalities are captured in use cases. So, it can say that uses cases are nothing but the system functionalities written in an organized manner. In this we use use case diagram to collect the data for predicting the output.

6.1.2 CLASS DIAGRAM

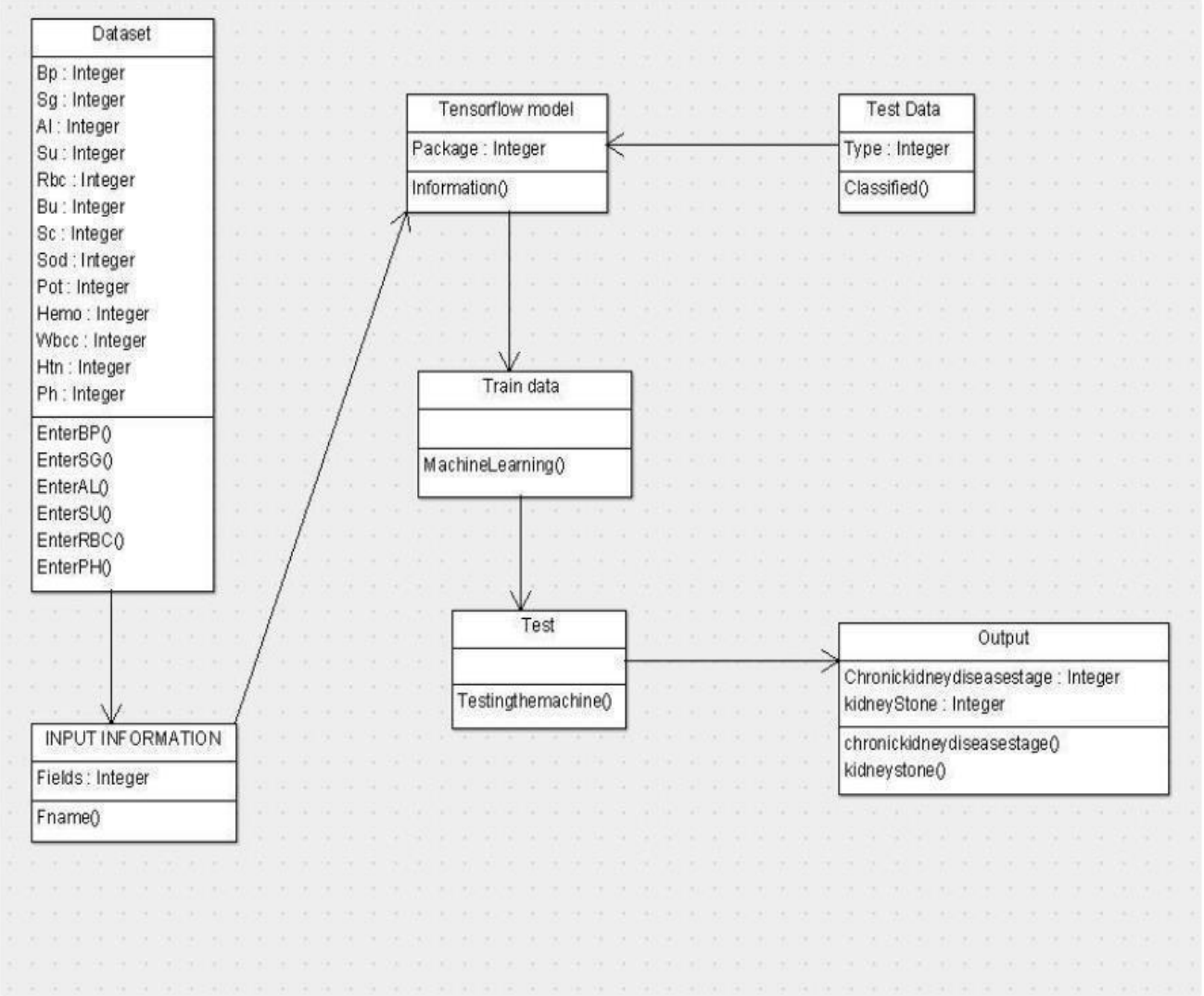


Fig 6.1.2 Class Diagram

EXPLANATION

Class diagram is basically a graphical representation of the static view of the system and represents different aspects of the application. So a collection of class diagrams represent the whole system. The name of the class diagram should be meaningful to describe the aspect of the system. Each element and their relationships should be identified in advance Responsibility (attributes and methods) of each class should be clearly identified for each class minimum number of properties should be specified and because, unnecessary properties will make the diagram complicated. Use notes whenever required to describe some aspect of the diagram and at the end of the drawingit should be understandable to the developer/coder. Finally, before making the final version, the diagram should be drawn on plain paper and rework as many times as possible to make it correct.

6.1.3 ACTIVITY DIAGRAM

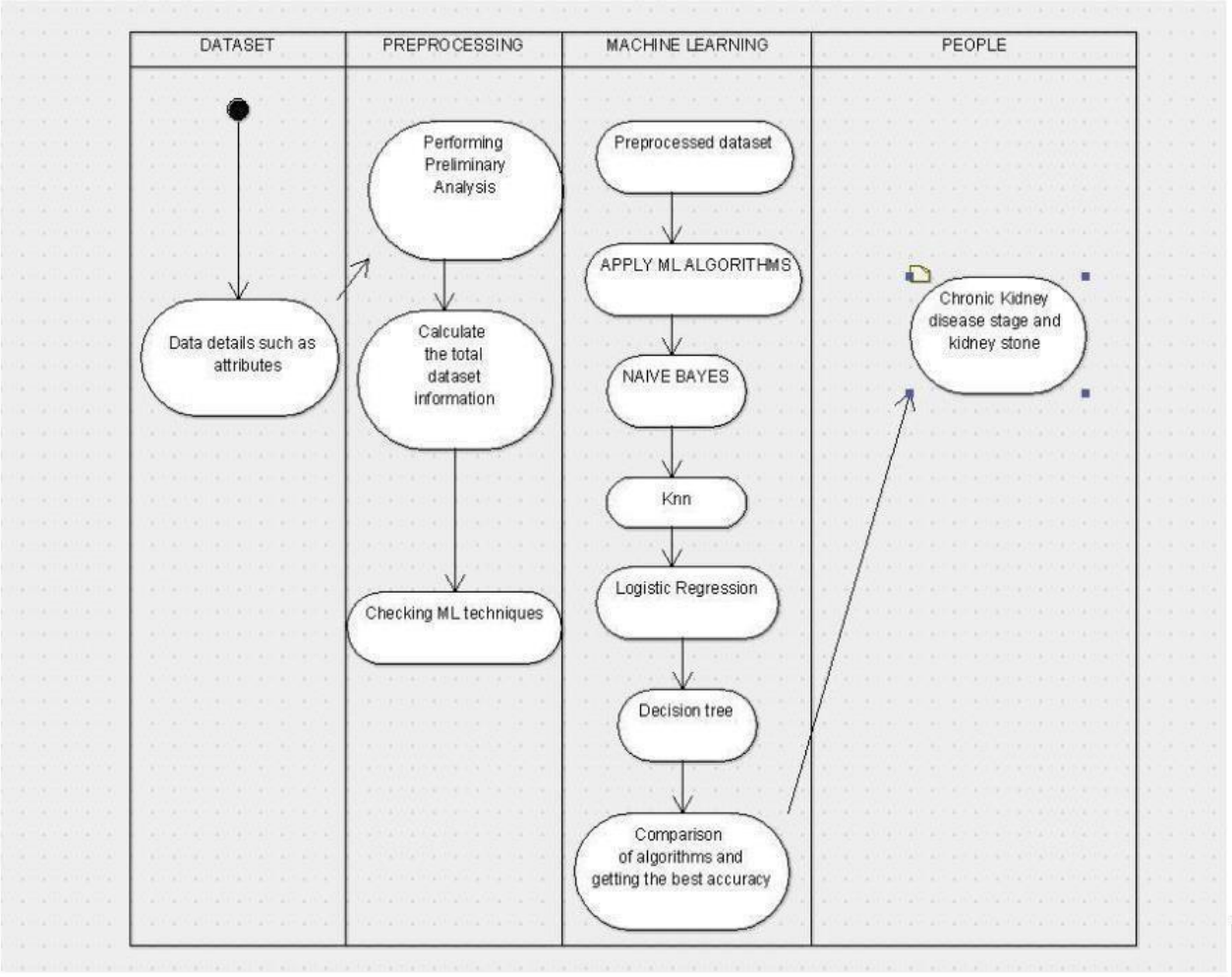


Fig 6.1.3 Activity Diagram

EXPLANATION

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing dynamic nature of a system but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in activity diagram is the message part. It does not show any message flow from one activity to another. Activity diagram is some time considered as the flowchart. Although the diagrams looks like a flow chart but it is not. It shows differentflow like parallel, branched, concurrent and single.

6.1.4 SEQUENCE DIAGRAM

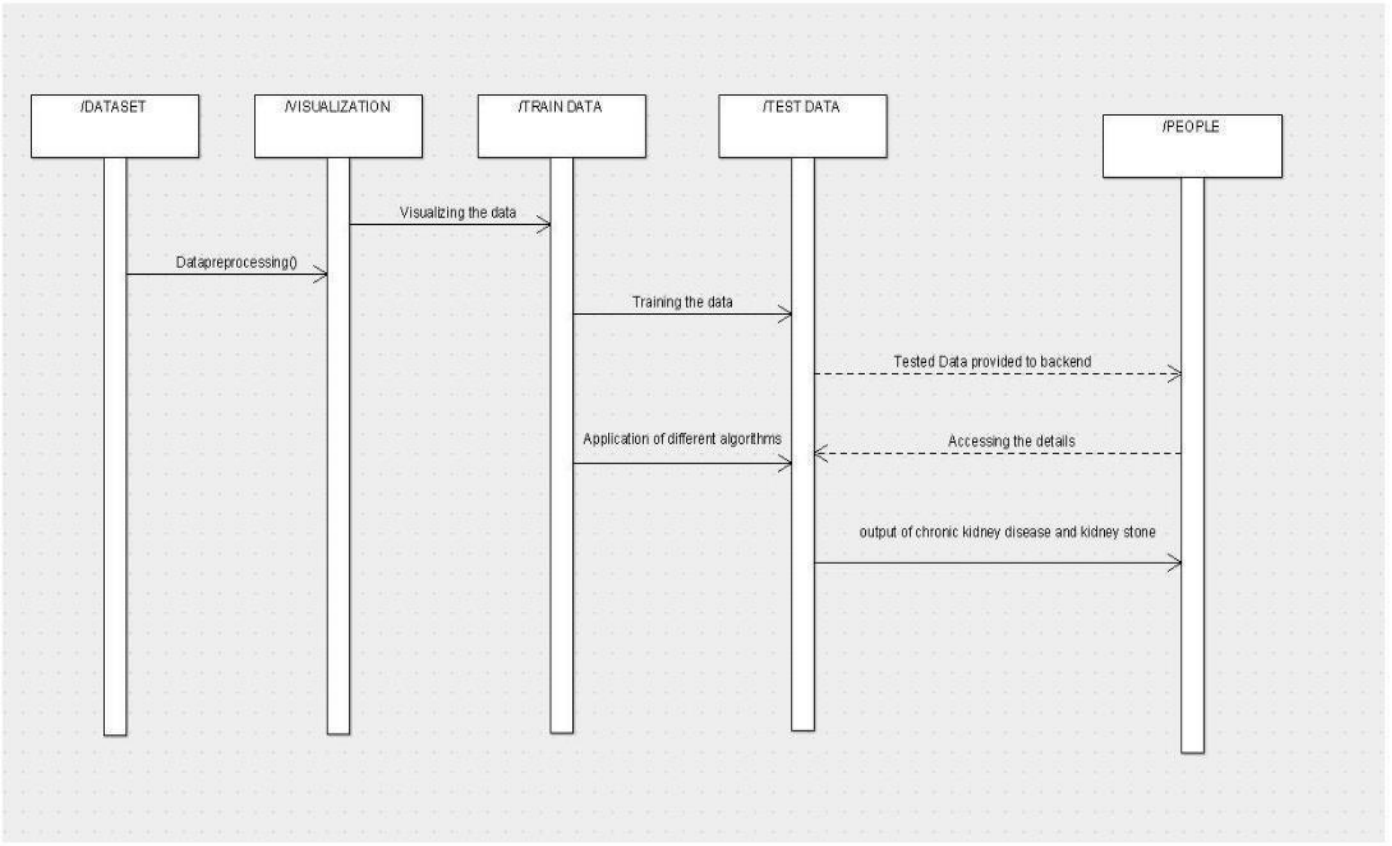


Fig 6.1.4 Sequence diagram

EXPLANATION

Sequence diagrams model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and are commonly used for both analysis and design purposes. Sequence diagrams are the most popular UML artifact for dynamic modeling, which focuses on identifying the behavior within your system. Other dynamic modeling techniques include activitydiagramming, communication diagramming, timing diagramming, and interaction overview diagramming. Sequence diagrams, along with class diagrams and physicaldata models are in my opinion the most important design-level models for modern business application development.

6.1.5 COLLABORATION DIAGRAM

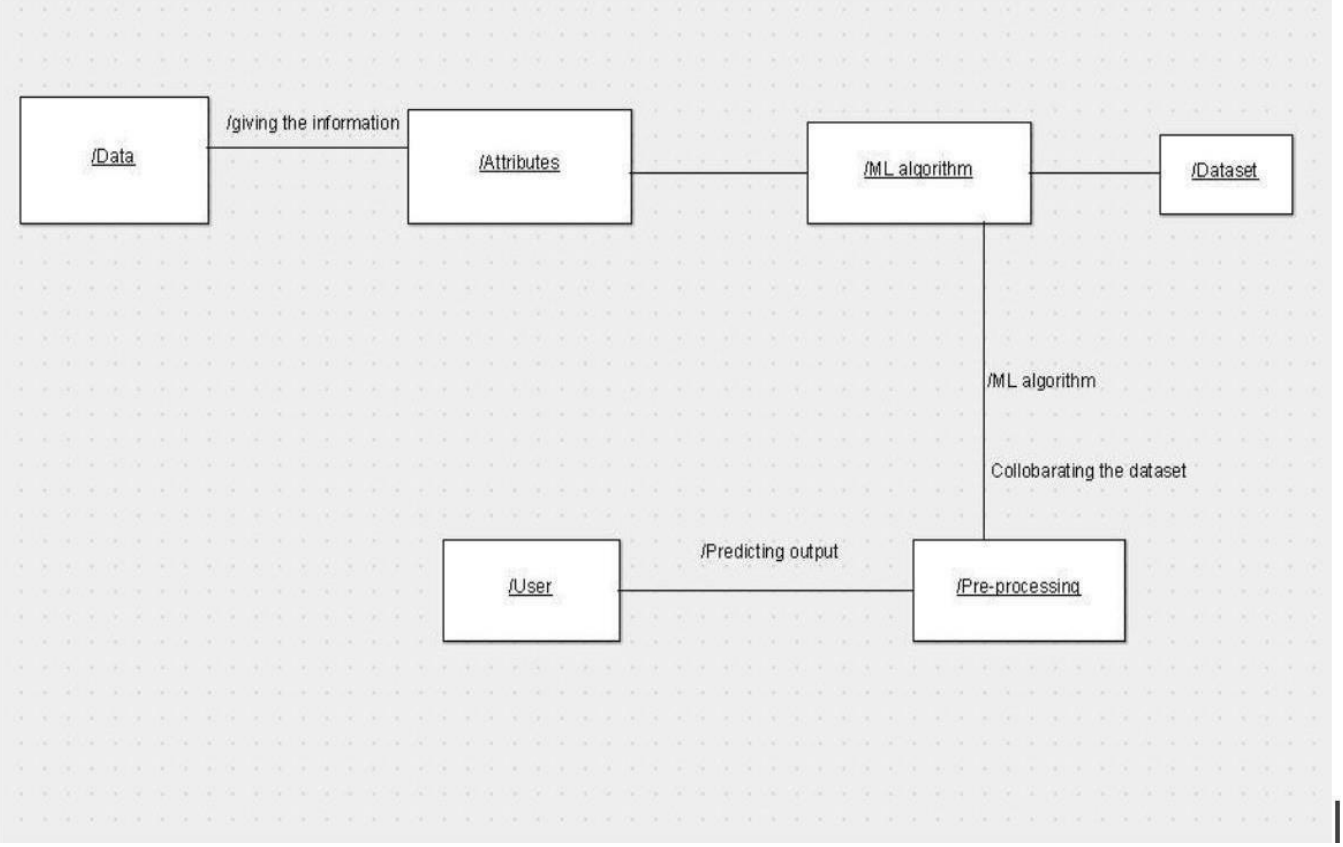


Fig 6.1.5 Collaboration diagram

EXPLANATION

A collaboration diagram is a type of visual presentation that shows how various software objects interact with each other within an overall IT architecture and how users (like doctor or patient) can benefit from this collaboration. A collaboration diagram often comes in the form of a visual chart that resembles a flow chart. It can show, at a glance, how a single piece of software complements other parts of a greater system.

6.1.6 ENTITY RELATIONSHIP DIAGRAM (ERD)

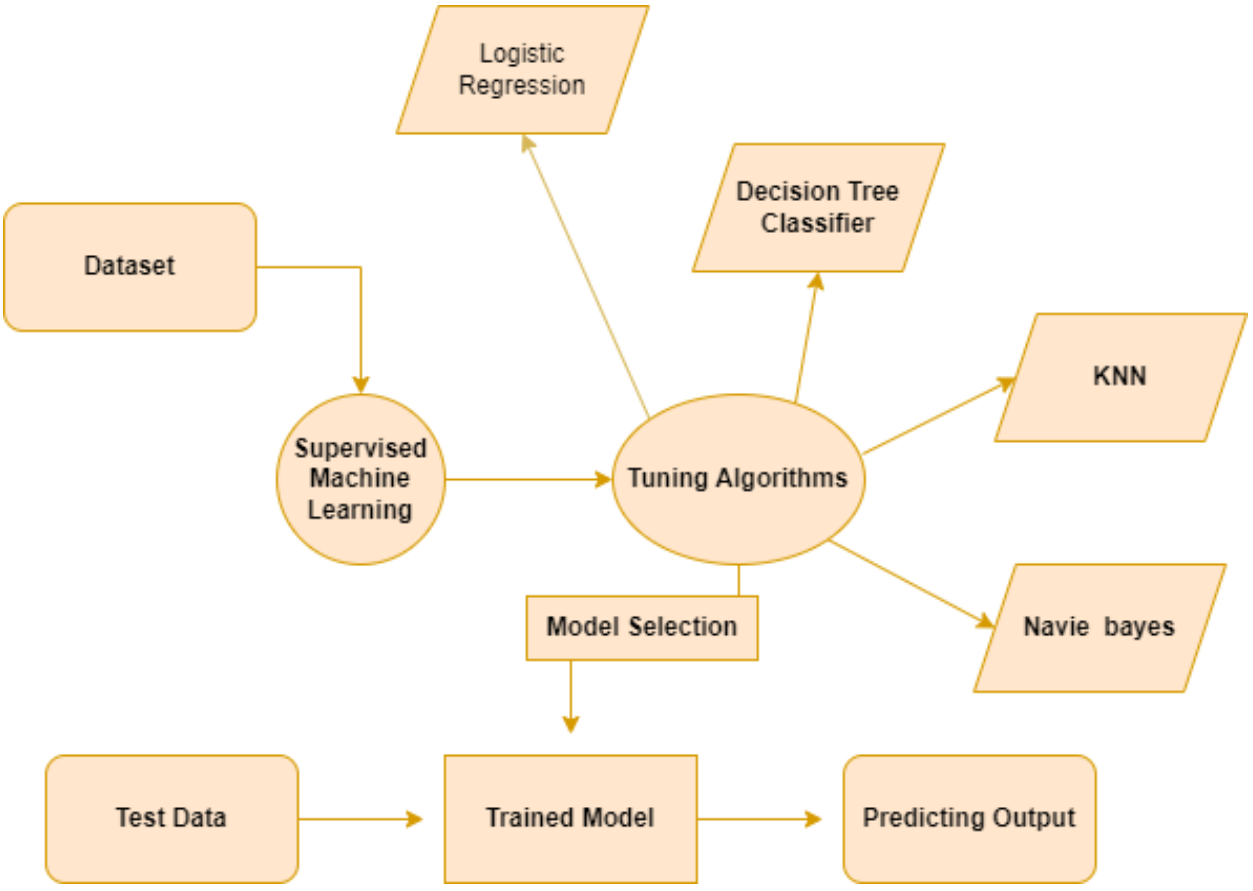


Fig 6.1.6 Entity Relationship Diagram(ERD)

EXPLANATION

An entity relationship diagram (ERD), also known as an entity relationship model, is a graphical representation of an information system that depicts the relationships among people, objects, places, concepts or events within that system. An ERD is a data modeling technique that can help define business processes and be used as the foundation for a relational database. Entity relationship diagrams provide a visual starting point for database design that can also be used to help determine information system requirements
throughout an organization. After a relational database is rolled out, an ERD can still serve as a referral point, should any debugging or business process re-engineering be needed later.

6.2TESTING

6.2.2 UNIT TESTING

Unit testing is conducted to verify the functional performance of each modular component of the software. Unit testing focuses on the smallest unit of the software design (i.e.), the module. The white-box testing techniques were heavilyemployed for unit testing.

6.2.3 SYSTEM TESTING

Testing is performed to identify errors. It is used for quality assurance. Testing is an integral part of the entire development and maintenance process. The Goal of the testing during phase is to verify that the specification has been accurately and completely incorporated into the design, as well as to ensure the correctness of thedesign itself. For example, the design must not have any logic faults in the design is detected before coding commences, otherwise the cost of fixing the faults will be considerably higher as reflected. Detection of design faults can be achieved by means of inspection as well as walk through. Testing is one of the important steps in the software development phase. Testing checks for the errors,as a whole of theproject testing involves the following test cases:

6.2.A Static analysis is used to investigate the structural properties ofthe Source code.

6.2.B Dynamic testing is used to investigate the behavior of the source codeby executing the program on the test.

6.2.4 FUNCTIONAL TESTING

Functional testing is a quality assurance (QA) process and a type of black box testing that bases its test cases on the specifications of the software component under test. Functions are tested by feeding them input and examining the output, and internal program structure is rarely considered (not like in white-box testing). Functional Testing usually describes what the system does. Functional testing differs from system testing in that functional testing "verifies a program by checking it against ... design document(s) or specification(s)", while system testing "validate a program by checking it against the published user or system requirements" (Kane, Falk, Nguyen 1999, p. 52). Functional testing typically involves five steps. The identification of functions that the software is expected to perform

6.2.4.1 The creation of input data based on the function's specifications

6.2.4.2 The determination of output based on the function's specifications

6.2.4.3 The execution of the test case

6.2.4.4 The comparison of actual and expected outputs.

6.2.5 PERFORMANCE TESTING

In general testing performed to determine how a system performs in terms of responsiveness and stability under a particular workload. It can also serve to investigate, measure, validate or verify other quality attributes of the system, such as scalability, reliability and resource usage. Performance testing is a subset of performance engineering, an emerging computer science practice which strives to build performance into the implementation, design and architecture of a system.

6.2.6 INTEGRATION TESTING

Integration testing is a systematic technique for constructing the program structurewhile at the same time conducting tests to uncover errors associated with. Individual modules, which are highly prone to interface errors, should not be assumed to work instantly when put together. The problem of course, is putting them together interfacing. There may be the chances of data lost across on another’s sub functions, when combined may not produce the desired major function; individually acceptable impression may be magnified to unacceptable levels; global data structures can present problems. Integration testing is the phase in software testing in which individual software modules are combined and tested as a group. Integration testing takes as its input modules that have been unit tested, groups them in larger aggregates, applies testsdefined in an integration test plan to those aggregates, and delivers as its output the integrated system ready. All the errors found in the system are corrected for the next phase.

The purpose of integration testing is to verify functional, performance, and reliability requirements placed on major design items. These "design items", i.e. assemblages (or groups of units), are exercised through their interfaces using black box testing, success and error cases being simulated via appropriate parameter and data inputs. Simulated usage of shared data areas and inter-process communication is tested and individual subsystems are exercised through their input interface. Test cases are constructed to test

whether all the components within assemblages interact correctlyfor example across procedure calls or process activations, and this is done after testing individual modules,
i.e. unit testing.

6.2.7 VALIDATION TESTING

Verification and Validation are independent procedures that are used together for checking that a product, service, or system meets requirements and specifications and that it full fills its intended purpose. These are critical components of a quality management system such as ISO 9000. The words "verification" and "validation" are sometimes preceded with "Independent" (or IV&V), indicating that the verification and validation is to be performed by a disinterested third party. It is sometimes said that validation can be expressed by the query "Are you building the right thing?" and verification by "Are you building it right?". In practice, the usage of these terms varies. Sometimes they are even used interchangeably.

6.2.8 USER ACCEPTANCE TESTING

User acceptance of a system is the factor for the success of any system. The system under consideration is tested for the user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changeswherever required.

- Input screen design.
- Output screen design.
- Online message to guide user.
- Format of the ad-hoc reports and other outputs.

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using the test data. While testing the system by using test data errors are again uncovered and correct.

6.3 TEST CASES

TEST CASE ID	MODULE	INPUT	EXPECTED OUTPUT	ACTUAL OUTPUT	RESULT
TC1	Collecting thedata	Dataset	Attribute segregation	Processed Data	PASS
TC2	Detecting The Stage	Attribute values like BP, SOD,POT, PH, etc	Stage Of Kidney Disease And Presence Of Kidney Stone	Stage 0,1 And 2 Kidney Stone PresentOr Not	PASS
TC3	Algorithm Implementation	Clean Data And Attribute Values	Accuracy And Average	Best Accuracy And Average	PASS
TC4	Webpage	Values For Attributes Present	Result Satges And Kidney Stone	Stage 0,1 And 2 Kidney Stone PresentOr Not	PASS

CHAPTER-7

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

Data preparation and processing, missing value analysis, exploratory analysis, and model construction and evaluation came first in the analytical process. The highest accuracy score on the public test set will be discovered. This software canassist in predicting kidney stonesand chronic renal illness.

7.2 FUTURE ENHANCEMENT

1. Hospitals aspire to automate (in real time) the process ofexcluding diseasedpeople from eligibility.
2. To automate this procedure by displaying the prediction outcome ina desktopor web application.
3. To streamline the work that has to be done in an AI setting.

APPENDICES

PAPER:

PREDICTION OF CHRONIC KIDNEY DISEASE STAGES AND CHRONIC KIDNEY STONES USING THE SML TECHNIQUE

[1A]M.SUMITHRA,[1B]DEEPA.S, [1C]GRACE NESAM.R, [1D]PAVITHRAA.S
[1A]msumithra@panimalar.ac.in,[1B]sdeepa1910@gmail.com,[1C]gracenesam01@gmail.com,[1D]spavithra0507@gmail.com

ABSTRACT

The phrase chronic kidney disease refers to kidney damage which is continuous for long time and may get worse over time. The kidney does not function properly if the harm is severe. This is mentioned as End-Stage Renal disease or Kidney failure. Patients with kidney disease may enter the chronic phase which is characterized by a gradual decline in kidney function. For determining whether kidney disease is severe or not, a variety of algorithm is used. For predicting the disease's stages, it is further taken into consideration if it is severe. Additionally, kidney stones are also diagnosed for its presence in the human body. In this project the modules are deployed to find the appropriate accuracy of the disease. Here, Supervised Machine Learning Techniques are used to predict the accuracies of the disease stages and the presence of stones.

1.INTRODUCTION

Using SML techniques, we choose the subject of knowledge set to improve detection of chronic kidney disease and kidney stone. People are all aware about CKD which is one of the most common disease among individuals nowadays. By gathering a large amount of CKD and kidney stone data, we are working hard to improve the spotting of disease in people. It usually affects the adults above the age of 30 and continues indefinitely. Males between age of thirty to thirty five have stage five CKD without a transplant, one could expect to live for fourteen years. Women of the same age affected with this disease are expected to live for thirteen years. Everybody has a four year life expectancy between the age of seventy to seventy five regardless of the gender. GFR rate is popular tool in identifying the majority of renal disorders.

2.LITERATURE SURVEY

When aberrant albumin excretion or impaired kidney function last for longer than three months, as determined by a measured or estimated glomerular filtration rate (GFR), CKD is present. According to the stage of the disease, treatments are suggested for CKD and dialysis patients. These treatments could lower these patients' morbidity and fatality rates. Although it has not been determined for sure, it is prudent to abide by FDA instructions.[1]

The originality in this study is in creating a system for diagnosing chronic renalillnesses. This study helps specialists about studying preventive methods for CKD through early diagnosis utilizing ML approaches. This paper's main ideology was to assess a dataset made up of 400 patients and 24 attributes. The missing nominal and numerical data were replaced using the mean and mode statistical analysis methods. sRFE helps to select the most crucial characteristics. Our systems' accuracy ranged from 100% with random forest to 97.3%with SVM. These papers don't accurately depict chronic renal disease.[2]

This article inspect CKD prediction in a different manner. In this paper, seven classifier methods were used. The results have been calculated for each classifier based on the particular criterion features. The linear support vector machine provided the maximum accuracy of98.46%. Logistic and KNN were not employed in SMOTE since they didnot produce the desired results.[3]

In this article they predict they predict the Chronic kidney disease stages and kidney stone using glomerular filtration rate. They have combined deep neural network and shallow neural network for determining the chronic kidney disease stages and kidney stone. They just capture the filtering rate to predict the output. The deep learning approach provides the maximum accuracy of 88.3% it focuses on GFR for predicting the output. In this they just capture the Filtering Rate and they don't use any machine learning algorithms. Another big drawback is they don't compare different algorithms and accuracy is reduced.[4]

In this paper they have predicted the chronic kidney disease by concentrating on several machine learning methods. Mainly they considered artificial neural network and support vector machines as classification methods for prediction. Both these model helped to produce accurate result of prediction. But artificial neural network produced better and accurate result when compared to SVM. In this paper they have used 10 fold cross validation algorithm. ANN produced for accurate result when compared to SVM. Result is not accurately predicted.[5]

In this paper, they have used heterogeneous modified artificial neural network for predicting CKD. They used digital images for prediction. First the MR and CT scan images of abdominal are segmented into left and right portion to validate and train dataset. In the images which are used for prediction there are lot of noisy and missing values to reduce that HMANN method is proposed and segmentation is helped for better identification of kidney stone produces precise prediction of kidney stone. Time taken for prediction is more.[6]

3.EXISTING SYSTEM

For the clinical diagnosis of kidney disease stages and stone formation, an accurate testing of Glomerular filtration rate (GFR) is essential. Computer learning methodologies such as deep neural networks offer a viable way for improving GFR estimation accuracy. In order to find GFR, they created a uniquearchitecture called a deep and shallow neural network (dlGFR). We then compared its working to that of estimated GFR derived from the MDRD and CKD-EPI equations for stages and kidney stone epidemiology.

4.PROPOSED SYSTEM

The proposed ideology is to create a ML model for categorizing renal disorder stages and kidney stones. The procedure begins with data collection, in which past information about kidney disease stages and stones is collected. With the help of attributes, we implement different algorithms and try to find the best accuracy between them. The one that produces the best accuracy is promoted to provide the output to the user.

4.1. Scope of The Project

The goal of the research is to use ML methods for studying a dataset collected from a hospital website for determining whether a patient has CKD or not and has any kidney stones as soon as feasible.

4.2. List of Modules

1. Data pre-processing
2. Data visualization
3. Implementing Algorithms

a. Logistic Regression
b. KNN
c. Decision Tree Classifier
d. Naïve Bayes

4. Deployment

4.3. Flow Diagram

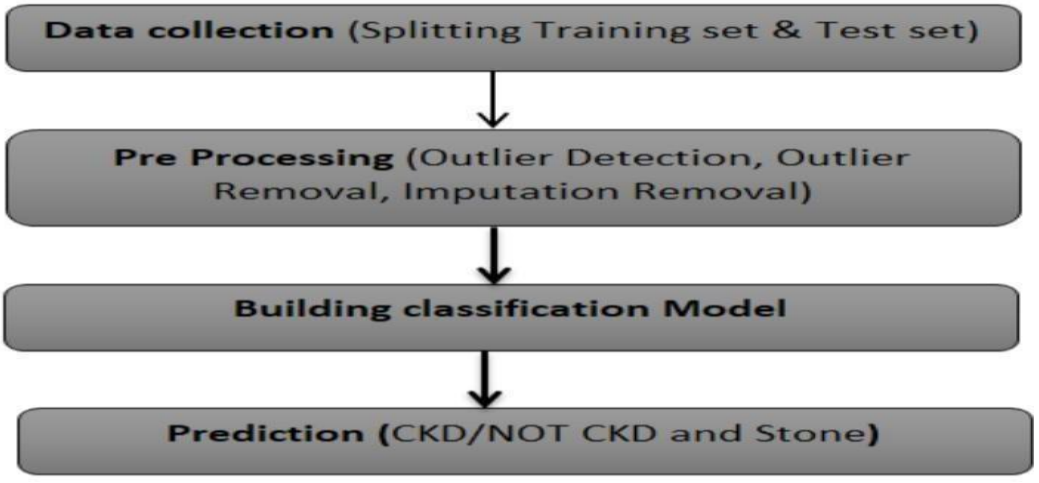


Fig 1:Flow Diagram of Proposed System

4.4.Working Model

According to the blood test report, the values for each attributes will be entered in the website which we have given. Then the User will click the "PREDICT" button. After then the process will be carried out in the backend to produce the results according to the data which the user entered. In that the first process will be "DATA PREPROCESSING". After Data Preprocessing, Data visualization will be performed for better understanding of the data. After Data visualization, We are separating the

dataset into training dataset and testing dataset. The output of the training dataset will be classification of ML algorithms. So with ML algorithms, we would test the dataset for the accuracies of different algorithms. After the accuracy report of each algorithms, we consider the best accuracy algorithm to find the stages and kidney stone according to the values entered by the user. For finding the stages and kidney stones, the average of attributes valuesare considered to predict the result as such like Class 0,Class 1, Class 2.

4.5 System Architecture

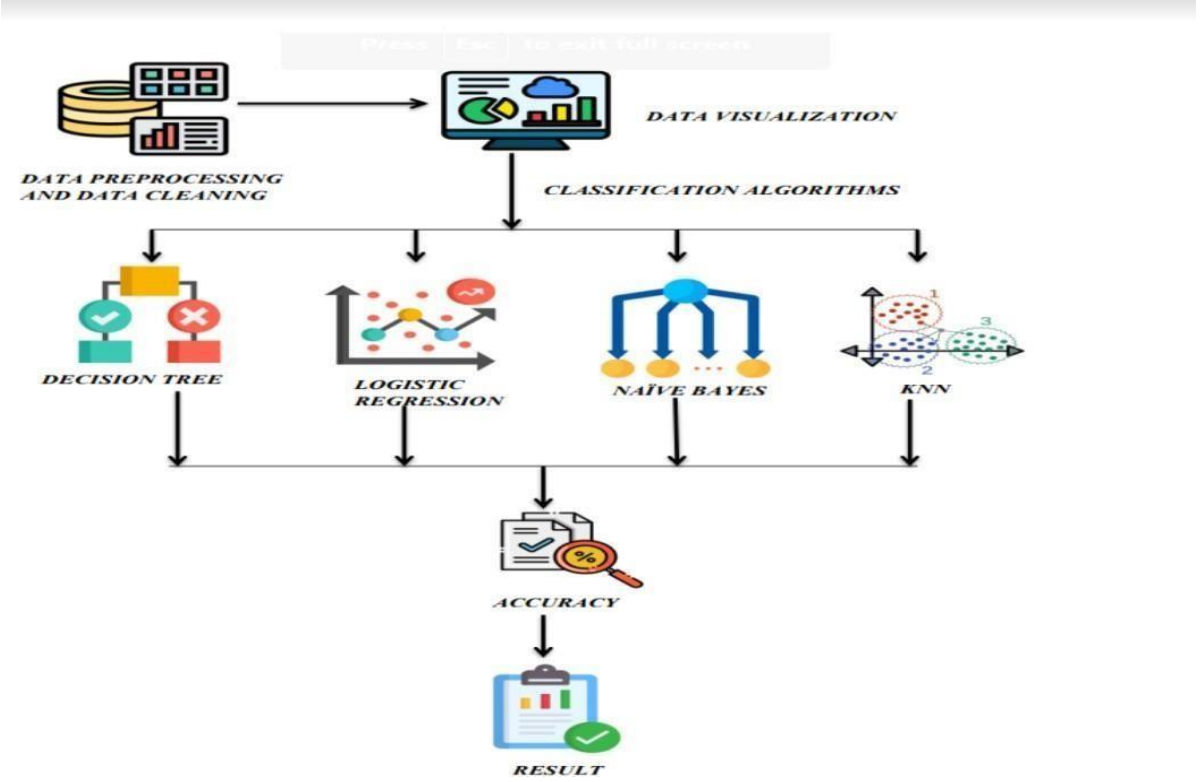


Fig 2: System Architecture

5.RESULT AND DISCUSSION

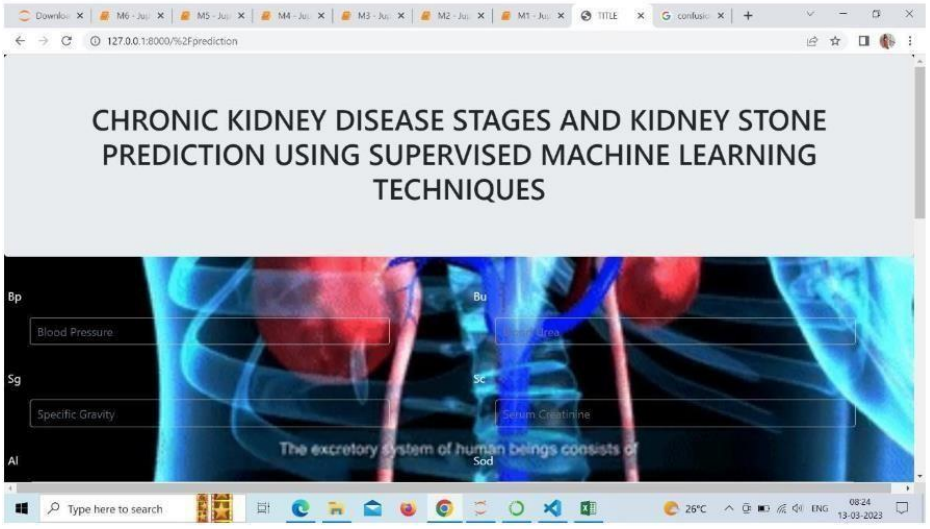


Fig 3:Prediction Page

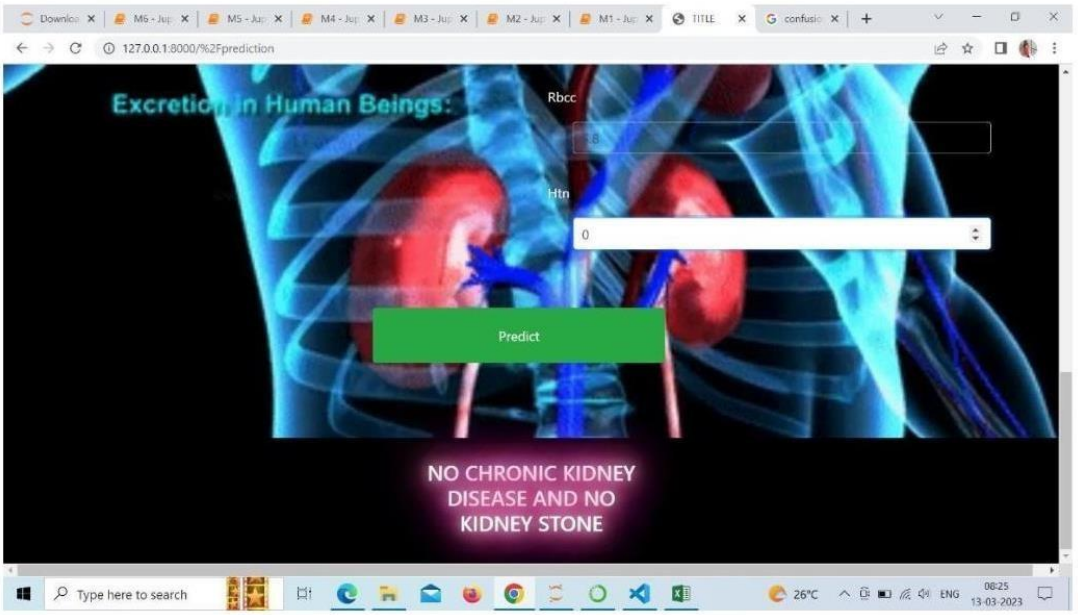


Fig 4: Stage 0

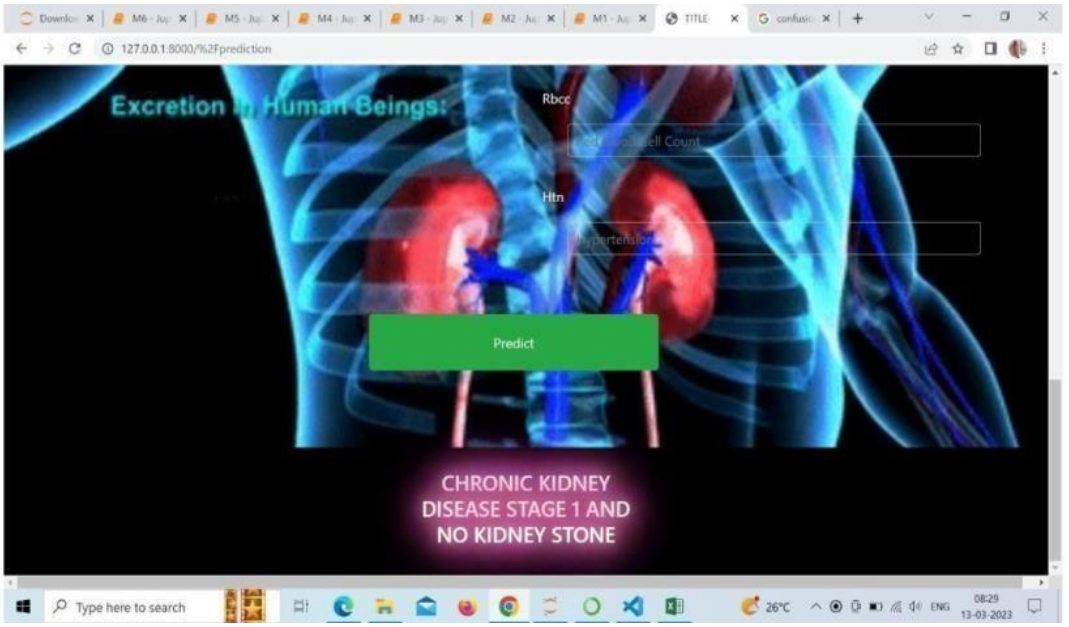


Fig 5: Stage 1

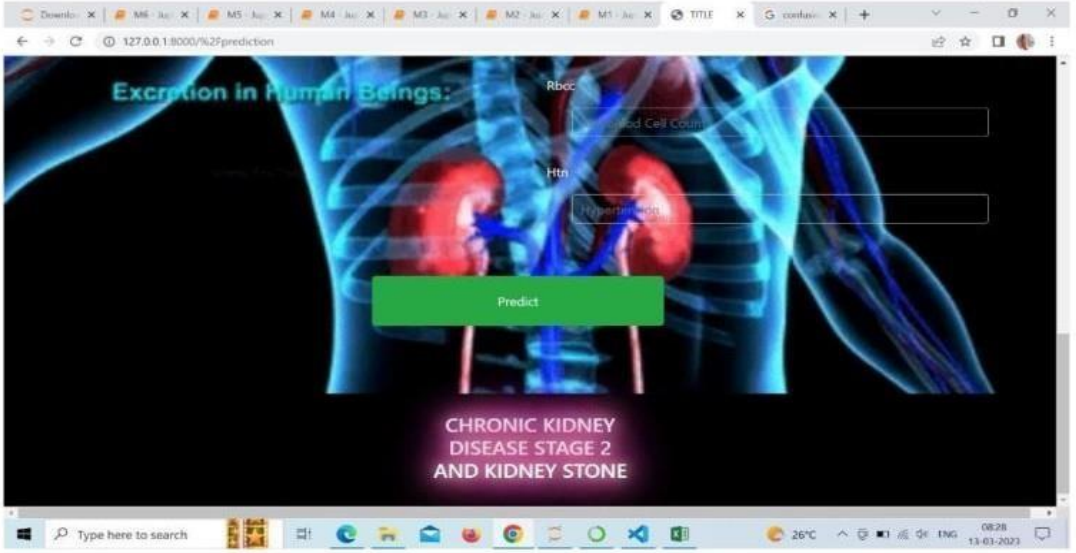


Fig 6: Stage 2

Figure 3 is the sample output for the webpage we have designed. Users will be able to interact and get their result using this webpage. Users will give their input values that is attributes like BP, Serum creatinine, potassium etc.. values according to their test results. Then after entering the attributes values the user will click the predict button which is shown in figure 4. Afterwards in the backend the code will be processed and show the result whether stage 0,1 or 2 which is shown in figure 8,9 or 10 according to the attributes values mentioned by the user**.**

6.CONCLUSION

In this project we tried to find the CKD disease's presence and the kidneystone presence.In this project we achieved accuracy of 97.9%.

7.FUTURE ENHANCEMENT

1.Hospitals aspire to automate (in real time) the process of excludingdiseased people from eligibility.

2. To automate this process by displaying the results in websites

3.To simplify the tasks that must be completed in an AI environment.

REFERENCE

[1] Robert Thomas, MD, AbbasKans, MD,John R. Sedor, MD-Chronic Kidney DiseaseAnd itsComplications.(2008)

[2] Mosleh Hmoud Al-Adhaileh,Fawaz Waselallah Alsaade,Theyazn H. H. Aldhyani,Ahmed Abdullah Alqarni-Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques.

[3] Pankaj Chittora, Sandeep Chaurasia, Prasun Chakrabarti, Gaurav Kumawat1,Tulika Chakrabarti -Prediction Of Chronic Kidney Disease – A Machine Learning Perspective

[4] Haishuai Wang, Benjamin Bowe, Zhicheng Cui, Hong Yang, S Joshua Swamidass, Yan Xie, Ziyad Al-Aly-A Deep Learning Approach for the Estimation of Glomerular Filtration Rate

[5] Njoud Abdullah Almansour,Hajra Fahim Syed,Nuha Radwan Khayat,Rawan Kanaan Altheeb- Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study

[6] panelFuzhe Ma a,Tao Sun a,Lingyun Liu- Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network

[7] B.Buvanswari and T.Kalpalatha Reddy, ―A Review of EEG Based Human Facial Expression Recognition Systems in Cognitive Sciences‖ International Conference on Enenrgy, Communication,Data analytics and SoftComputing(ICECDS),CFP17M55-PRJ:978-1-5386-1886-8‖,August 2017.

[8]M. Sumithra and Dr. S. Malathi, ‖ Modified Global Flower Pollination Algorithm-based image fusion for medical diagnosis using computed tomography and magnetic resonance imaging‖, International Journal of Imaging Systems and Technology, Vol. 31, Issue No.1, pp. 223-235, 2021

[9]K. Sridharan , and Dr. M. Chitra "SBPE: A paradigm Approach for proficient Information Retrieval , Jokull Journal" , Vol 63, No. 7;Jul 2013
[10] M. Sumithra and Dr. S. Malathi, ―3D Densealex NET Model with Back Propagation for Brain Tumor Segmentation‖, International Journal OfCurent Research and Review, Vol. 13, Issue 12, 2021.
[11]B.Buvaneswari and Dr.T. Kalpalatha Reddy,―EEG signal classification using soft computing techniques for brain disease diagnosis‖,Journal of International Pharmaceutical Research ,ISSN : 1674-0440,Vol.46,No.1,Pp.525-528,2019.
[12]K. Sridharan , and Dr. M. Chitra "Web Based Agent And Assertion Passive Grading For Information Retervial", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 16, September 2015 pp:7043-7048
[13]M. Sumithra and Dr. S. Malathi, ―Segmentation Of Different Modalitites Using Fuzzy K-Means And Wavelet ROI‖, International Journal Of Scientific &Technology Research, Vol. 8, Issue 11, pp. 996-1002, November 2019.
[14]M. Sumithra and S. Malathi, ― A Survey of Brain Tumor Segmentation Methods with Different Image Modalitites‖, International Journal of Computer Science Trends and Technology (IJCST) – Vol. 5 Issue 2, Mar – Apr 2017
[15]B.Buvaneswari and Dr.T. Kalpalatha Reddy, ―High Performance Hybrid Cognitive Framework for Bio-Facial Signal Fusion Processing for the Disease Diagnosis‖, Measurement,ISSN: 0263-2241, Vol. 140, Pp.89-99,2019.
[16]M. Sumithra and Dr. S. Malathi, ―A Brief Survey on Multi Modalities Fusion‖, Lecture Notes on Data Engineering and Communications Technologies, Springer, 35, pp. 1031-1041,2020.
[17]M. Sumithra and S. Malathi, ―A survey on Medical Image Segmentation Methods with Different Modalitites‖, International Journal of Engineering Research and Technology (IJERT) – Vol. 6 Issue 2, Mar 2018.
[18]B.Buvaneswari and Dr.T. KalpalathaReddy,―ELSA- A Novel Technique to Predict Parkinson's Disease in Bio-Facial‖,International Journal of Advanced Trends in Computer Science and Engineering, ISSN 2278-3091,Vol.8,No.1,Pp. 12- 17,2019
[19] K. Sridharan , and Dr. M. Chitra , Proficient Information Retrieval Using Trust Based Search On Expert And Knowledge Users Query Formulation System, Australian Journal of Basic and Applied Sciences, 9(23) July 2015, Pages: 755-765.
[20]B.Buvaneswari and Dr.T. Kalpalatha Reddy, ―ACPT- An Intelligent Methodology for Disease Diagnosis‖,Journal of Advanced Research in Dynamical and Control Systems,ISSN : 0974-5572,Vol.11,No.4,Pp.2187-2194,2019.
[21]Sumithra, M., Shruthi, S., Ram, S., Swathi, S., Deepika, T., "MRI image classification of brain tumor using deep neural network and deployment using web framework", Advances in Parallel Computing, 2021, 38, pp. 614–617.
[22]K. Sridharan , and Dr. M. Chitra "RSSE: A Paradigm for Proficient Information Retrieval using Semantic Web" , Life Science Journal 2013;10(7s), pp: 418-42

REFERENCES

[1] N. R. Hill, S. T. Fatoba, J. L. Oke, J. A. Hirst, C. A. O'Callaghan, D. S. Lasserson, and F. R. Hobbs, ‒Global prevalence of chronic kidney disease–a systematic review and metaanalysis,‖ PloS one, vol. 11, no. 7, p. e0158765, 2016.

[2] L. A. Stevens, J. Coresh, T. Greene, and A. S. Levey, ‒Assessing kidney function—measured and estimated glomerular filtration rate,‖ New England Journal of Medicine, vol. 354, no. 23, pp. 2473– 2483, 2006.

[3] A. S. Levey, L. A. Inker, and J. Coresh, ‒Gfr estimation: from physiology to public health,‖ American Journal of Kidney Diseases, vol. 63, no. 5, pp. 820–834, 2014. 4.A. S. Levey, L. A. Stevens, C. H. Schmid, Y. L. Zhang, A. F. Castro, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene et al., ‒A new equation to estimate glomerular filtration rate,‖ Annals of internal medicine, vol. 150, no. 9, pp. 604–612, 2009.

[4] A. S. Levey, J. P. Bosch, J. B. Lewis, T. Greene, N. Rogers, and D. Roth, ‒A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation,‖Annals of internal medicine, vol. 130, no. 6, pp. 461–470, 1999.

[5] A. Levin, P. E. Stevens, R. W. Bilous, J. Coresh, A. L. De Francisco, P. E. De Jong, K. E. Griffith, B. R. Hemmelgarn, K. Iseki, E. J. Lamb et al., ‒Kidney disease: Improving global outcomes (kdigo) ckd work group. kdigo 2012 clinical practice guideline for the evaluation and man agement of chronic kidney disease,‖ Kidney International Supplements, vol. 3, no1 pp. 1–150, 2013

[6] M. Sumithra and Dr. S. Malathi, ‒A Novel Distributed Matching Global and Local Fuzzy Clustering (DMGLFC) FOR 3D Brain Image Segmentation for Tumor Detection‖, IETE Journal ofResearch, doi.org/10.1080/03772063.2022.2027284, 2021

[7] B.Buvanswari and T.Kalpalatha Reddy, ‒A Review of EEG Based Human Facial Expression Recognition Systems in Cognitive Sciences‖ International Conference on Enenrgy, Communication,Data analytics and SoftComputing(ICECDS),CFP17M55-PRJ:978-1-5386-1886-8‖,August 2017.

[8] M. Sumithra and Dr. S. Malathi, ‖ Modified Global Flower Pollination Algorithm-based imagefusion for medical diagnosis using computed tomography and magnetic resonance imaging‖, International Journal of Imaging Systems and Technology, Vol. 31, Issue No.1, pp. 223-235, 2021

[9] K. Sridharan , and Dr. M. Chitra "SBPE: A paradigm Approach for proficient InformationRetrieval , Jokull Journal" , Vol 63, No. 7;Jul 2013

[10] M. Sumithra and Dr. S. Malathi, ‒3D Densealex NET Model with Back Propagation for Brain Tumor Segmentation‖, International Journal OfCurent Research and Review, Vol. 13, Issue 12, 2021.

[11] B.Buvaneswari and Dr.T. Kalpalatha Reddy,‒EEG signal classification using soft computing techniques for brain disease diagnosis‖,Journal of International Pharmaceutical Research ,ISSN : 1674-0440,Vol.46,No.1,Pp.525-528,2019.

[12] K. Sridharan , and Dr. M. Chitra "Web Based Agent And Assertion Passive Grading For Information Retervial", ARPN Journal of Engineering and Applied

Sciences, VOL. 10, NO. 16,September 2015 pp:7043-7048

[13] M. Sumithra and Dr. S. Malathi, Segmentation Of Different Modalitites Using Fuzzy K Means And Wavelet ROI‖, International Journal Of Scientific & Technology Research, Vol. 8, Issue 11, pp.996-1002, November 2019.

[14] M. Sumithra and S. Malathi, ― A Survey of Brain Tumor Segmentation Methods with DifferentImage Modalitites‖, International Journal of Computer Science Trends and Technology (IJCST) –Vol. 5 Issue 2, Mar – Apr 2017

[15] B.Buvaneswari and Dr.T. Kalpalatha Reddy, ―High Performance Hybrid Cognitive Framework for Bio-Facial Signal Fusion Processing for the Disease Diagnosis‖, Measurement,ISSN: 0263-2241,Vol. 140, Pp.89-99,2019.

[16] M. Sumithra and Dr. S. Malathi, ―A Brief Survey on Multi Modalities Fusion‖, Lecture Notes on Data Engineering and Communications Technologies, Springer, 35, pp. 1031-1041,2020.

[17] M. Sumithra and S. Malathi, ―A survey on Medical Image Segmentation Methods with Different Modalitites‖, International Journal of Engineering Research and Technology (IJERT) – Vol. 6 Issue 2, Mar 2018.

[18] B.Buvaneswari and Dr.T. KalpalathaReddy,―ELSA- A Novel Technique to Predict Parkinson's Disease in Bio-Facial‖,International Journal of Advanced Trends in Computer Science and Engineering, ISSN 2278-3091,Vol.8,No.1,Pp. 12-17,2019

[19] K. Sridharan , and Dr. M. Chitra , Proficient Information Retrieval Using Trust Based SearchOn Expert And Knowledge Users Query Formulation System, Australian Journal of Basic and Applied Sciences, 9(23) July 2015, Pages: 755-765.

[20] B.Buvaneswari and Dr.T. Kalpalatha Reddy, ―ACPT- An Intelligent Methodology for Disease Diagnosis‖,Journal of Advanced Research in Dynamical and Control Systems,ISSN : 0974- 5572,Vol.11,No.4,Pp.2187-2194,2019.

[21] Sumithra, M., Shruthi, S., Ram, S., Swathi, S., Deepika, T., "MRI image classification of braintumor using deep neural network and deployment using web framework", Advances in Parallel Computing, 2021, 38, pp. 614–617.

[22] K. Sridharan , and Dr. M. Chitra "RSSE: A Paradigm for Proficient Information Retrieval usingSemantic Web" , Life Science Journal 2013;10(7s), pp: 418-425