A Project Report

on

DIAGNOSING CHRONIC KIDNEYDISEASE USING MACHINE LEARNING

Submitted in partial fulfillment of the requirements for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

Computer Science & Engineering

by

J.SWETHA	174G1A0596
S.REKHA	164G1A0580
K.SAI KRISHNA	174G1A0568
K.BHARATH	174G1A05C0

Under the Guidance of

Mr.C.Sudheer Kumar, M.Tech,(Ph.D)

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(B.Tech Program Accredited by NBA)

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY:
ANANTAPURAMU
(Accredited by NAAC with ‘A’ Grade, Affiliated to JNTUA, Approved by AICTE, New Delhi)

2020-2021

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY: ANANTAPURAMU

(Accredited by NAAC with 'A' Grade, Affiliated to JNTUA, Approved by AICTE, New Delhi)

Rotarypuram Village, B K Samudram Mandal, Ananthapuramu – 515701

Certificate

This is to certify that the project report entitled DIAGNOSING CHRONIC KIDNEY DISEASE USING MACHINE LEARNING is the bonafide work carried out by **J.Swetha** bearing Roll Number 174G1A0596, **S.Rekha** bearing Roll Number 164G1A0580, **K.Sai Krishna** bearing Roll Number 174G1A0568 and **K.Bharath** bearing Roll Number 174G1A05C0 in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering** during the academic year2020-2021.

Guide	**Head of the Department**
Mr.C.Sudheer Kumar,M.Tech.,(Ph.D)	Dr. G.K.V. Narasimha Reddy ,Ph..D
Assistant Professor	Professor & HOD

Date:

Place: Rotarypuram **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion ofany task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that we have now the opportunity to express my gratitude for all ofthem.

It is with immense pleasure that we would like to express my indebted gratitude to my Guide **Mr.C.Sudheer Kumar,** M.Tech.,**(Ph.D)**, **Assistant Professor, Computer Science & Engineering**, who has guided me a lot and encouraged me in every step of the project work. We thank her for the stimulating guidance, constant encouragement and constructive criticism which have made possible to bring out this project work.

We express our deep-felt gratitude to **Dr. P. Chitralingappa, M.Tech, Ph.D, Associate Professor,** And **Mrs. M.Soumya, Assistant Professor** project coordinator valuable guidance and unstinting encouragement enable us to accomplish our project successfully in time.

We are very much thankful to **Dr. G.K.V.Narasimha Reddy, Ph.D, Professor & Head of the Department, Computer Science & Engineering,** for his kind support and for providing necessary facilities to carry out the work.

We wish to convey my special thanks to **Dr.G.BalaKrishna, Ph.D, Principal, Srinivasa Ramanujan Institute of Technology** for giving the required information in doing our project work. Not to forget, we thank all other faculty and non-teaching staff, and my friends who had directly or indirectly helped and supported us in completing our project in time.

We also express our sincere thanks to the Management for providing excellent facilities**.**

Finally, we wish to convey our gratitude to our family who fostered all the requirements and facilities that we need.

Project Associates

DECLARATION

We, Ms. J.Swetha with reg no: 174G1A0596, Ms. S.Rekha with reg no: 164G1A0580, Mr. K. Sai Krishna with reg no: 174G1A0568, Mr. K.Bharath with reg no: 174G1A05C0 students of SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY, Rotarypuram, hereby declare that the dissertation entitled "DIAGNOSING CHRONIC KIDNEY DISEASE USING MACHINE LEARNING" embodies the report of our project work carried out by us during IV year Bachelor of Technology under theguidance of Mr.C.Sudheer Kumar,M.Tech,(Ph.D) Department of CSE, SRINIVASA RAMANUJANINSTITUTE OF TECHNOLOGY, and this work has been submitted for the partial fulfilment of the requirements for the award of the Bachelor of Technology degree.

The results embodied in this project have not been submitted to any other University of Institute for the award of any Degree or Diploma.

J.Swetha	Reg no: 174G1A0596
S.Rekha	Reg no: 164G1A0580
K.Sai krishna	Reg no: 174G1A0568
K.Bharath	Reg no: 174G1A05C0

CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

CSV	Comma-separated values
SFS	Sequential feature selection
SBE	Sequential Backward Elimination
SVM	Support Vector Machine
KNN	K-Nearest Neighbour
SRS	Software Requirement Specification
UML	Unified modelling language
Numpy	Numerical Python
ML	Machine Learning

ABSTRACT

Chronic kidney disease (CKD) is a global health problem with high morbidity and mortality rate, and it induces other diseases. Since there are no obvious symptoms during the early stages of CKD, patients often fail to notice the disease. Early detection of CKD enables patients to receive timely treatment to better the progression of this disease. To achieve this machine learning model is used.

Machine learning model can effectively aid clinicians achieve this goal due to their fast and accurate recognition performance. The Chronic kidney disease data set are obtained, machine learning algorithm Random Forest were used to establish model andKNN imputation in filling missing values of dataset, that could be applicable to more complicated clinical data for disease diagnosis.

CHAPTER 1

INTRODUCTION

1.1 CKD Prediction

Chronic kidney disease (CKD) is a global public health problem affecting approximately 10% of the world’s population. The percentage of prevalence of CKD in China is 10.8% , and the range of prevalence is 10%-15% in the United States. According to another study, this percentage has reached 14.7% in the Mexican adult general population. This disease is characterized by a slow deterioration in renal function, which eventually causes a complete loss of renal function. CKD does not show obvious values in its early stages. Therefore, the disease may not be detected until the kidney loses about 25% of its function. In addition, CKD has high morbidity and mortality, with a global impact on the human body. It can induce the occurrence of cardiovascular disease. CKD is a progressive and irreversible pathologic syndrome. Hence, the prediction and diagnosis of CKD in its early stages is quite essential, it may be able to enable patients to receive timely treatment to ameliorate the progression of the disease. Machine learning refers to a computer program, which calculates and deduces the information related to the task and obtains the characteristics of the corresponding pattern. This technology can achieve accurate and economical diagnoses of diseases; hence, it might be a promising method for diagnosing CKD. It has become a new kind of medical tool with the development of information technology and has a broad application prospect because of the rapid development of electronic health record. In the medical field, machine learning has already been used to detection human body status, analyze the relevant factors of the disease and diagnose various diseases. For example, the models built by machine learning algorithms were used to diagnose heart disease diabetes and retinopathy acute kidney injury cancer and other diseases.In these models, algorithms based on regression, tree, probability, decision surface and neural network were often effective. In the field of CKDdiagnosis,

Hodneland et al. utilized image registration to detect renal morphologic changes. Vasquez-

Morales et al. established a classifier based on neural network using large-scale CKD data, and the accuracy of the model on their test data was 95% In addition, most of the previous studies utilized the CKD data set that was obtained from the UCI machine learning repository. Chen et al. used k-nearest neighbor (KNN), support vector machine (SVM) and soft independent modelling of class analogy to diagnose CKD, KNN and SVM achieved the highest accuracy of 99.7%. In addition, they used fuzzy rule-building expert system, fuzzy optimal associative memory and partial least squares discriminant analysis to diagnose CKD, and therange of accuracy in those models was 95.5%-99.6%. Their studies have achieved good results in the diagnosis of CKD. In the above models, the mean imputation is used to fill in the missing values and it depends on the diagnostic categories of the samples. Asa result, their method could not be used when the diagnostic results of the samples are unknown. In reality, patients might miss some measurements for various reasons before diagnosing. In addition, for missing values in categorical variables, data obtained using mean imputation might have a large deviation from the actual values. For example, for variables with only two categories, we set the categories to 0 and 1, but the mean ofthe variables might be between 0 and 1. Polat et al. developed an SVM based on feature selection technology, the proposed models reduced the computational costthrough feature selection, and the range of accuracy in those models was from 97.75%-98.5%. J. Aljaaf et al. used novel multiple imputation to fill in the missing values, and then MLP neural network (MLP) achieved an accuracy of 98.1%. Subas et al. used MLP, SVM, KNN, C4.5 decision tree and random forest (RF) to diagnose CKD, and the RF achieved an accuracy of 100%. In the models established by Boukenze et al., MLP achieved the highest accuracy of 99.75%. The studies of focus mainly on the establishment of models and achieve an ideal result. However, a complete process of filling in the missing values is not described in detail, and no feature selection technology is used to select predictors as well. Almansour et al. used SVM and neural network to diagnose CKD, and the accuracy of the models was 97.75% and 99.75%, respectively. In the models established by Gunarathneet al., zero was used to fill out the missing values and decision forest achieved the best performance with the accuracy was 99.1%.

1.2 Problem Definition

Chronic kidney disease (CKD) is a global public health problem affecting approximately 10% of the world's population. The percentage of prevalence of CKD in China is 10.8%, and the range of prevalence is 10%-15% in the United States . According to another study, this percentage has reached 14.7% in the Mexican adult general population. This disease is characterised by a slow deterioration in renal function, which eventually causes a complete loss of renal function. CKD does not show obvious values in its early stages. Therefore, the disease may not be detected until the kidney loses about 25% of its function. In addition, CKD has high morbidity and mortality, with a global impact on the human body. It can induce the occurrence of cardiovascular disease. CKD is a progressive and irreversible pathologic syndrome. Hence, the prediction and diagnosis of CKD in its early stages is quite essential, it may be able to enable patients to receive timely treatment to ameliorate the progression of the disease.

1.3 Project Purpose

Machine learning refers to a computer program, which calculates and deduces the information related to the task and obtains the characteristics of the corresponding pattern. This technology can achieve accurate and economical diagnoses of diseases; hence, it might be a promising method for diagnosing CKD. It has become a new kind of medical tool with the development of information technology and has a broad application prospect because of the rapid development of electronic health record. In the medical field, machine learning has already been used to detection human body status analyze the relevant factors of the disease and diagnose various diseases. For example, the models built by machine learning algorithms were used to diagnose heart disease, diabetes and retinopathy, acute kidney injury cancer and other diseases. In these models, algorithms based on regression, tree, probability, decision surface and neural network were often effective, In the field of CKD diagnosis.

1.4 Project Features

The features of Chronic Kidney Disease Prediction Using Machine Learning are as follows.

- This Project will predict the diseases of the patients based on the values and other general information using the datasets.
- This is done based on the previous datasets of the hospitals so after comparing it can provide up to 80% of accurate results, and the project is still developing further to get the 100% accurate results.
- With the help of Disease prediction, it can predict the disease of the patient and can solve various problems and prevents from various aspects.
- It provides security for the system so that no one can break into that and no one can make any changes in the system.
- The disease is predicted using the algorithms and the user has to enter the values from the given drop-down menu, in order to get correct accuracy, the user has to enter all the values.
- Here we can easily prepare the data and transform that data into algorithm, which will reduce the overall work of the project.
- To make user more application friendly rather than discussing with others for their disease.
- It provides the necessary options to choose from the types andattributes.
- Here the user has to register first, in order to use the prediction and then login to the system using the credentials such as username and password.
- Once user open the system to login user needs to register by clicking on register/signup button.
- After which user needs to provide some basic details of signup and then the details of user are saved in system

CHAPTER-2

LITERATURE SURVEY

2.1 Machine Learning

Tom Mitchell states machine learning as “A computer program is said to learn from experience and from some tasks and some performance on, as measured by, improves with experience”. Machine Learning is combination of correlations and relationships, most machine learning algorithms in existence are concerned with finding and/or exploiting relationship between datasets. Once Machine Learning Algorithms can pinpointon certain correlations, the model can either use these relationships to predict future observations or generalize the data to reveal interesting patterns. In Machine Learning there are various types of algorithms such as Regression, Linear Regression, Logistic Regression, Naive Bayes Classifier, Bayes theorem, KNN (K-Nearest Neighbor Classifier), Decision Tress, Entropy, ID3, SVM (Support Vector Machines), K-means Algorithm, Random Forest and etc.,

The name machine learning was coined in 1959 by Arthur Samuel. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

Machine learning tasks Machine learning tasks are typically classified into several broad categories:

Supervised learning:

The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs. As special cases, the input signal can be only partially available, or restricted to special feedback.

Semi-supervised learning:

The computer is given only an incomplete training signal: a training set with some (often many) of the target outputs missing.

Active learning:

The computer can only obtain training labels for a limited set of instances(based on a budget), and also has to optimize its choice of objects to acquire labels for. When used interactively, these can be presented to the user for labelling.

Unsupervised learning:

No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

Reinforcement learning:

Data (in form of rewards and punishments) are given only as feedback to the program's actions in a dynamic environment, such as driving a vehicle or playing a game against an opponent.

2.1.1 Features of ML

- It is nothing but automating the Automation.
- Getting computers to program themselves.

- Writing Software is bottleneck.

- Machine leaning models involves machines learning from data without the help of humans or any kind of human intervention.

- Machine Learning is the science of making of making the computers learn and act like
- humans by feeding data and information without being explicitly programmed.

- Machine Learning is totally different from traditionally programming, here data and output is given to the computer and in return it gives us the programwhich provides solution to the various problems. Below is the figure 2.1.

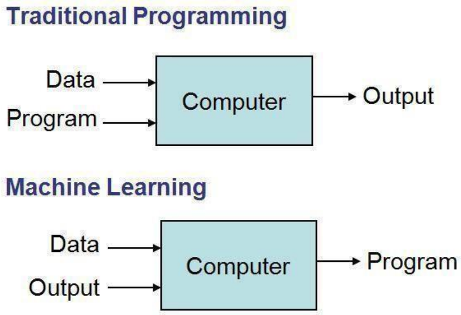
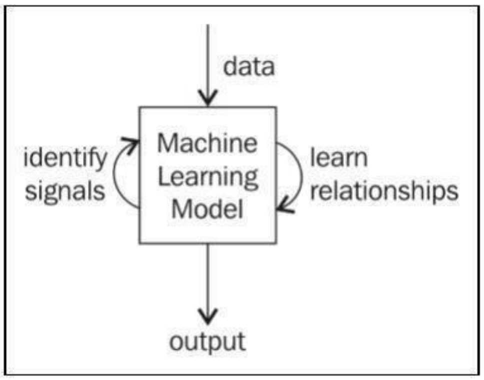


Fig 2.1: Traditional Programming vs Machine Learning

- Machine Learning is a combination of Algorithms, Datasets, and Programs.

- There are Many Algorithms in Machine Learning through which we will provide us the exact solution in predicting the disease of the patients.



An overview of machine learning models

Fig 2.2: Machine Learning Model

How Does Machine Learning Works?, shown above fig 2.2

- Solution to the above question is Machine learning works by taking in data, finding relationships within that data and then giving the output.
- There are various applications in which machine learning is implemented such as Web search, computing biology, finance, e-commerce, space exploration, robotics, social networks, debugging and much more.
- There are 3 types of machine learning supervised, unsupervised, and reinforcement.

2.1 EXISTING SYSTEM

Prediction using traditional methods and models involves various risk factors and it consists of various measures of algorithms such as datasets, programs and much more to add on. High-risk and Low-risk patient classification is done on the basis of the tests that are done in group. But these models are only valuable in clinical situations and not inbig industry sector. So, to include the disease predictions in various health related industries, we have used the concepts of machine learning and supervised learning methods to build the predictions system.

After doing the research and comparison of all the algorithms and theorems of machine learning we have come to conclusion that all those algorithms such as Decision Tree, KNN, Naïve Bayes, Regression and Random Forest Algorithm all are important in building a Chronic Kidney Disease Prediction system which predicts the disease of the patients from which he/she is suffering from and to do this we have used some performance measures like ROC, KAPPA Statistics, RMSE, MEA and various other tools.

After using various techniques such as neural networks to make predictions of the diseases and after doing that we come to conclusion that it can predicts up to 90% accuracy rate after doing the experimentation and verifying the results. The information of patient statistics, results, disease history in recorded in EHR, which enables toidentify the potential data centric solution, which reduces the cost of medical case studies. Existing system can predict the disease but not the sub type of the disease and itfails to predict the condition of the people, the predictions of disease have been indefinite and non- specific.

2.2 PROPOSED SYSTEM

The proposed system of Chronic Kidney Disease Prediction using machine learn that we have used many techniques and algorithms and all other various tools to build a system which predicts the disease of the patient using the values and by taking those values we are comparing with the system’s dataset that is previously available. By taking those datasets and comparing with the patient’s disease we will predict the accurate.

percentage disease of the patient. The dataset and values go to the prediction model of the system where the data is pre-processed for the future references and then the feature selection is done by the user where he will enter the various values.

Then the classification of those data is done with the help of algorithms and techniques such Random Forest and etc.

Then the data goes in the recommendation model, there it shows the risk analysis that is involved in the system and it also provides the probability estimation of the system such that it shows the variousprobability like how the system behaves whenthere are n number of predictions are done and it also does the recommendations for the patients from their final result and also fromtheir values like it can show what to use andwhat not to use from the given datasets and the final results.

Here we have combined the overall structure and unstructured formof data forthe overall risk analysis that is required for doingthe prediction of the disease.Using the structured analysis, we can identify the chronic types of disease in a particular region and particular community. In unstructured analysis we select the features automatically with the help of algorithms and techniques.

This system takes values from the user and predicts the disease accordingly based on the values that it takes and also from the previous datasets, it also helps in continuous evaluation of viral diseases, heart rate, blood pressure, sugar level and much more which is in the system and along with other external values its predicts the appropriate and accurate disease.

So, to include the disease predictions in various health related industries, we have used the concepts of machine learning and supervised learning methods to buildthe predictions system. .

2.3 SOFTWARE DESCRIPTION

2.3.1 PYTHON

Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of its features support functional programming and aspect-oriented programming (including by metaprogramming and metaobjects. Many other paradigms are supported via extensions, including design by contract and logic programming.

Python uses dynamic typing and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution.

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development Domain. I will list down some of the key advantages of learning Python:

- Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- Python is a Beginner's Language − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

Python's developers strive to avoid premature optimization, and reject patches to non-critical parts of CPython that would offer marginal increases in speed at the cost ofclarity. When speed is important, a Python programmer can move time-critical functions to extension modules written in languages such as C, or use PyPy, a just-in-time compiler. Cython is also available, which translates a Python script into C and makes direct C- level API calls into the Python interpreter.

An important goal of Python's developers is keeping it fun to use. Python's design offers some support for functional programming in the Lisp tradition. It has filter, map, and reduce functions, list comprehensions, dictionaries, sets, and generator expressions. The standard library has two modules (itertools and functools) that implement functional tools borrowed from Haskell and Standard ML.

2.3.2 BENEFITS OF PYTHON

- Presence of Third-Party Modules
- Extensive Support Libraries
- Open Source and Community Development
- Learning Ease and Support Available
- User-friendly Data Structures
- Productivity and Speed
- Highly Extensible and Easily Readable Language.

2.3.3 TKINTER INTERFACE

Tkinter is a Python binding to the Tk GUI toolkit. It is the standard Python interface tothe Tk GUI toolkit and is Python's de facto standard GUI. Tkinter is included with standard Linux, Microsoft Windows and Mac OS X installs of Python. The name Tkinter comes fromTk interface. Tkinter was written by Fredrik Lundh. Tkinter is free software released undera Python license.

As with most other modern Tk bindings, Tkinter is implemented as a Python wrapper around a complete Tool Command Language (TCL) interpreter embedded in the Python interpreter. Tkinter calls are translated into Tcl commands which are fed to this embedded interpreter, thus making it possible to mix Python and TCL in a single application.In Tkinter, the Frame widget is the basic unit of organization for complex layouts.

Python offers multiple options for developing GUI (Graphical User Interface).

Out of all the GUI methods, tkinter is most commonly used method. It is a standard Python interfaceto the Tk GUI toolkit shipped with Python.
Python with tkinter outputs the fastest and easiest way to create the GUI applications.

To create a tkinter:

Importing the module – tkinter Create the main window (container)

Add any number of widgets to the main windowApply the event Trigger on the widgets.

Importing tkinter is same as importing any other module in the python code.
Note that the name of the module in Python 2.x is ‘Tkinter’ and in Python 3.x is ‘tkinter’.

CHAPTER-3
REQUIREMENTS ANALYSIS

3.1 FUNCTIONAL REQUIREMENTS

A Functional requirement defines a function of a system or its component. A function is described as a set of inputs, the behaviour, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish.

Behavioural requirements describing all cases where the system uses the functional requirements arecaptured in use cases. Functional requirements are supported by non- functional requirements (also known as quality requirements), which impose constraints on the design or implementation (such as performance requirements, security, or reliability).

As defined in requirements engineering, functional requirements specify particular results of a system. This should be contrasted with non-functional requirements which specify overall characteristics such as cost and reliability. Functional requirements drive the application architecture of a system, while non-functional requirements drive the technical architecture of a system.

- Functional Requirements concerns with the specific functions delivered by the system.So, Functional requirements are statements of the services that the system must provide.
- The functional requirements of the system should be both complete and consistent
- Completeness means that all the services required by the user should be defined.
- Consistency means that requirements should not have any contradictory definitions.
- The requirements are usually described in a fairly abstract way. However, functional system requirements describe the system function in details, its inputs and outputs, exceptions and so on.

- Take user id and password match it with corresponding file entries. If a match is found then continue else raise an error message.

3.2 NON-FUNCTIONAL REQUIREMENTS

- Non-functional Requirements refer to the constraints or restrictions on the system. They may relate to emergent system properties such as reliability, response time and store occupancyor the selection of language, platform, implementation techniques and tools.

- The non-functional requirements can be built on the basis of needs of the

user, budget constraints, organization policies and etc.

1. **Performance requirement:** All data entered shall be up to mark and no flawsshall be there for the performance to be 100%.
2. **Platform constraints:** The main target is to generate an intelligent system topredict the adult height.
3. **Accuracy and Precision**: Requirements are accuracyand precision of the data
4. **Modifiability:** Requirements about the effort required to make changes in the software. Often, the measurement is personnel effort (person- months).
5. **Portability:** Since mobile phone is handy so it is portable and can be carried and used whenever required.
6. **Reliability**: Requirements about how often the software fails. The definition of a failure must be clear. Also, don't confuse reliability with availability whichis quite a different kind of requirement. Be sure to specify the consequencesof software failure, how to protect from failure, a strategy for error Prediction,and a strategy for correction.
7. **Security**: One or more requirements about protection of your system and its data.

8. **Usability:** Requirements about how difficult it will be to learn and operate the time or similar metrics.system. The requirements are often expressed in learning

ACCESSIBILITY:

Accessibility is a general term used to describe the degree to which a product, device, service, or environment is accessible by as manypeople as possible. In our project peoplewho have registered with the cloud can access the cloud to store and retrieve their datawith the help of a secret key sent to their email ids. User interface is simple and efficientand easy to use.

MAINTAINABILITY:

In software engineering, maintainability is the ease with which a software product can bemodified in order to include new functionalities can be added in the project based ontheuser requirements just by adding the appropriate files to existing project using .net and programming languages. Since the programming is very simple, it is easier to find and correct the defects and to make the changes in the project.

SCALABILITY:

System is capable of handling increase total throughput under an increased load when resources (typically hardware) are added. System can work normally under situations such as low bandwidth and large number of users.

PORTABILITY:

Portabilityis one of the key concepts of high-level programming. Portability is the software code base feature to be able to reuse the existing code instead of creating new code when moving software from an environment to another. Project can be executed under different operation conditions provided it meet its minimum configurations. Only system files and dependant assemblies would have to be configured in such case.

VALIDATION:

It is the process of checking that a software system meets specifications and that it fulfils its intended purpose. It may also be referred to as software quality control. It is normally the responsibility of software testers as part of the software development lifecycle.Software validation checks that the software product satisfies or fits the intended use (high-level checking), i.e., the software meets the user requirements, not as specification artefacts or as needs

3.3 HARDWARE REQUIREMENTS

- System : Pentium 4, Intel Core i3, i5, i7 and 2 GHz Minimum
- RAM: 512Mb or above
- Hard Disk : 10 GB or above
- Input Device : Keyboard and Mouse
- Output Device : Monitor or PC

3.4 SOFTWARE REQUIREMENTS

- Operating System: Windows 7, 10 or Higher Versions
- Platform : Jupiter Notebook
- Front End : Python falsk
- Back End : Python and Files
- Programming Lang: Python
- Database : Mysql

CHAPTER-4

DESIGN

4.1 DESIGN GOALS

The Design goals consist of various design which we have implemented in our system Chronic Kidney Disease Predictionusing machine learning. This system has built with various designs suchas data flow diagram, sequence diagram, class diagram, use case diagram, component diagram, activity diagram, state chart diagram, deployment diagram. After doing these various diagrams and based on these diagrams we have done our project.

We have designed our system in such a way that whenever user log in into the system, the user has to register to the system, and new user cannot use the system without registering in the system. After that for registration the user requires basic credentials such as username, age, email, phone, password. Then the user has to login to the system using the same username and password. Here are the things that this system can perform.

a. Entering Values

b. Disease Prediction

Entering Values:

Once user successfully logged in to the system then he/she has to select the values from the given drop-down menu.

Disease prediction:

The predictive model predicts the disease of a person he might have,based on the user entered values.

.

4.2 SYSTEM ARCHITECTURE

Chronic Kidney Disease Prediction using machine learning predicts the presence of the disease for the user based on various values and the information the user gives such as sugar level, haemoglobin level and many more such general information throughthe values. The architecture of the system Chronic Kidney Disease Prediction using machine learning consist of various datasets through which we will compare the values of the user and predicts it, then the datasets are transformed into the smaller sets and from there

The purpose of system **architecture** activities is to define a comprehensive solution based on principles, concepts, and properties logically related to and consistent with each other. The solution architecture has features, properties, and characteristics which satisfy, as far as possible, the problem or opportunity expressed by a set of system requirements (traceable to mission/business and stakeholder requirements) and life cycle concepts (e.g., operational, support) and which are implementable through technologies (e.g., mechanics, electronics, hydraulics, software, services, procedures, human activity).

System Architecture is abstract, conceptualization-oriented, global, and focused to achieve the mission and life cycle concepts of the system. It also focuses on high-level structure in systems and system elements. It addresses the architectural principles, concepts, properties, and characteristics of the system-of-interest. It may also be applied to more than one system, in some cases forming the common structure, pattern, and set of requirements for classes or families of similar or related systems.

Design decisions and technological solutions are selected according to performance criteria and non-functional requirements, such as operational conditions and life cycle constraints (e.g., environmental conditions, maintenance constraints, realization constraints, etc.), as illustrated in Figure 1.

Creating intermediate models, such as logical architecture models, facilitates the validation of functional, behavioral, and temporal properties of the system against the system requirements that have no major technological influence impacts during the life of the system, the physical interfaces, or the technological layer without completely questioning the logical functioning of the system. .

it gets classified based on the classification algorithms later on the classified data is then processed into the machine learning technologies through which the data gets processed and goes in to the Chronic Kidney Disease Prediction model using all the inputs from the user that is mentioned above.

Then after user entering the above information and overall processed data combines and compares in the prediction model of the system and finally predicts the disease. An architecture diagram is a graphical representation of a set of concepts, that are part of an architecture, including their principles, elements and components. The diagram explains about the system software in perception of overview of the system. Shown in below this fig 4.1

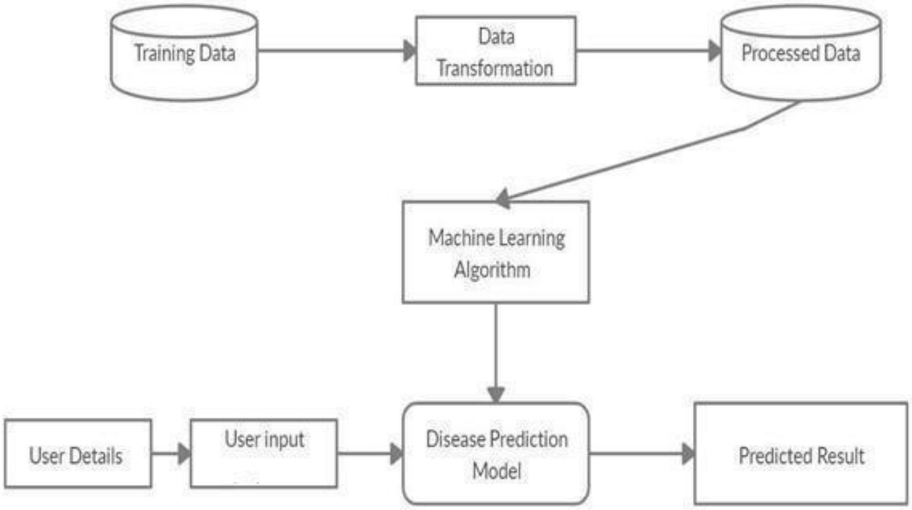


Fig 4. 1 System Architecture

4.3 DATA FLOW DIAGRAM

The dataflow diagram of the project Chronic Kidney Disease Prediction using machine learning consist of all the various aspects a normal flow diagram requires. This dataflowdiagram shows how from starting the model flows from one step to another, like he enterinto the system then enters all the information’s and all other general information alongwith the values that goes into the system, compares with the prediction model and if trueis predicts the appropriate results otherwise it shows in below fig 4.2, the details where the user if gone wrong while entering the information.

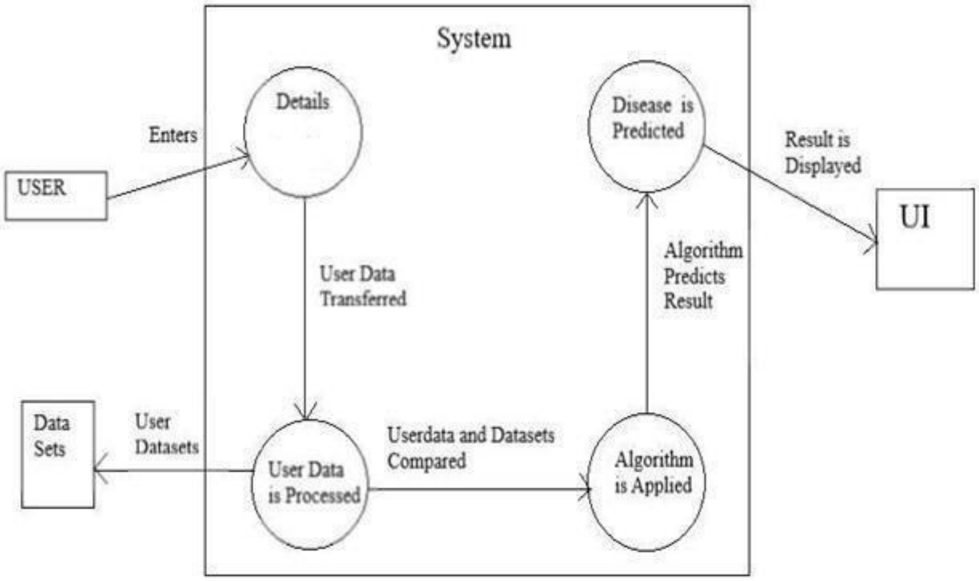


Fig 4.2 Data Flow Diagram

4.4 CLASS DIAGRAM

Chronic Kidney Disease Predictionusing machine learning consist of class diagram thatall the other application that consists the basic class diagram, here the class diagram is the basic entitythat is required in order to carry on with the project. Class diagram consistinformation about all the classes that is used and all therelated datasets, and all the othernecessary attributes and their relationships with other entities, all these information is necessary inorder to use the concept of the prediction, where the user will enter all necessary information such as user name, email, phone number, and many more attributes that is required in order to login intothe system and using the files concept we will store the information of the users whoare registering into the system and retrieves those information later while logging into the system.Shown in below fig 4.3

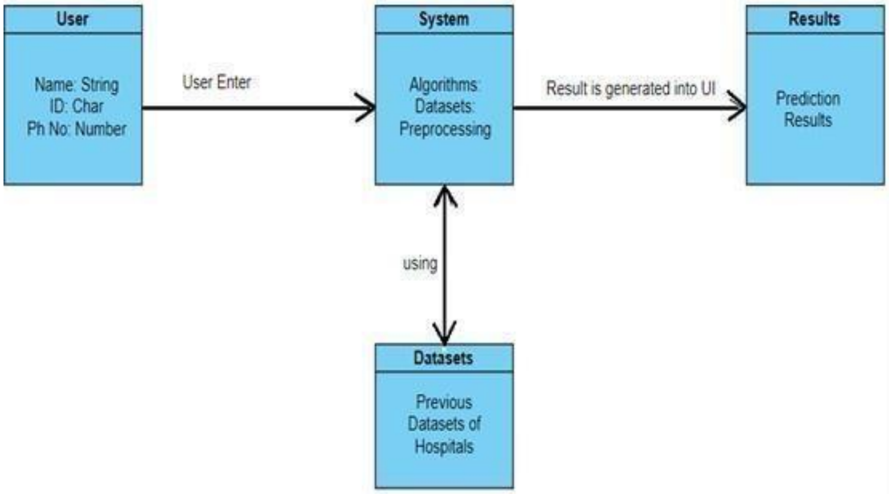


Fig 4.3 Class Diagram

4.5 SEQUENCE DIAGRAM

The Sequence diagram of the project Chronic Kidney Disease Prediction using machine learning consist of all the various aspects a normal sequence diagram requires. This sequence diagram show show from starting the model flows from one step to another, like he enter into the system then enters all the information's and all other general information along with the values that goes into the system, compares with the prediction model and if true is predicts the appropriate results otherwise it shows the details where the user if gone wrong while entering the information's and it also shows the appropriate precautionary measure for the user to follow. Here the sequence of all the entities are linked to each other where the user gets started with the system. Shown in below fig 4.4

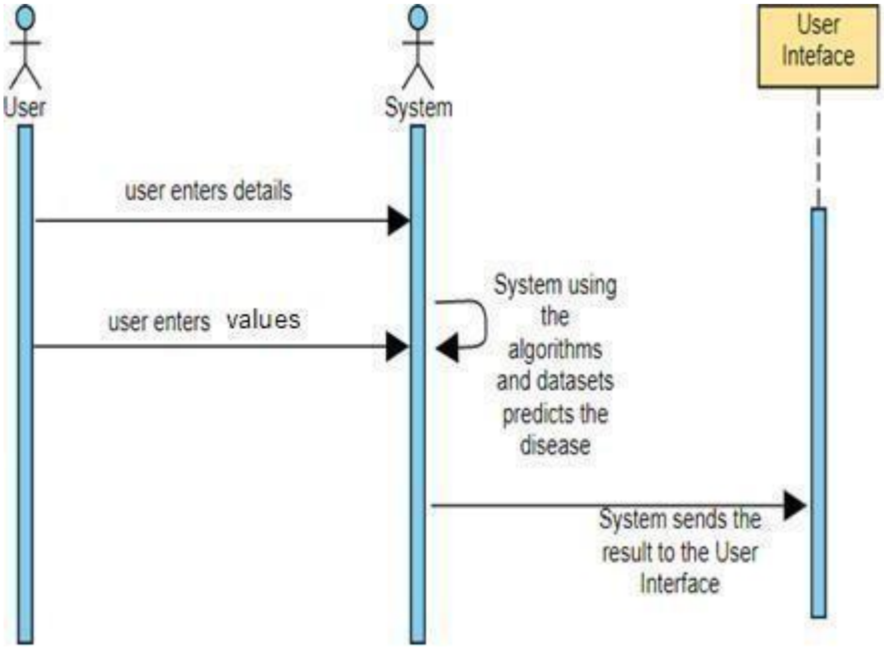

User
System
User Inteface
user enters details
user enters values
System using the algorithms and datasets predicts the disease
System sends the result to the User Interface

Fig 4.4 Sequence Diagram

4.6 USE CASE DIAGRAM

The Use Case diagram of the project Chronic Kidney Disease Prediction using machine learning consist of all the various aspects a normal use case diagramrequires. This use case diagram shows how from starting the model flows from one step to another, like he enter into the system hen enters all the information's and all other general information along with the values that goes into the system, compares with the prediction model and if true is predicts the appropriate results otherwise it shows the details where the user if gone wrong while entering the information's and it also shows the appropriate precautionary measure for the user to follow. Here the use case diagram of all the entities are linked to each other where the user gets startedwith the system.

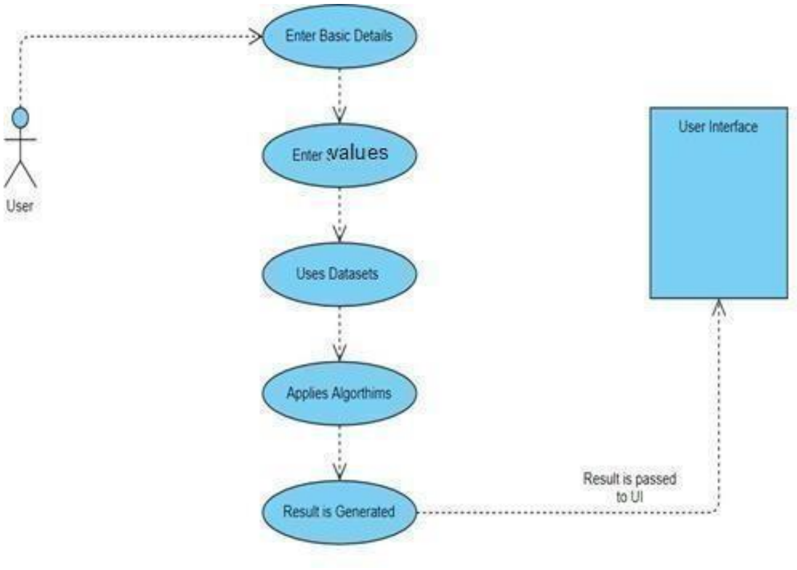


Fig 4.5 Use Case Diagram

4.7 ACTIVITY DIAGRAM

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another. Here in this diagram the activity starts from user where the user registers into the system then login using the credentials and then the credentials are matched in the system and if its true, then the user proceeds to the prediction phase where the prediction happens. Then finally after processing the data from datasets the analysis will happen then the correct result will be displayed that is nothing but the Output.

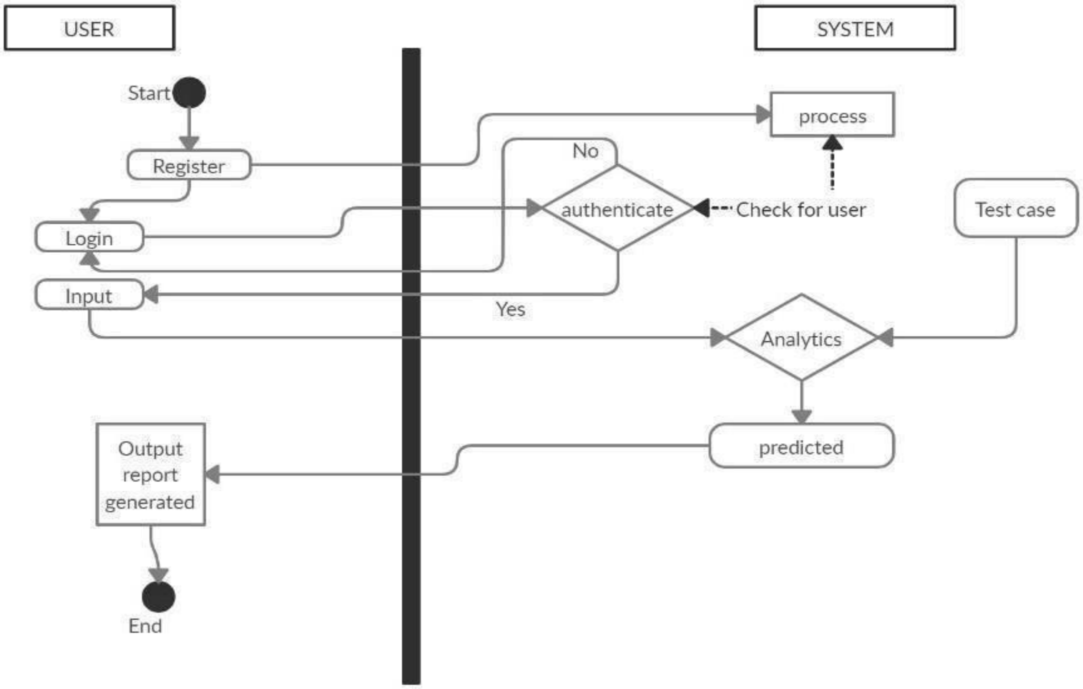


Fig 4.6 Activity Diagram

4.8 COMPONENT DIAGRAM

A component diagram, also known as a UML component diagram, describes the organization and wiring of the physical components in a system. Component diagrams areoften drawn to help model implementation details and double-check that every aspectofthe system's required function is covered by planned development. Here component diagram consists of all major components that is used to built a system. So, Design, Algorithm, File System and Datasets all are linked to one another. Datasets are used tocompare the results and algorithm is used to process those results and give a correct accuracy and design UI is used to show the result in an appropriate way in the system andfile system is used to store the user data. So, like this all components are interlinked to each other.

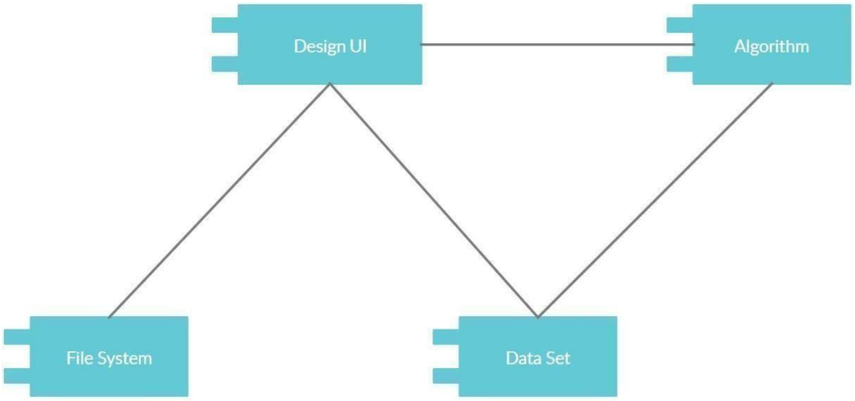


Fig 4.7 Component Diagram

4.9 STATE CHART DIAGRAM

A State chart diagram describes the behaviour of a single object in response to a series of events in a system. Sometimes it's also known as a Harel state chart or a state machine diagram. This UML diagram models the dynamic flow of control from state to state of a particular object within a system. It is similar to activity diagram but here there are only few rules like how it starts and how it end all are denoted with the help of the symbol given below, the system starts with the registration and then login comes, if the login is successful then it will go to the next step and if login is incorrect then comes back to same page stating incorrect details. After the successful login the user needs to enter the values and then press the prediction button, at the same time the backend process will do their work and the correct result is predicted.

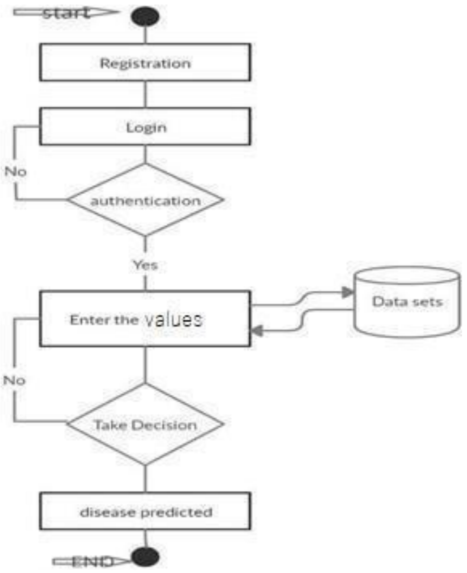


Fig 4.8 State Chart Diagram

4.10 COLLABORATION DIAGRAM

A collaboration diagram, also known as a communication diagram, is an illustration ofthe relationships and interactions among software objects in the Unified Modelling Language(UML). These diagrams can be used to portray the dynamic behaviour of a particular use case and define the role of each object. Here this diagram shows how all the models are connected to show the correct result starting from user, where he opens the system thenusing the system he does registration and that registration data is saved into file system and the using those data user logs in to the system and then he providesall the necessaryinformation in order to get the accurate result, then system evaluates the user entered information and finally gives the correct result,shown in the below fig 4.9

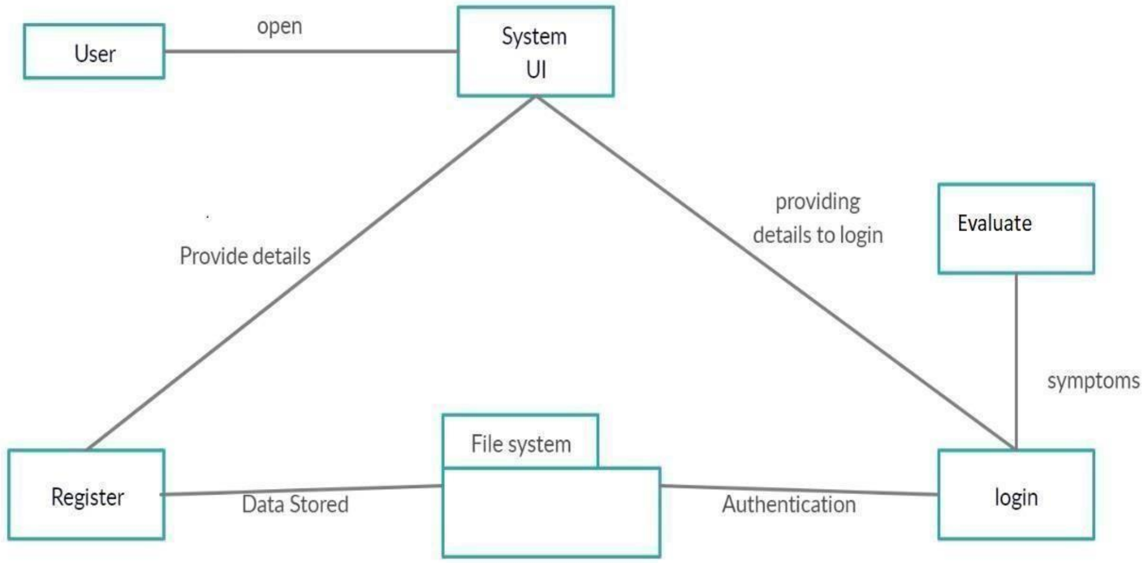


Fig 4.9 Collaboration Diagram

4.11 DEPLOYMENT DIAGRAM

A deployment diagram shows the configuration of run time processing nodes and the components that live on them. Deployment diagrams is a kind of structure diagram usedin modelling the physical aspects of an object-oriented system. Here the deployment diagram show the final stage of the project and it also shows how the model looks like after doing all the processes and deploying in the machine. Starting from the system how it processes the user entered information and then comparing that information with the help of datasets, then training and testing those data using the algorithms such as decision tree, naïve Bayes, random forest. Then finally processing all those data and information the system gives the desired result in the interface, shown in the below fig 4.10

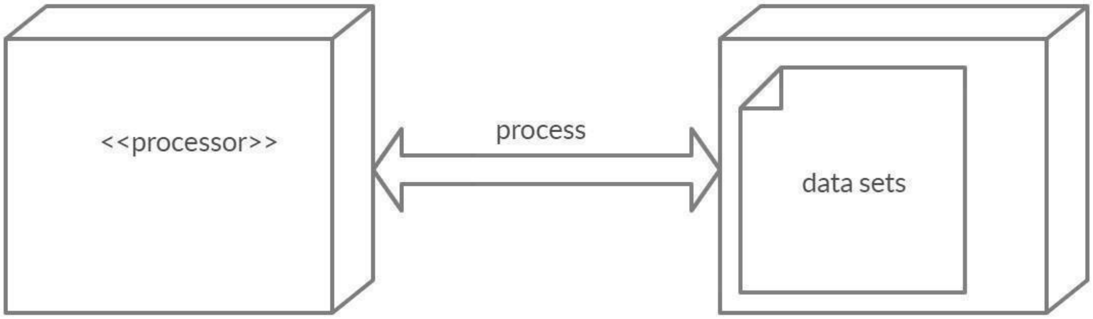


Fig 4.10 Deployment Diagram

4.12 INTERFACE AND FRAMEWORK DIAGRAM

Flask Web Framework?

A web framework is an architecture containing tools, libraries, and functionalities suitable to build and maintain massive web projects using a fast and efficient approach. Shown in fig 4.11, They are designed to streamline programs and promote code reuse. To create the server-side of the web application, you need to use a server-side language. Python is home to numerous such frameworks, famous among which are Django and Flask. Python Flask Framework is a lightweight micro- framework based on Werkzeug, Jinja2. It is called a micro framework because it aims to keep its core functionality small yet typically extensible to cover an array of small and large applications. Flask Framework depends on two external libraries: The Jinja2 template,Werkzeug WSGI toolkit. Even though we have a plethora of web apps at our disposal,Flask tends to be better suited due to –

- Built-in development server, fast debugger.
- Integrated support for unit testing.
- RESTful request dispatching.
- Jinja2 Templating.
- Support for secure cookies.
- Lightweight and modular design allows for a flexible framework.

Initialization: flask applications must create an application instance. The web serverpasses all the requests it receives from clients to objects for handling using a protocol for WSG from flask import Flask app = Flask (name) (An application instance is an object of class Flask.)

The layout of the Python Flask Framework

- Module Init – (project_root/app_name/admin/ init .py) – required to enable the app
- Module URL – (project_root/app_name/admin/url.py) – Url definitions of each

module

○ App root Init – (project_root/app_name/_init_.py) – Not necessary to define the entire app within_____init .py

○ Module Views – (project_root/app_ame/admin/views.py) – Defines views foreach module. Separate '.py.' Files as the project scale to ensure they are accessible to URLs.

Module Templates – (project_root/app_name/admin/templates/admin/main.html) – Normal template folder. HTTP Methods

Request

To process incoming data in Flask, you need to use the request object, including mime-type, IP address, and data. HEAD: Un-encrypted data sent to server w/o response.

GET

Sends data to the server requesting a response body.

POST

Read form inputs and register a user, send HTML data to the server are methods handled by the route.

Flask attaches methods to each route so that different view functions can handle different request methods to the same URL.

Response

Flask invokes a view function. It has to return a response value to the client. HTTP requires it to be more than a string response, a status code.

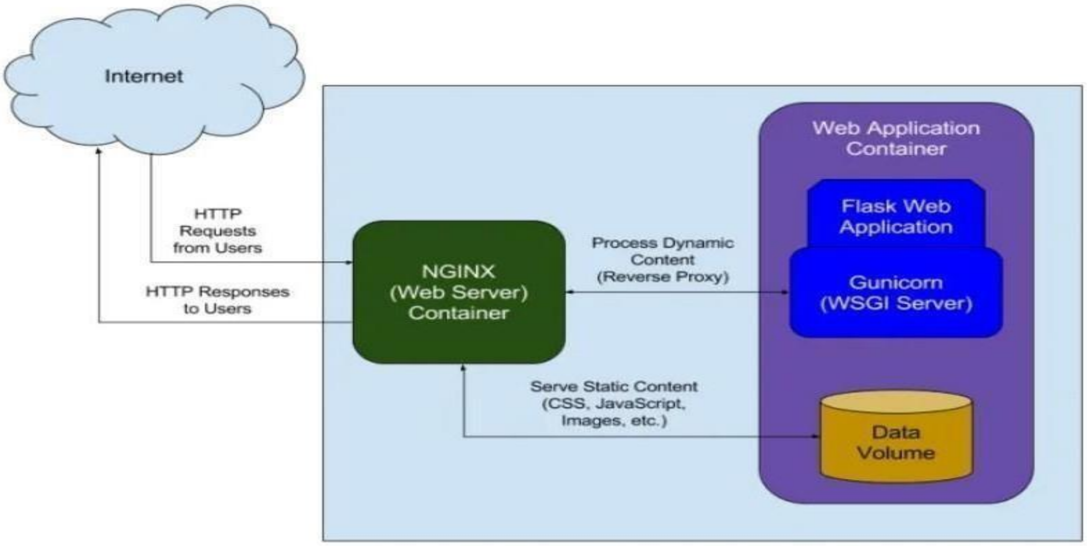

Internet
HTTP Requests from Users
HTTP Responses to Users
NGINX (Web Server) Container
Process Dynamic Content (Reverse Proxy)
Serve Static Content (CSS, JavaScript, Images, etc.)
Web Application Container
Flask Web Application
Gunicorn (WSGI Server)
Data Volume

Fig 4.11 Interface And Framework Diagram

CHAPTER-5
IMPLEMENTATION

5.1 OVERVIEW

The project Chronic Kidney Disease Prediction using Machine Learning is developed to overcome generaldisease in earlier stages as we all know in competitive environment of economic development the mankind has involved so much that he/she is not concerned about health according to research there are 40% peoples how ignores aboutgeneral disease which leads to harmful disease later. The Project “Chronic Kidney Disease Predictionusing Machine Learning” isimplemented using python completely.

Even the interface of this project is done using python’s library interface called Tkinter.Here first the user needs to register into the system in order to use the prediction, user needs to register with username, email-id, phone, age and password.

All these values are stored into the file system respectively, then user has option to move forward or leave, then user needs to login to the system using the username and password which he/she provided during the time of registration.

If he/she enter incorrect username and correct password then the error message will prompt stating incorrect username and if he/she enters incorrect password and correct username then the error message will prompt stating incorrect password, so both username and password is necessary in order to login to the system.

After logging inthe user needs to the name and needs to select thevalues from given drop-down menu, for more accurate result the user needs to enter allthe given values, then thesystem will provide the accurate result. This prediction is basically done with the help of 3 algorithms of machine learning such as Decision Tree,Random Forest and Naïve Bayes.

When user enter all the values then he needs to press the buttons of respective algorithm, for example there are 3 buttons for 3 algorithms, ifuser enters all values and presses only Random forest's button then the result will be provided only calculating using that algorithm, like this we have used 3 algorithms to provide more clear picture of the results and user needs to be satisfied with his Predicted result.

The project is designed user friendly and also secure to use ever user requires a authentication to enter into

the system after which it provides the result based on the user input let me explain the

complete implementation and working of project step wise below

- Once user open the system to login user needs to register by clicking on register/signup button
- After which user needs to provide some basic details of signup and then the details of user are saved in system
- Then user needs to login to have a checkup of his/her health
- When user tries to login if he provides wrong user name the system will provide a prompt message stating that the user is not found.
- And if user tries to enter the wrong password the system will prompt stating that password is in correct hence the user needs to enter the correct user id and password to get in to the system.
- After user enters the system user has to provide the values which he/she is going through based on which we have several algorithms which predict thedisease and also displays the percentage of accuracy.
- The user needs to enter all the columns of values to get the accurate result. Data collection and dataset preparation This will involve collection of medical information from various sources like hospitals, then pre-processing is applied on dataset which will remove all the unnecessary data and extract important features from data.

Developing a probabilistic model for Chronic Kidney Disease Prediction in this step probabilistic model and deep learning approach based on Random Forest Algorithm is to be developed it will run effectively on extensive databases of healthcare.And generate decision tree also it can deal with a huge number of information variables withoutvariable deletion.

Training and experimentation on datasets The Chronic Kidney Disease Prediction model will be trained on the dataset of diseases to do the prediction accurately and produce Confusion matrix. In this project 3 different algorithms were used –

- Random Forest Algorithm
- Deployment and analysis on real life scenario the trained and tested prediction model will be deployed in a real-life scenario made by the human experts & will be leveraged for further improvement in the methodology.
- The working and basic explanation of those 3 algorithms Random Forest, DecisionTree and Naïve Bayes is given below.

5.2 RANDOM FOREST ALGORITHM

1. It is an ensemble classifier using many decision trees models; it can be used for regression as well as classification.

2. Accuracy and variable importance information can be provided with the results.

3. A random forest is the classifier consisting of a collection of tree structured classifiers k,where the k is independently, identically distributed random trees and each random treeconsist of the unit of vote for classification of input.

4. Random forest uses the Gini index for the classification and determining the final class in each tree.

5. The final class of each tree is aggregated and voted by the weighted values to construct the final classifier.

6. The working of random forest is, A random seed is chosen which pulls out at a random,a collection of samples from the training datasets while maintaining the class

distribution. Shown in the fig below 5.1

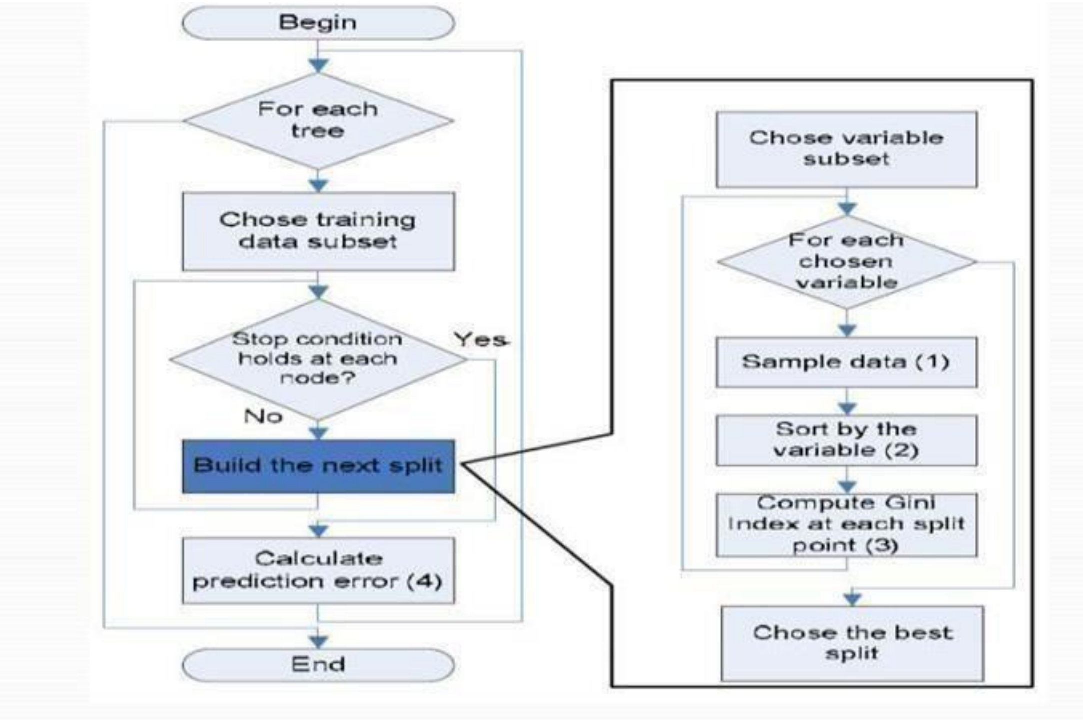


Fig 5.1 Random Forest Example

CHAPTER-6

TESTING

Types Of Tests

6.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logicis functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and isinvasive. Unit tests perform basic tests at component level and test a specificbusiness process, application, and/or system configuration.

6.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if theyactually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination ofcomponents.

6.3 VALIDATION TESTING

An engineering validation test (EVT) is performed on first engineering prototypes, to ensure that the basic unit performs to design goals and specifications. It is important in identifying design problems, and solving them as early in the design cycle as possible, is the key to keeping projects on time and within budget. Too often, product design and performance problems are not detected until late in the product development cycle — when the product is ready to be shipped. The old adage holds true: It costs a penny to make a change in engineering, a dime in production and a dollar after a product is in the field. Verification is a Quality control process that is used to evaluate

whether or not a product, service, or system complies with regulations, specifications, or conditions imposed at thestart of a development phase. Verification can be in development, scale- up, or production.This is often an internal process.

Validation is a Quality assurance process of establishing evidence that provides a high degree of assurance that a product, service, or system accomplishes its intended requirements. This often involves acceptance of fitness for purpose with end users and other product stakeholders.

The testing process overview is as follows:

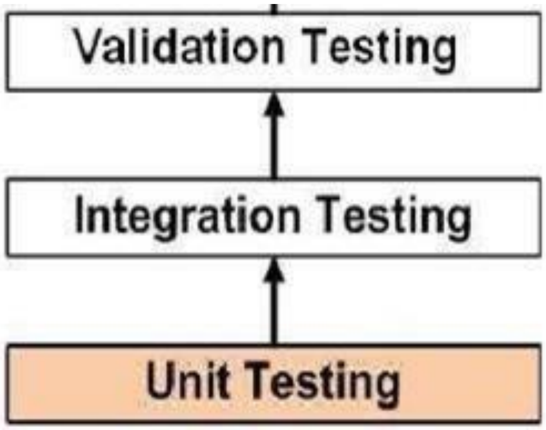


Fig 6.1 The Testing Process

6.4 SYSTEM TESTING

System testing of software or hardware is testing conducted on a complete, integrated system to evaluate the system's compliance with its specified requirements. System testing falls within the scope of black box testing, and as such, should require no knowledge of the inner design of the code or logic.

As a rule, system testing takes, as its input, all of the "integrated" software components that have successfully passed integration testing and also the software system itself integrated with any applicable hardware system. System testing is a more limited type oftesting; it seeks to detect defects both within the "inter-assemblages" and also within thesystem as a whole. System testing is performed on the entire system in the context of a Functional Requirement Specification (FRS) or System Requirement Specification (SRS).

6.5 TESTING OF INITIALIZATION ANDUICOMPONENTS

Serial Number of Test Case	TC 01
Module Under Test	User Registration
Description	A user enters their details for registeringthemselves to the System
Input	Details of Users such as username, email,phone, age, password.
Output	If the user’s details are correct, user is registered. If the user’s details are incorrect, Displays error message. If the user isalready registered, Displays error message.
Remarks	Test Successful.

Table 6.5.1 Test Case for User Registration

Serial Number of Test Case	TC 03
Module Under Test	Prediction Result
Description	User needs to enter the name andvalues to get the prediction result.
Input	Name and Values
Output	If user enters all 5 correct values then the accuracy will be high. If user enters only few values then accuracy will be low.
Remarks	Test Successful.

Table 6.5.2 Test Case for Prediction Result

CHAPTER-7

SNAPSHOTS

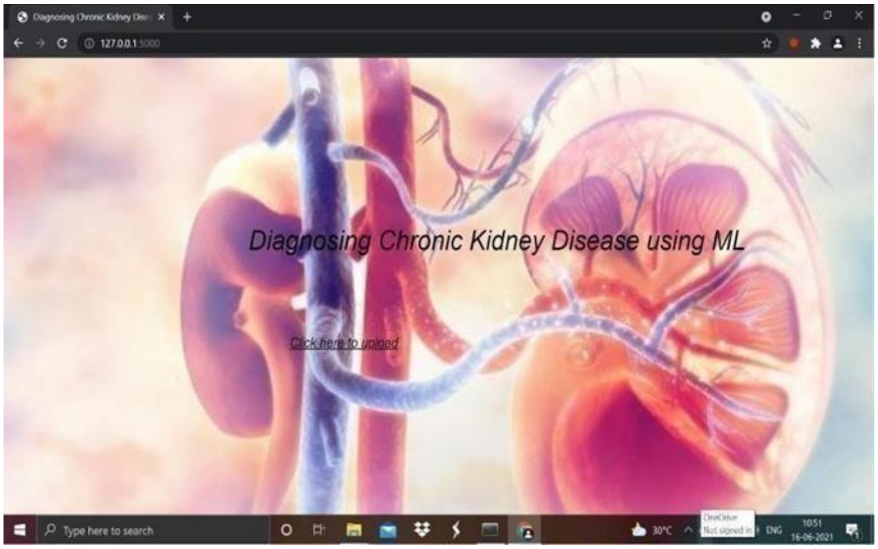
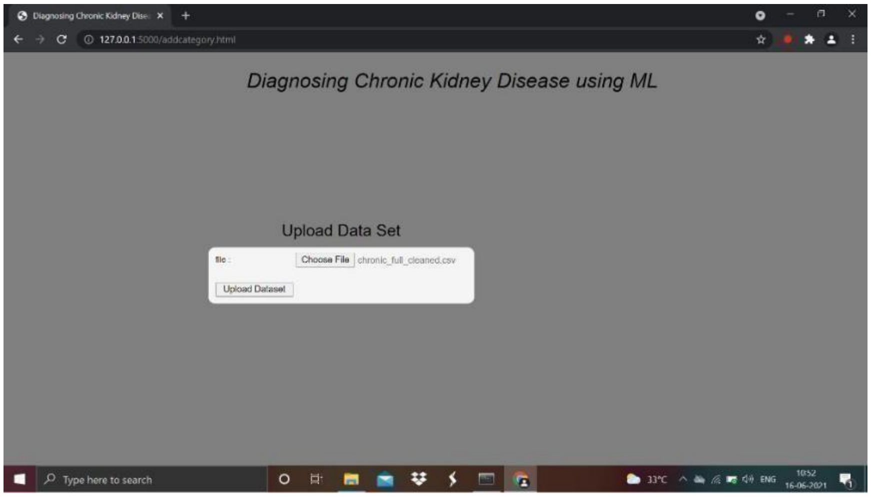
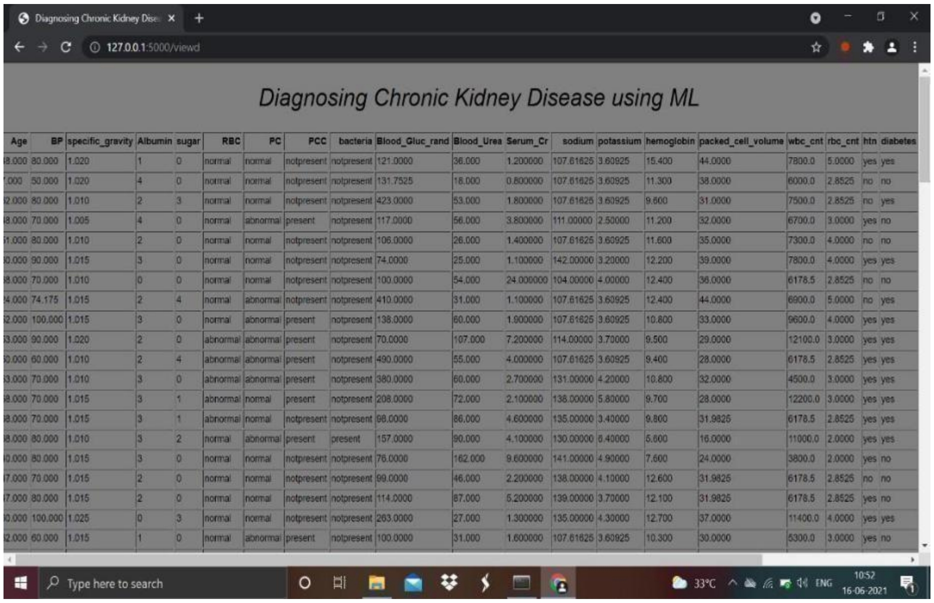

Diagnosing Chronic Kidney Dise...
127.0.0.1:5000
Diagnosing Chronic Kidney Disease using ML
Click here to upload
Type here to search

Figure 7.1: Main Page


Diagnosing Chronic Kidney Dise...
127.0.0.1:5000/addcategory.html
Diagnosing Chronic Kidney Disease using ML
Upload Data Set
Choose File
Upload Dataset
Type here to search

file : chronic_full_cleaned.csv

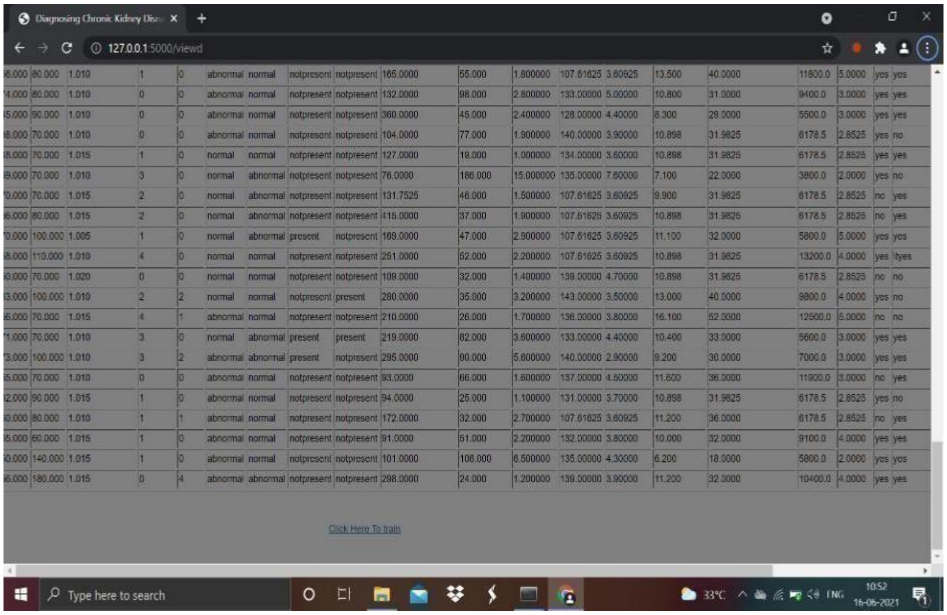
Figure 7.2: Upload file



Diagnosing Chronic Kidney Disease using ML

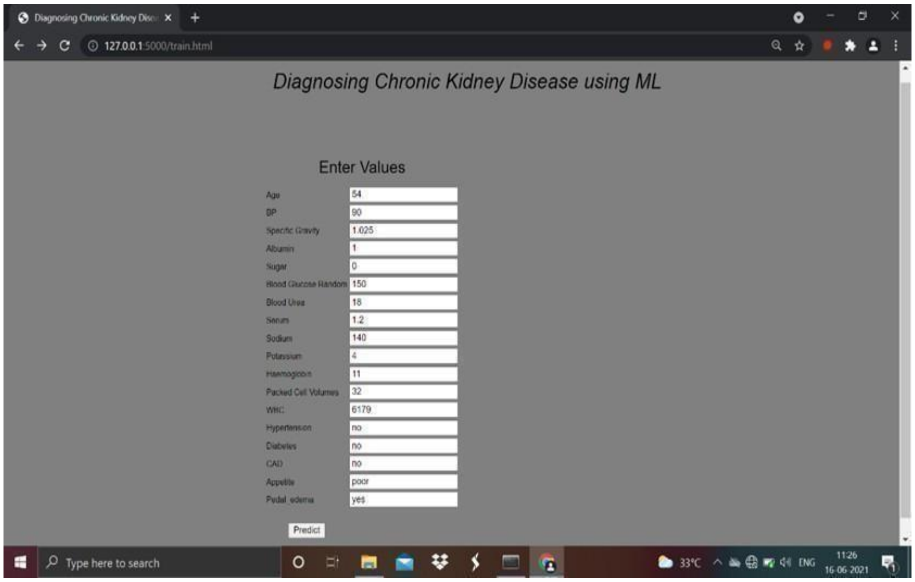
Age	BP	specific_gravity	Albumin	sugar	RBC	PC	PCC	bacteria	Blood_Gluc_rand	Blood_Urea	Serum_Cr	sodium	potassium	hemoglobin	packed_cell_volume	wbc_cnt	rbc_cnt	htn	diabetes
8.000	80.000	1.020	1	0	normal	normal	notpresent	notpresent	121.0000	36.000	1.200000	107.61625	3.60925	15.400	44.0000	7800.0	5.0000	yes	yes
.000	50.000	1.020	4	0	normal	normal	notpresent	notpresent	131.7525	18.000	0.800000	107.61625	3.60925	11.300	38.0000	6000.0	2.8525	no	no
2.000	80.000	1.010	2	3	normal	normal	notpresent	notpresent	423.0000	53.000	1.800000	107.61625	3.60925	9.600	31.0000	7500.0	2.8525	no	yes
8.000	70.000	1.005	4	0	normal	abnormal	present	notpresent	117.0000	56.000	3.800000	111.00000	2.50000	11.200	32.0000	6700.0	3.0000	yes	no
1.000	80.000	1.010	2	0	normal	normal	notpresent	notpresent	106.0000	26.000	1.400000	107.61625	3.60925	11.600	35.0000	7300.0	4.0000	no	no
0.000	90.000	1.015	3	0	normal	normal	notpresent	notpresent	74.0000	25.000	1.100000	142.00000	3.20000	12.200	39.0000	7800.0	4.0000	yes	yes
8.000	70.000	1.010	0	0	normal	normal	notpresent	notpresent	100.0000	54.000	24.000000	104.00000	4.00000	12.400	36.0000	6178.5	2.8525	no	no
4.000	74.175	1.015	2	4	normal	abnormal	notpresent	notpresent	410.0000	31.000	1.100000	107.61625	3.60925	12.400	44.0000	6900.0	5.0000	no	yes
2.000	100.000	1.015	3	0	normal	abnormal	present	notpresent	138.0000	60.000	1.900000	107.61625	3.60925	10.800	33.0000	9600.0	4.0000	yes	yes
3.000	90.000	1.020	2	0	abnormal	abnormal	present	notpresent	70.0000	107.000	7.200000	114.00000	3.70000	9.500	29.0000	12100.0	3.0000	yes	yes
0.000	60.000	1.010	2	4	abnormal	abnormal	present	notpresent	490.0000	55.000	4.000000	107.61625	3.60925	9.400	28.0000	6178.5	2.8525	yes	yes
3.000	70.000	1.010	3	0	abnormal	abnormal	present	notpresent	380.0000	60.000	2.700000	131.00000	4.20000	10.800	32.0000	4500.0	3.0000	yes	yes
8.000	70.000	1.015	3	1	abnormal	normal	present	notpresent	208.0000	72.000	2.100000	138.00000	5.80000	9.700	28.0000	12200.0	3.0000	yes	yes
8.000	70.000	1.015	3	1	abnormal	normal	notpresent	notpresent	98.0000	86.000	4.600000	135.00000	3.40000	9.800	31.9825	6178.5	2.8525	yes	yes
8.000	80.000	1.010	3	2	normal	abnormal	present	present	157.0000	90.000	4.100000	130.00000	6.40000	5.600	16.0000	11000.0	2.0000	yes	yes
0.000	80.000	1.015	3	0	normal	normal	notpresent	notpresent	76.0000	162.000	9.600000	141.00000	4.90000	7.600	24.0000	3800.0	2.0000	yes	no
7.000	70.000	1.015	2	0	normal	normal	notpresent	notpresent	99.0000	46.000	2.200000	138.00000	4.10000	12.600	31.9825	6178.5	2.8525	no	no
7.000	80.000	1.015	2	0	normal	normal	notpresent	notpresent	114.0000	87.000	5.200000	139.00000	3.70000	12.100	31.9825	6178.5	2.8525	yes	no
0.000	100.000	1.025	0	3	normal	normal	notpresent	notpresent	263.0000	27.000	1.300000	135.00000	4.30000	12.700	37.0000	11400.0	4.0000	yes	yes
2.000	60.000	1.015	1	0	normal	abnormal	present	notpresent	100.0000	31.000	1.600000	107.61625	3.60925	10.300	30.0000	5300.0	3.0000	yes	no

Figure 7.3: Dataset



6.000	80.000	1.010	1	0	abnormal	normal	notpresent	notpresent	165.0000	55.000	1.800000	107.61625	3.60925	13.500	40.0000	11800.0	5.0000	yes	yes
4.000	80.000	1.010	0	0	abnormal	normal	notpresent	notpresent	132.0000	98.000	2.800000	133.00000	5.00000	10.800	31.0000	9400.0	3.0000	yes	yes
5.000	90.000	1.010	0	0	abnormal	normal	notpresent	notpresent	360.0000	45.000	2.400000	128.00000	4.40000	8.300	29.0000	5500.0	3.0000	yes	yes
8.000	70.000	1.010	0	0	abnormal	normal	notpresent	notpresent	104.0000	77.000	1.900000	140.00000	3.90000	10.898	31.9825	6178.5	2.8525	yes	no
8.000	70.000	1.015	1	0	normal	normal	notpresent	notpresent	127.0000	19.000	1.000000	134.00000	3.60000	10.898	31.9825	6178.5	2.8525	yes	yes
9.000	70.000	1.010	3	0	normal	abnormal	notpresent	notpresent	76.0000	186.000	15.000000	135.00000	7.60000	7.100	22.0000	3800.0	2.0000	yes	no
0.000	70.000	1.015	2	0	normal	abnormal	notpresent	notpresent	131.7525	46.000	1.500000	107.61625	3.60925	9.900	31.9825	6178.5	2.8525	no	yes
6.000	80.000	1.015	2	0	normal	abnormal	notpresent	notpresent	415.0000	37.000	1.900000	107.61625	3.60925	10.898	31.9825	6178.5	2.8525	no	yes
0.000	100.000	1.005	1	0	normal	abnormal	present	notpresent	169.0000	47.000	2.900000	107.61625	3.60925	11.100	32.0000	5800.0	5.0000	yes	yes
8.000	110.000	1.010	4	0	normal	normal	notpresent	notpresent	251.0000	52.000	2.200000	107.61625	3.60925	10.898	31.9825	13200.0	4.0000	yes	\tyes
0.000	70.000	1.020	0	0	normal	normal	notpresent	notpresent	109.0000	32.000	1.400000	139.00000	4.70000	10.898	31.9825	6178.5	2.8525	no	no
3.000	100.000	1.010	2	2	normal	normal	notpresent	present	280.0000	35.000	3.200000	143.00000	3.50000	13.000	40.0000	9800.0	4.0000	yes	no
6.000	70.000	1.015	4	1	abnormal	normal	notpresent	notpresent	210.0000	26.000	1.700000	136.00000	3.80000	16.100	52.0000	12500.0	5.0000	no	no
1.000	70.000	1.010	3	0	normal	abnormal	present	present	219.0000	82.000	3.600000	133.00000	4.40000	10.400	33.0000	5600.0	3.0000	yes	yes
3.000	100.000	1.010	3	2	abnormal	abnormal	present	notpresent	295.0000	90.000	5.600000	140.00000	2.90000	9.200	30.0000	7000.0	3.0000	yes	yes
5.000	70.000	1.010	0	0	abnormal	normal	notpresent	notpresent	93.0000	66.000	1.600000	137.00000	4.50000	11.600	36.0000	11900.0	3.0000	no	yes
2.000	90.000	1.015	1	0	abnormal	normal	notpresent	notpresent	94.0000	25.000	1.100000	131.00000	3.70000	10.898	31.9825	6178.5	2.8525	yes	no
0.000	80.000	1.010	1	1	abnormal	normal	notpresent	notpresent	172.0000	32.000	2.700000	107.61625	3.60925	11.200	36.0000	6178.5	2.8525	no	yes
5.000	60.000	1.015	1	0	abnormal	normal	notpresent	notpresent	91.0000	51.000	2.200000	132.00000	3.80000	10.000	32.0000	9100.0	4.0000	yes	yes
0.000	140.000	1.015	1	0	abnormal	normal	notpresent	notpresent	101.0000	106.000	6.500000	135.00000	4.30000	6.200	18.0000	5800.0	2.0000	yes	yes
6.000	180.000	1.015	0	4	abnormal	abnormal	notpresent	notpresent	298.0000	24.000	1.200000	139.00000	3.90000	11.200	32.0000	10400.0	4.0000	yes	yes

Figure 7.4: Click on train

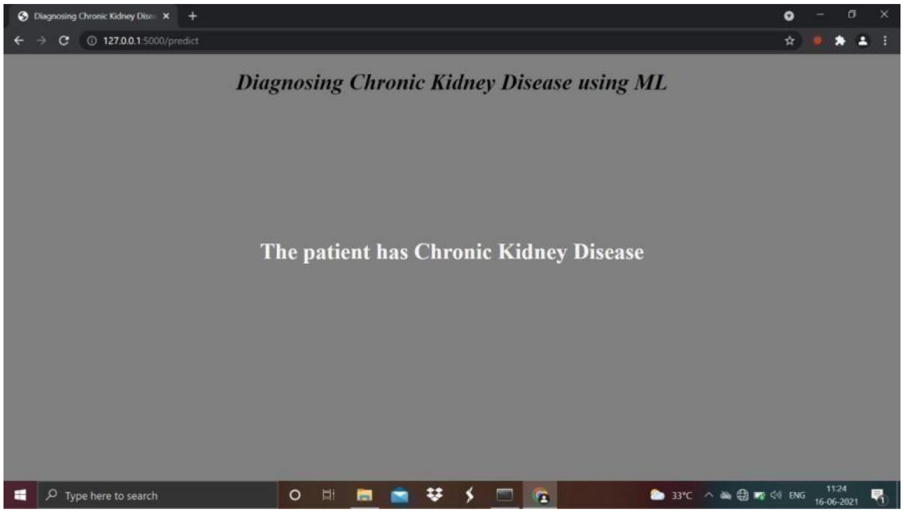


Diagnosing Chronic Kidney Disease using ML

Enter Values

Age	54
BP	90
Specific Gravity	1.025
Albumin	1
Sugar	0
Blood Glucose Random	150
Blood Urea	18
Serum	1.2
Sodium	140
Potassium	4
Haemoglobin	11
Packed Cell Volumes	32
WBC	6179
Hypertension	no
Diabetes	no
CAD	no
Appetite	poor
Pedal_edema	yes

Figure 7.5: Test Data



Diagnosing Chronic Kidney Disease using ML

The patient has Chronic Kidney Disease

Fig 7.5: Prediction Page

CHAPTER-8

CONCLUSION AND FUTURE ENHANCEMENT

Conclusion

The proposed CKD diagnostic methodology is feasible in terms of data imputation and samples diagnosis. After unsupervised imputation of missing values in the data set by using preprocessing steps and use Random Forest Classifier to predict result, the integrated model could achieve a satisfactory accuracy. Hence, we speculate thatapplying this methodology to the practical diagnosis of CKD would achieve a desirableeffect. In addition, this methodology might be applicable to the clinical data of the otherdiseases in actual medical diagnosis..

Future Enhancement

- Facility for modifying user detail.
- More interactive user interface.
- Facilities for Backup creation.
- Can be done as stand alone.
- Can be done as Mobile Application.
- More Details and Latest Diseases.

REFERENCES

1. Chronic Kidney Disease Prediction and Doctor Recommendation System by www.irjet.net

2. Disease Prediction Based on Prior Knowledge by www.hcup- us.ahrq.gov/nisoverview.jsp

3. GDPS- General Chronic Kidney Disease Prediction System by www.irjet.net

4. Chronic Kidney Disease Prediction Using Machine Learning byInternational Research Journal of Engineering and Technology(IRJET).

5. Kaveeshwar, S.A., and Cornwall, J., 2014, "The current state of disease mellitus in India".AMJ,7(1),pp.45-48.

6. Dean,L.,McEntyre,J.,2004,"The Genetic Land scape of Disease[Internet]. Bethesda (MD): National Center for Biotechnology Information (US); Chapter 1,Introduction to Disease.2004 Jul7.

7. Machine Learning Methods Used in Disease by www.wikipedia.com

8. https://www.researchgate.net/publication/325116774_disease_prediction_using_machine_learning_techniques

9. https://ieeexplore.ieee.org/document/8819782/disease_prediction

10. Algorithms Details from www.dataspirant.com

11. https://www.youtube.com/disease_prediction

12. https://www.slideshare.com/disease_prediction

13. https:/en.wikipedia.org/machine_learning_algorithms

14. https://en.wikipedia.org/wiki/Python_(programming_language)

15. https://wiki.python.org/TkInter

16. https://creately.com/lp/uml-diagram-tool/

17. https://app.diagrams.net/