### A Project Report on

# CHRONIC KIDNEY DISEASE PREDICTION MODEL USING MACHINE LEARNING

Submitted in partial fulfilment for the award of the degree of

B.Tech (Branch)
By

RAJIV GUPTA 18BCE2203 VAIBHAV SINGH 18BCE2313

Under the guidance of Dr. Sweta Bhattacharya Assistant Professor (Senior)



SCHOOL OF COMPUTER SCIENCE & ENGINEERING April 2020

#### Table of Contents

#### **Abstract**

### Acknowledgement

- 1. Introduction
  - 1.1. Motivation
  - 1.2. Aim of the proposed Work
  - 1.3. Objective(s) of the proposed work
  - 1.4. Report Organization
- 2. Literature Survey
  - 2.1. Survey of the Existing Models/Work
  - 2.2. Summary/Gaps identified in the Survey
- 3. Proposed System Requirements Analysis and Design
  - 3.1. Introduction
  - 3.2. Requirement Analysis
    - 3.2.1. Stakeholder Identification
    - 3.2.2. Functional Requirements
    - 3.2.3. Non Functional Requirements
    - 3.2.4. System Requirements
      - 3.2.4.1. H/W Requirements(details about Application-Specific Hardware)
      - 3.2.4.2. S/W Requirements(details about Application-Specific Software)
    - 3.2.5. Software Requirement Specification document
    - 3.2.6. Work breakdown structure
    - 3.2.7. Pert chart (with task table)
    - 3.2.8. Gantt Chart
- 4. Design of the Proposed System
  - 4.1. Introduction
  - 4.2. High level Design (Framework, Architecture or Module for the Proposed System(with explanation))
    - 4.2.1. Architecture design (choose the appropriate pattern with justification)
    - 4.2.2. Architecture diagram (explanation)
    - 4.2.3. UI design
  - 4.3. Detailed Design (ER Diagram/UML Diagram/Mathematical Modeling)
    - 4.3.1. ER Diagram
    - 4.3.2. UML diagram (Use case, class, Statechart, Activity and interaction diagrams)
- 5. Implementation and Testing (Snap shots with description)
  - 5.1. Implementation details (snapshots)

- 5.2. Testing
  - 5.2.1. Types of Testing
  - 5.2.2. Testcases (for all modules as per the template)
- 6. Conclusion, Limitations and Scope for future Work
- 7. References

### **ABSTRACT**

Chronic kidney disease (CKD) is an important public health issue because CKD patients have an increased risk of end-stage renal disease (ESRD). Dialysis has charged the national health insurance system billions of dollars in recent years, and the costs are rising. We will develop a system by using artificial neural networks, and other data mining algorithms that could predict the disease at early stages. Chronic kidney disease (CKD) is common. Kidney disease severity can be classified by estimated glomerular filtration rate (GFR) and albuminuria, but more accurate information regarding risk for progression to kidney failure is required for clinical decisions about testing, treatment, and referral.

### **ACKNOWLEDGEMENT**

I take immense pleasure in thanking **Dr. G. Viswanathan**, my beloved **Chancellor**, VIT University and respected **Dean**, **Dr. Saravanan R**, **SCHOOLOF COMPUTER SCIENCE AND ENGINEERING**, for having permitted meto carry out the project.

I express gratitude to my guide, **SWETA BHATTACHARYA.**, for guidance and suggestions that helped me to complete the project on time. Words are inadequate to express my gratitude to the faculty and staff members who encouraged and supported me during the project. Finally, I would like to thank my ever-loving parents for their blessings and my friends for their timely help and support.

### **INTRODUCTION**

Chronic Kidney Disease (CKD) is a major public health concern with rising prevalence. Kidney disease is when the kidneys are damaged and could not filter the blood properly. This damage could cause the wastes to build up in the body. There are five stages of CKD, the most serious one is stage 5 because, at this stage, the kidneys are unable to do most of their functions. It is difficult to pinpoint the CKD stage of each patient especially at the early-stages. It also causes a high possibility of death within a short period of time, a patient must be hospitalized and appropriately cured. The most common causes of kidney disease are diabetes and high blood pressure.

Machine learning is a field of computer science that gives the ability of machines to learn without being explicitly programmed. By using computational methods, machine learning has shown success in providing solutions for earlystage diagnosis in a variety of medical domains. These methods are used to find hidden patterns from data and mine these data for decision- makers.

#### Motivation

Chronic kidney disease (CKD) affects 8–16% of the populationworldwide and is associated with an increased risk of cardio-vascular disease (CVD), muscle wasting, decreased phys-ical function and overall poorer quality of life (QOL). Exercise is being increasingly recognized for its therapeuticbenefits in patients with CKD, which include improved physicalfitness, cardiovascular health and better QOL. Most of theresearch regarding such benefits has been conducted in dialysispatients but the potential of exercise to modulate a num-ber of factors related to disease progression, as well as addressco-morbidities, makes it a particularly interesting and theoretically important treatment for all patients with CKD. Current international guidelines recommend that CKD patients should engage in an exercise programme that is compat-ible with cardiovascular health for 30 min, 5 days of the week. Despite this, CKD patients are known to lead insufficientlyactive lifestyles. Identifying barriers and asking participants to strategizeways to overcome them is a popular technique used in behav-iour change interventions.

### AIM OF THE PROPOSED WORK

The proposed work operates on three stages: pre-processing, implementation and conclusion. The pre-processing stage is the primary process since the database may contain redundant and noise data. By examining the data, different processes take place such as data cleaning, filling missing values, removing excessive data because the missing values and excessive data degrade the performance. For the implementation stage we will work with three most common neural network algorithms (Probabilistic Neural Networks, Multilayer Perceptron algorithm, Support vector machine algorithm) and visualization tools to map the results, in conclusion stage by analysing the graphs various results could be drawn.

### **OBJECTIVE(S) OF THE PROPOSED WORK**

Chronic kidney disease (CKD) is common, and associated with increased risk of cardiovascular disease and end-stage renal disease, which are potentially preventable through early identification and treatment of individuals at risk, this product will develop models that can help in the prediction of the disease at an early stage so that sufficient health facilities could be provided to the patients.

### **Literature Survey**

### **Survey of the Existing Models/Work**

The following are some of the earlier works in the field of using machine-learning algorithms to diagnose CKD. They used the same dataset from UCI Machine Learning Repository with different machine learning algorithms. The dataset has been collected from Apollo hospital (Tamilandu) [6]. It has 25 attributes and 400 total instances, out of which 250 instances are classified as CKD and 150 instances as nonCKD [6]. Jena and Kamila proposed a method for predicting chronic kidney disease using SVM, Naïve Bayes, Multilayer Perceptron, J48 (a type of decision tree), Conjunctive Rule and Decision Table. From the experimental result, the Multilayer Perception algorithm gives a better classification accuracy of 99.7%. The performance of these algorithms was measured by classification accuracy, the time taken to build the model, the time taken to test the model, and the mean absolute error [7]. In another related work, Manish Kumar has made some research on other authors' studies and reported that SVM performed best compared to other classifiers. The authors used six machine learning algorithms namely: Random Forest (RF), Naïve Bayes, Sequential Minimum Optimization (SMO), Radial Basis Function (RBFClassifier), Multilayer Perceptron Classifier (MLPC) and Simple Logistic (SLG). The author compared the performance of the six classifiers with SVM. The results showed that RF achieved a performance of 100% classification accuracy, while SMO and RBF achieved a lower classification accuracy [8].

### Summary/Gaps identified in the Survey

- A. Artificial Neural Network Artificial neural networks (ANNs) are a branch of machine learning that are statistical-based learning algorithms which were designed to simulate the properties of the biological neural networks [13]. One of the most widely used neural network is Multilayer Perceptron (MLP). MLP consists mainly of three types of layers made up of artificial neurons and connected by weighted links as shown in figure 1 [14]. Depending on the weights and a specific value called the activation value, some neurons will be activated to some value and others will not. The activation pattern of a layer affects that of the next layer [15].
- B. Support Vector Machine The original Support Vector Machine (SVM) algorithm was first introduced by Russian mathematicians Vladimir Vapnik and Alexey Chervonenkis based on decades of research in computational learning theory [16]. SVM is a supervised machine learning algorithm used in both classification and regression problems [17]. Focusing on classification problems, SVM became a popular choice among many researchers, since it often outperforms other classification algorithms [17]. Moreover, SVM performs well even when the number of samples is low [17]. SVM has been used in many fields such as optical character recognition [18], email spam detection [19], and medical diagnosis [20]. For further details readers should refer to [16], [21], [22].
- C. K-Nearest Neighbors K-NN is a simple classifier which increasingly became a popular choice in practice. The idea of K-NN was first introduced by Fix and Hodges in 1951 as a non-parametric method for pattern classification [24]. It is considered a lazy classifier as it does not require building a training model, instead, for each query, the k nearest neighbors are located from the training database regardless of their class label, then the majority vote predicts the class label of the instance [25].

#### PROPOSED SYSTEM REQUIREMENTS ANALYSIS AND DESIGN

### **INTRODUCTION**

### **Purpose**

This project will help doctors and other users to predict whether the patient has a kidney failure. The CKD model is made using python and machine learning to detect the kidney failure.

#### **Document Conventions**

DS = Dataset

UI = User Interface

CKD = Chronic Kidney Disease

ANN = Artificial Neural Network

### **Intended Audience and Reading Suggestions**

This document is meant to be read by anyone who is working on this project as well as for the people who would like to refer this project for future work.

### **Product Scope**

Chronic kidney disease (CKD) is common, and associated with increased risk of cardiovascular disease and end- stage renal disease, which are potentially preventable through early identification and treatment of individuals at risk, this product will develop models that can help in the prediction of the disease at an early stage so that sufficient health facilities could be provided to the patients.

#### Features:

- Commenting and interacting between other users.
- Uploading files, documents, datasets and relevant information.
- Validation of expert users and normal users.

#### Milestones:

- Designing CKD models and relationships.
- Designing the UI/UX and its webpage layout

#### Deliverables:

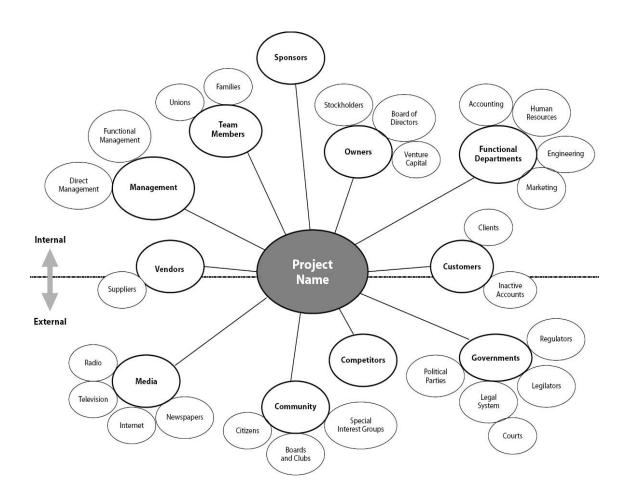
- Basic API for interaction.
- Fully fledged model with many features.

### **Technical Requirements:**

- Responsive UI
- Should be a scalable model that is able to handle a large dataset.
- Should facilitate to interaction with users.

### **REQUIREMENT ANALYSIS**

### STAKEHOLDER IDENTIFICATION



<b>STAKEHOLDERS</b>	<u>INTERESTS</u>	ESTIMATED PRIORITY		
OWNER	<ul> <li>Achieve targets</li> </ul>	1		
	<ul> <li>Liability</li> </ul>			
	<ul> <li>Increase sales margin</li> </ul>			
SPONSOR	<ul> <li>Suscessfully addresses needs of</li> </ul>	3		
	adjunct customers			
	<ul> <li>Appears competent among peers</li> </ul>			
	<ul> <li>Provides new market to expand</li> </ul>			
	ventures			
TEAM	<ul> <li>New product excitement</li> </ul>	2		
MEMEBERS OR	<ul> <li>Retain and expand skill level</li> </ul>			
DEVELOPERS	<ul> <li>Keep bringing new changes and</li> </ul>			
	updates to the product			
CUSTOMERS	Tests and uses the product	3		
	<ul> <li>Check for feasibility and usability</li> </ul>			

### FUNCTIONAL REQUIREMENTS

### **Product Perspective**

The proposed project operates on three stages: pre-processing, implementation and conclusion. The pre-processing stage is the primary process since the database may contain redundant and noise data. By examining the data, different processes take place such as data cleaning, filling missing values, removing excessive data because the missing values and excessive data degrade the performance. For the implementation stage we will work with three most common neural network algorithms (Probabilistic Neural Networks, Multilayer Perceptron algorithm, Support vector machine algorithm) and visualization tools to map the results, in conclusion stage by analysing the graphs various results could be drawn.

#### **Product Functions**

Predict kidney failure.
Show a comparison study between all the patients.
Shows how accurate the model is.
Graphical visualization of the model and the results

### **User Classes and Characteristics**

- Scientists
- Engineers
- Doctors
- Entrepreneurs
- Innovative/Curious folk
- Students

### **Operating Environment**

This product will be launched for Window, Linux and Mac operating systems. The basic hardware required to support these operating systems are (4GB ram,4core processor), (1GB ram, 2core processor) and 2GB ram,4 core processor) respectively.

### **Design and Implementation Constraints**

- The algorithms used should be able to train the model properly to increase the accuracy of model.
- The dataset used should be extensive enough to train the model properly.
- Data Cleaning should be done properly to get proper results.

### NON FUNCTIONAL REQUIREMENTS

### **Performance Requirements**

The website should be fully functional even when scaled to 1000's of users. The latency periods for retrieving data should be under 1ms per 1000 queries. 48 threads should be running on the Jupyter Notebook server.

The model should be properly fitted and accurate to avoid large size that increases latency.

### **Safety Requirements**

There are no safety compromises in this project.

### **Security Requirements**

Sufficient measures should be taken to avoid the following security attacks:

- Cross site request Forgery
- XML scripting
- SQL injections
- Dataset checking

### **Software Quality Attributes**

This project aims to be easy to use for end users and also aims to attract experts on to the site. For an overall better experience.

The main focus is to increase the accuracy of the model to get better results.

### **SYSTEM REQUIREMENTS**

### **H/W Requirements(details about Application-Specific Hardware)**

The website can be accessed through any of computers with a proper internet connection. Preferred configuration is i7 intel processor, 8gb Ram, 2 Gb Graphics card.

### S/W Requirements(details about Application-Specific Software)

The website can be accessed through any of the common browsers such as mozilla Firefox, Google Chrome, Microsoft edge, Safari etc. Additional softwares Anaconda, Python, Jupyter Notebook also had to be installed for the model to work.

If the User is not logged in through proper channel, then the screen will appear for which the user can be authenticated and then directed to the main page.

The main page will contain the users feed or relevant posts or projects on which the user can comment and interact. The main page will also contain links to the logout page, documentation page FAQ page. The user will also be able to request help from one of the experts in the site.

The interface works using machine learning libraries and runs on Jupyter Notebook.

### SOFTWARE REQUIREMENT SPECIFICATION DOCUMENT

### 1. Introduction

### 1.1 Purpose

This project will help doctors and other users to predict whether the patient has a kidney failure. The CKD model is made using python and machine learning to detect the kidney failure.

### **1.2 Document Conventions**

DS = Dataset

UI = User Interface

CKD = Chronic Kidney Disease

ANN = Artificial Neural Network

### **1.3** Intended Audience and Reading Suggestions

This document is meant to be read by anyone who is working on this project as well as for the people who would like to refer this project for future work.

### 1.4 Product Scope

Chronic kidney disease (CKD) is common, and associated with increased risk of cardiovascular disease and endstage renal disease, which are potentially preventable through early identification and treatment of individuals at risk, this product will develop models that can help in the prediction of the disease at an early stage so that sufficient health facilities could be provided to the patients.

#### Features:

- Commenting and interacting between other users.
- Uploading files, documents, datasets and relevant information.
- Validation of expert users and normal users.

#### Milestones:

- Designing CKD models and relationships.
- Designing the UI/UX and its webpage layout

#### Deliverables:

- Basic API for interaction.
- Fully fledged model with many features.

### **Technical Requirements:**

- Responsive UI
- Should be a scalable model that is able to handle a large dataset.

### 1.5 References

- 1. https://thesai.org/Downloads/Volume10No8/Paper\_1Detection\_of\_Chronic\_Kidney\_Disease.pdf
- 2. <a href="https://medium.com/@randerson112358/chronic-kidney-disease-prediction-detection-using-machine-learning-29cc7e3eba96">https://medium.com/@randerson112358/chronic-kidney-disease-prediction-detection-using-machine-learning-29cc7e3eba96</a>
- 3. <a href="https://wonderfulengineering.com/this-artificial-kidney-eliminates-the-need-for-kidney-ialysis/?fbclid=IwAR19Olftga9vr0HEhwl51c89OPN6RGsxPNDDJB2rzhqdKkIVGYYXeGCF1Ds">https://wonderfulengineering.com/this-artificial-kidney-eliminates-the-need-for-kidney-ialysis/?fbclid=IwAR19Olftga9vr0HEhwl51c89OPN6RGsxPNDDJB2rzhqdKkIVGYYXeGCF1Ds</a>

### 2. Overall Description

### 2.1 Product Perspective

The proposed project operates on three stages: pre-processing, implementation and conclusion. The pre-processing stage is the primary process since the database may contain redundant and noise data. By examining the data, different processes take place such as data cleaning, filling missing values, removing excessive data because the missing values and excessive data degrade the performance. For the implementation stage we will work with three most common neural network algorithms (Probabilistic Neural Networks, Multilayer Perceptron algorithm, Support vector machine algorithm) and visualization tools to map the results, in conclusion stage by analysing the graphs various results could be drawn.

### 2.2 Product Functions

- Predict kidney failure.
- Show a comparison study between all the patients.
- Shows how accurate the model is.
- Graphical visualization of the model and the results.

### 2.3 User Classes and Characteristics

- Scientists
- Engineers
- Doctors

- Entrepreneurs
- Innovative/Curious folk
- Students

### **2.4** Operating Environment

This product will be launched for Window, Linux and Mac operating systems. The basic hardware required to support these operating systems are (4GB ram,4core processor), (1GB ram, 2core processor) and 2GB ram,4 core processor) respectively.

### 2.5 Design and Implementation Constraints

- The algorithms used should be able to train the model properly to increase the accuracy of model.
- The dataset used should be extensive enough to train the model properly.
- Data Cleaning should be done properly to get proper results.

### 2.6 User Documentation

Please refer section 3.1 for user interfaces.

### 2.7 Assumptions and Dependencies

It is assumed that the hardware designed will work correctly with the third-party operating system Windows 10 and other Operating Systems and the developed software.

### 3. External Interface Requirements

### 3.1 User Interfaces

If the User is not logged in through proper channel, then the screen will appear for which the user can be authenticated and then directed to the main page.

The main page will contain the users feed or relevant posts or projects on which the user can comment and interact. The main page will also contain links to the logout page, documentation page FAQ page. The user will also be able to request help from one of the experts in the site.

The interface works using machine learning libraries and runs on Jupyter Notebook.

### 3.2 Hardware Interfaces

The website can be accessed through any of computers with a proper internet connection. Preferred configuration is i7 intel processor, 8gb Ram, 2 Gb Graphics card.

### 3.3 Software Interfaces

The website can be accessed through any of the common browsers such as mozilla Firefox, Google Chrome, Microsoft edge, Safari etc. Additional softwares Anaconda, Python, Jupyter Notebook also had to be installed for the model to work.

### 3.4 Communications Interfaces

This software will make use of HTTP connections for website access and will use SMTP for email related functionality. It will also use Jupyter Notebook and various machine learning libraries to work properly.

### 4. System Features

### 4.1 Upload Project Content

### 4.1.1 Description and Priority

This allows the users to predict whether the patient is suffering from kidney failure or not using various attributes and medical terms.

Priority:

This is the main priority of the whole project and crucial to implement.

#### 4.1.2 Stimulus/Response Sequences

- 1) User inputs the dataset
- 2) User enters his requirements.
- 3) User enters related documents and clicks run.

### 4.1.3 Functional Requirements

- 1) The system should provide the option of an upload to upload the dataset.
- 2) User needs to be authenticated before uploading.
- 3) The system should support all file type uploads relevant to the project.
- 4) There should be a UI for writing descriptions in the website itself.

### 4.2 Expert help

### 4.1.1 Description and Priority

Experts are people who are well versed in the subjects and can help the other users on the platform.

Priority:

This is a medium priority feature.

### 4.1.2 Stimulus/Response Sequences

- 1) User presses the help button
- 2) User browses all experts presently available to help.

3) User can then interact with the expert and ask for help.

### 4.1.3 Functional Requirements

- 1) The system should provide the option of a help button.
- 2) Experts have to be certified and verified before becoming an authenticated expert on the website.
- 3) The system should support all communicating factors such as file sharing or image sharing.
- 4) There should be a UI for interacting with experts in the website itself.

### 5. Other Non-functional Requirements

### **5.1 Performance Requirements**

The website should be fully functional even when scaled to 1000's of users. The latency periods for retrieving data should be under 1ms per 1000 queries. 48 threads should be running on the Jupyter Notebook server.

The model should be properly fitted and accurate to avoid large size that increases latency.

### 5.2 Safety Requirements

There are no safety compromises in this project.

### 5.3 Security Requirements

Sufficient measures should be taken to avoid the following security attacks:

- Cross site request Forgery
- XML scripting
- SQL injections
- Dataset checking

### **5.4 Software Quality Attributes**

This project aims to be easy to use for end users and also aims to attract experts on to the site. For an overall better experience.

The main focus is to increase the accuracy of the model to get better result

### 6. Other Requirements

Cloud storage for faster access to data and easy to maintain server

**Appendix A:** 

**Glossary** 

**Appendix B:** 

**Analysis** 

**Models** 

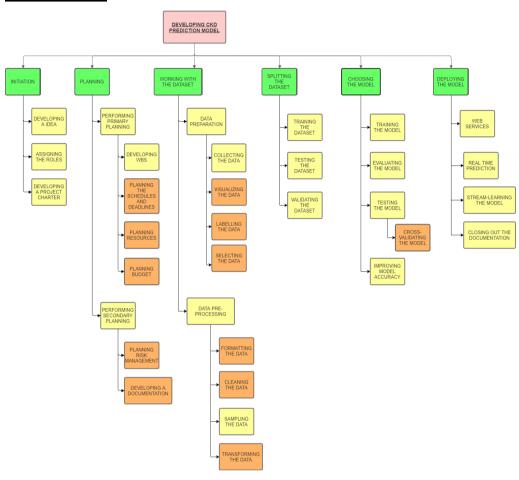
Refer the Graphical representations below.

### **Appendix C: To Be Determined List**

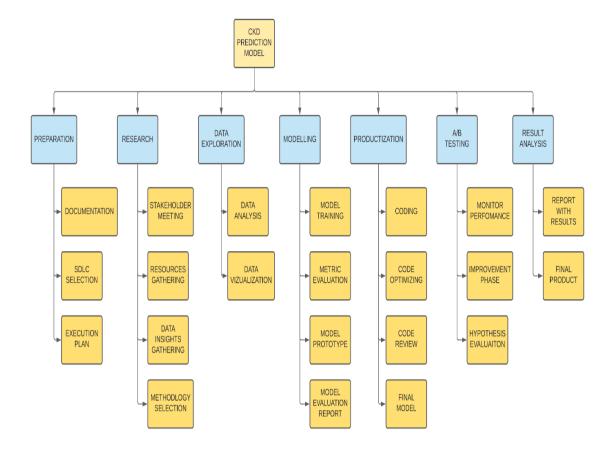
- internet assistant.
- Recommendation algorithms for users.

### WORK BREAKDOWN STRUCTURE

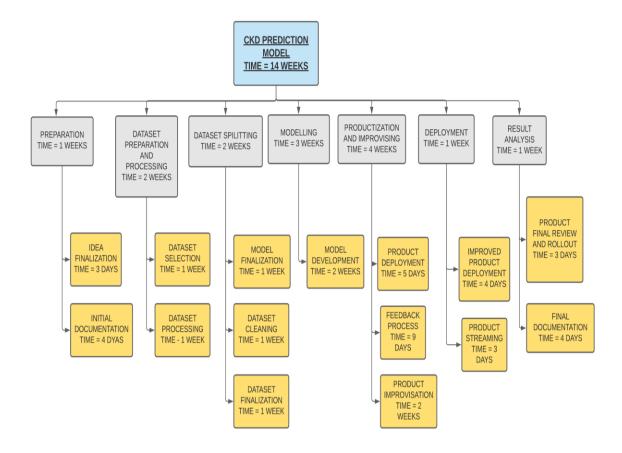
### **VERB BASED**



### **NOUN BASED**



### **TIME BASED**



### **GANTT CHART**

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Requirements Gathering						
System Design						
Design Phase 1						
Design Phase 2						
Testing and Verification						
Deployment						

### **DESIGN OF THE PROPOSED SYSTEM**

#### **INTRODUCTION**

We will be using RAPID APPLICATION DEVELOPEMNT MODEL (RAD)

The **RAD** (**Rapid Application Development**) model is based on prototyping and iterative development with no specific planning involved. The process of writing the software itself involves the planning required for developing the product.

Rapid Application Development focuses on gathering customer requirements through workshops or focus groups, early testing of the prototypes by the customer using iterative concept, reuse of the existing prototypes (components), continuous integration and rapid delivery.

Rapid application development is a software development methodology that uses minimal planning in favour of rapid prototyping. A prototype is a working model that is functionally equivalent to a component of the product.

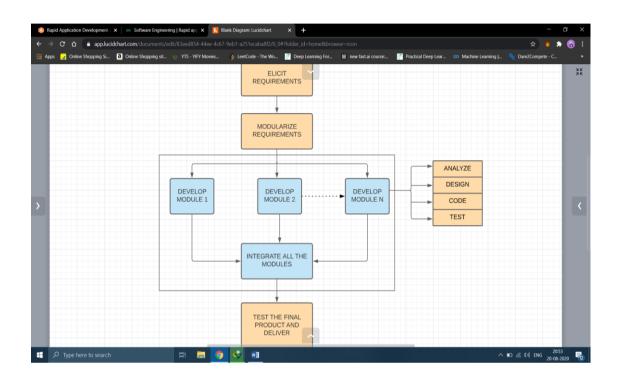
In the RAD model, the functional modules are developed in parallel as prototypes and are integrated to make the complete product for faster product delivery. Since there is no detailed preplanning, it makes it easier to incorporate the changes within the development process.

RAD projects follow iterative and incremental model and have small teams comprising of developers, domain experts, customer representatives and other IT resources working progressively on their component or prototype.

The most important aspect for this model to be successful is to make sure that the prototypes developed are reusable.

## HIGH LEVEL DESIGN (FRAMEWORK, ARCHITECTURE OR MODULE FOR THE PROPOSED SYSTEM(WITH EXPLANATION))

## ARCHITECTURE DESIGN (CHOOSE THE APPROPRIATE PATTERN WITH JUSTIFICATION)

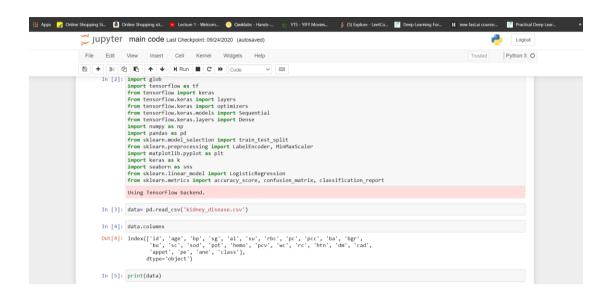


### **ARCHITECTURE DIAGRAM (EXPLANATION)**

#### JUSTIFICATION FOR USING RAD MODEL

- 1. We use RAD model because there is a need to create a system that can be modularized in 2-3 months of time.
- 2. Also it increases the reusability of components and review of our work can be done quickly.
- 3. The other main purpose of using this model is because it allows integration from very beginning which solves a lot of integration issues.
- 4. Due to limited amount of time and man power, this will allow us to deliver our project in small pieces, that is whole project can be divided into number of smaller components.
- 5. The technical risks are low and also the requirements of the product are known well beforehand.
- 6. Also there is a good scope of getting reliable feedback on our deliverables for further improvement.

### **UI DESIGN**



### Detailed Design (ER Diagram/UML Diagram/Mathematical Modeling)

### **ER Diagram**

### Logistic regression algorithm

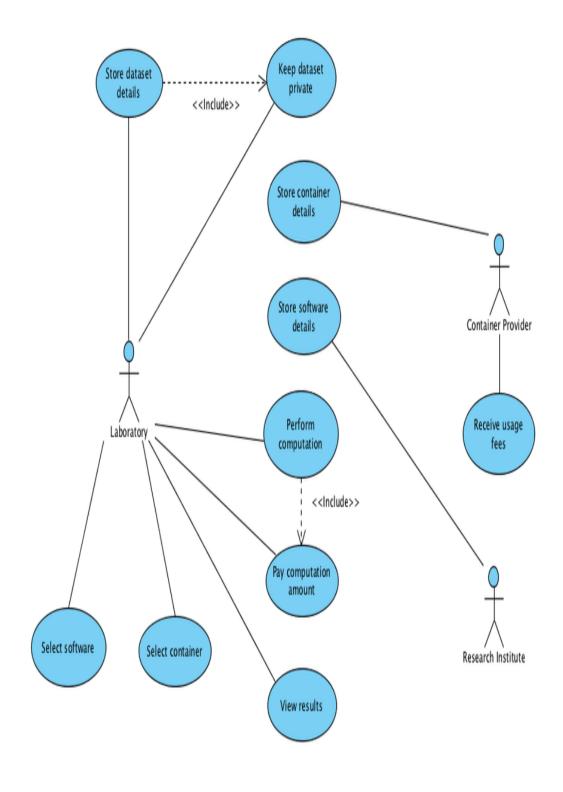
```
X = data.iloc[:,:-1]
y = data['class']
logreg = LogisticRegression(max_iter = 10000)
X_train, X_test, y_train, y_test = train_test_split(X,y, stratify = y, shuffle = True)
logreg.fit(X_train,y_train)
test_pred = logreg.predict(X_test)
train_pred = logreg.predict(X_train)
print('Train Accuracy: ', accuracy_score(y_train, train_pred))
print('Test Accuracy: ', accuracy_score(y_test, test_pred))
```

### Ann algorithm

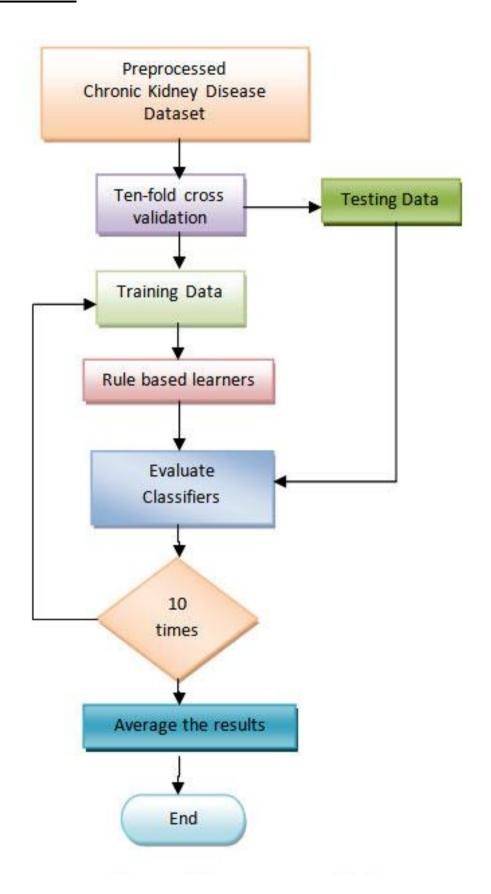
```
#build model
model = Sequential()
#first layer
model.add(Dense(256,input_dim= len(X.columns),kernel_initializer=
k.initializers.random_normal(seed= 13),activation= 'relu'))
#second layer
model.add(Dense (1, activation = 'hard_sigmoid'))
#compiling the model
#opt = adam(lr=0.001, decay=1e-6)
model.compile(loss = 'binary_crossentropy', optimizer = 'adam', metrics = ['accuracy'])
```

## <u>UML DIAGRAM (USE CASE, CLASS, STATECHART, ACTIVITY AND INTERACTION DIAGRAMS)</u>

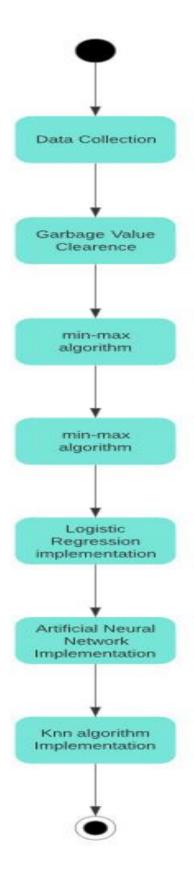
### **USE CASE**



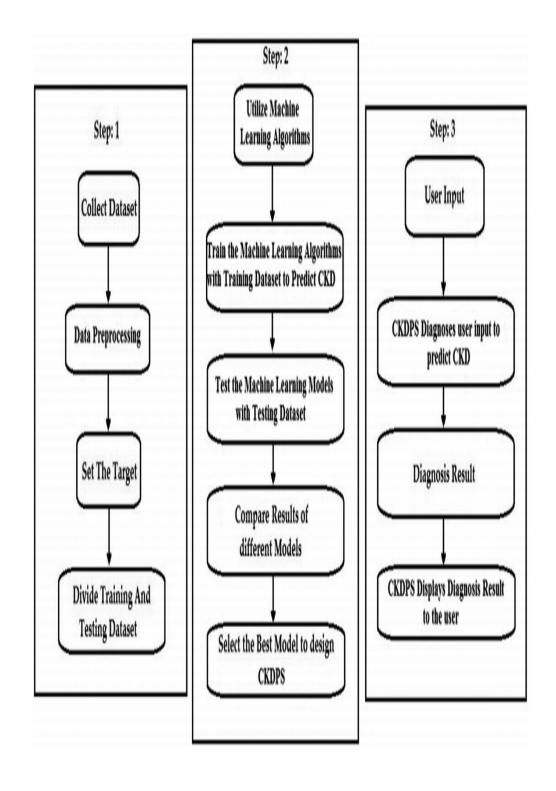
### **STATESCHART**



### **ACTIVITY DIAGRAM**

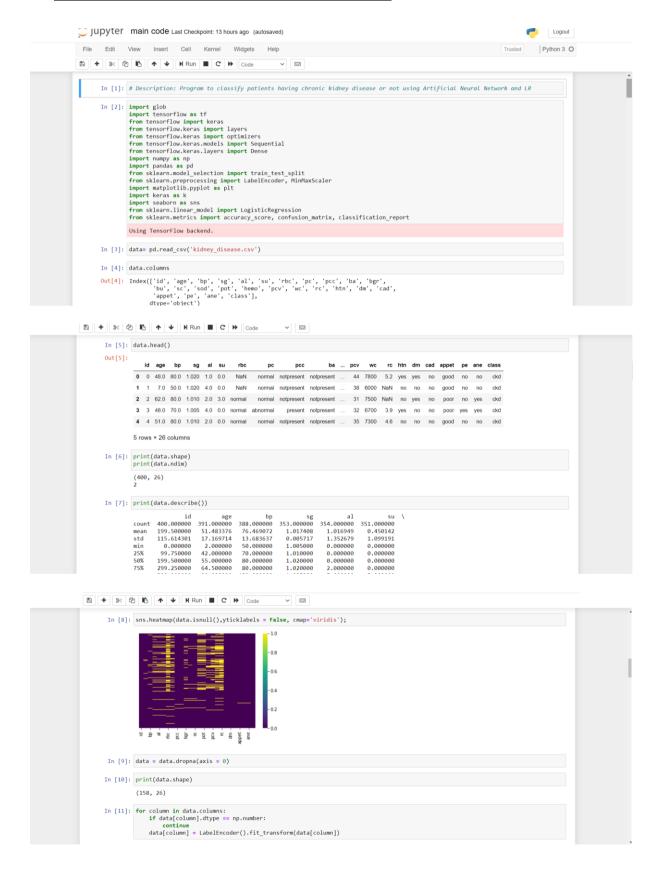


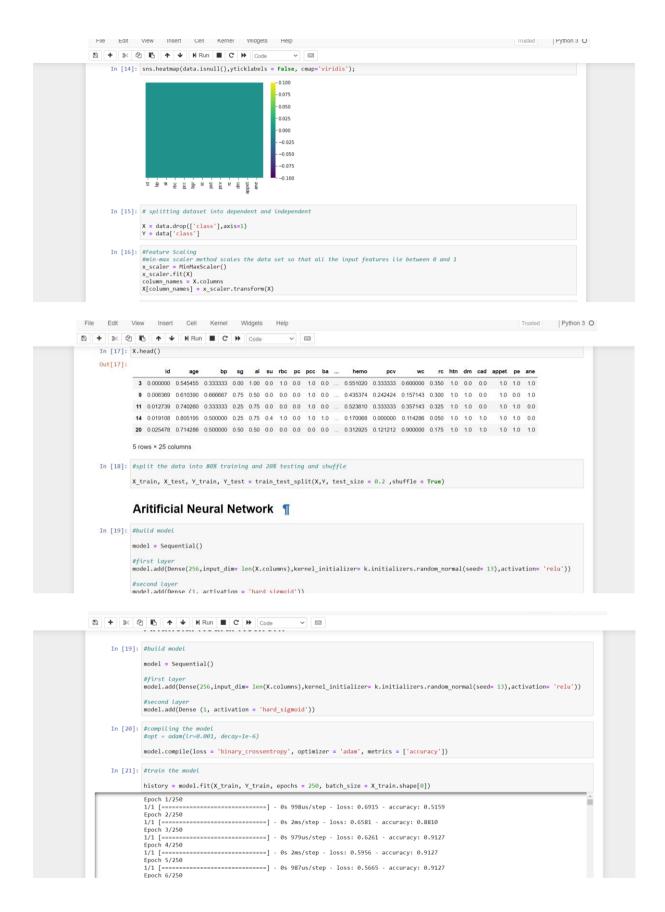
### **INTERACTION DIAGRAM**



### IMPLEMENTATION AND TESTING (SNAP SHOTS WITH DESCRIPTION)

### **IMPLEMENTATION DETAILS (SNAPSHOTS)**

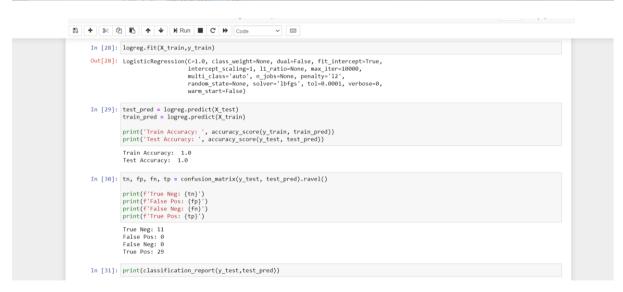




```
In [24]: Print("shape of training data: ",X_train.shape)

print("shape of training data: ",X_train.shape)
shape of training data: (126, 25)
shape of test data: (32, 25)
```

```
| Solution | Manage |
```



#### **TESTING**

#### TESTCASES (FOR ALL MODULES AS PER THE TEMPLATE)

#### **ORIGINAL DATA**

File Edit View Language current mod

```
37,77,80,,,,,,,notpresent,notpresent,137,65,3.4,141,4.7,9.7,28,6900,2.5,yes,yes,no,poor,no,yes,"ckd=""
38,69,01,102,3,0,ahonomal,notpresent,notpresent,193,4.1,132,5.9,12.5,,yes,no,no,good,no,no,ckd
40,46,90,1.01,2,2,normal,notpresent,notpresent,198,4.1,131,25.9,12.5,yes,no,no,good,no,no,ckd
41,45,70,1.01,0,0,normal,notpresent,notpresent,pore,99,80,2.1,1,11,13,2,100,4,1yes,no,"—no",good,no,no,ckd
41,45,70,1.01,0,0,normal,notpresent,notpresent,20,0.7,,,,no,no,no,good,yes,no,ckd
42,47,100,1.01,0,0,normal,notpresent,notpresent,20,0.7,,,,no,no,no,good,yes,no,ckd
43,35,80,1.01,1,0,ahonomal,notpresent,notpresent,79,202,10.8,134,3.4,7.9,24,7900,3.1,no,yes,no,good,no,no,ckd
44,54,80,1.02,3,0,abnormal,notpresent,notpresent,79,77,6.3,134,4.8,9.7,28,,yes,yes,no,poor,yes,no,ckd
44,54,80,1.02,3,0,abnormal,notpresent,notpresent,70,77,6.3,134,4.8,9.7,28,,yes,yes,no,poor,yes,no,ckd
44,54,80,1.02,3,0,abnormal,notpresent,notpresent,124,24,1.2,142,4.2,12.4,37,6400,4.7,no,yes,no,good,no,no,ckd
47,70,1.01,80,0,normal,notpresent,notpresent,170,32,0,9125,410,29,18900,35,yes,yes,no,good,yes,no,dod,94,77,01,01,80,0,normal,notpresent,notpresent,70,32,0,9125,410,29,18900,35,yes,yes,no,good,yes,no,ckd
48,70,1.015,0,0,normal,notpresent,notpresent,170,32,0,9125,410,29,18900,35,yes,yes,no,good,yes,no,ckd
49,70,10,10,2,0,normal,notpresent,notpresent,170,52,0,9125,410,29,18900,35,yes,yes,no,good,yes,no,ckd
50,53,60,,,,notpresent,notpresent,120,61,61,364,410,13,33,,yes,yes,no,poor,yes,yes,ckd
51,53,60,,,,notpresent,notpresent,notpresent,130,64,8500,47,9es,yes,no,poor,yes,yes,ckd
52,53,90,1015,0,0,normal,notpresent,notpresent,101,31,64,0,8500,47,yes,yes,no,poor,yes,yes,no,ckd
53,15,4100,1015,3,0,normal,notpresent,notpresent,13,13,64,9800,47,yes,yes,no,poor,yes,yes,yes,no,ckd
54,63,80,1015,0,0,notpresent,notpresent,notpresent,130,49,800,47,yes,yes,no,non,no,ckd
55,75,90,1,015,0,0,normal,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresent,notpresen
```

| The | Edit | View | Language |

```
304, 33, 80, 1.025, 0, 0, normal, normal, notpresent, notpresent, 128, 38, 0.6, 135, 3.9, 13.1, 45, 6200, 4.5, no, no, no, good, no, no, notckd 307, 305, 41, 80, 1.02, 0, 0, normal, notpresent, notpresent, 122, 25, 0.8, 138, 5, 17.1, 2, 29, 4300, 5.7, no, no, good, no, no, notckd 308, 306, 52, 80, 1.02, 0, 0, normal, notpresent, notpresent, 128, 30, 1.2, 140, 4.5, 15, 12, 20, 4300, 5.7, no, no, no, good, no, no, notckd 309, 307, 47, 60, 1.02, 0, 0, normal, notpresent, notpresent, 128, 30, 1.2, 140, 4.5, 15, 13.5, 44, 7900, 4.5, no, no, no, good, no, no, notckd 310, 308, 43, 80, 1.025, 0, 0, normal, notpresent, notpresent, 81, 46, 0.6, 135, 49, 49, 86, 690, 4.9, no, no, no, good, no, no, notckd 311, 309, 51, 60, 1.02, 0, 0, ., notpresent, notpresent, 129, 27, 0, 7, 142, 4, 9, 13, 44, 11080, 5.2, 4, no, no, no, good, no, no, notckd 311, 315, 56, 60, 1.02, 0, 0, normal, normal, notpresent, notpresent, 102, 27, 0, 7, 142, 4, 9, 13, 44, 11080, 5.4, 4, no, no, no, good, no, no, notckd 312, 80, 70, 1.02, 0, normal, normal, notpresent, notpresent, 132, 18, 1.1, 147, 4.7, 13.7, 45, 7500, 5.6, no, no, no, good, no, no, notckd 312, 80, 70, 1.02, 0, normal, normal, notpresent, notpresent, 104, 28, 109, 142, 48, 117, 3.52, 8200, 4.8, no, no, no, good, no, no, notckd 314, 39, 70, 1.025, 0, normal, normal, notpresent, notpresent, 132, 18, 1.1, 147, 4.7, 13.7, 45, 7500, 5.6, no, no, no, good, no, no, notckd 314, 319, 70, 1.025, 0, normal, normal, notpresent, notpresent, 104, 28, 109, 142, 48, 117, 13, 52, 2200, 4.8, no, no, no, good, no, no, notckd 316, 314, 39, 70, 1.025, 0, normal, normal, notpresent, notpresent, 1, 135, 44, 11, 15, 3, 48, 5800, 4.7, 10, no, no, no, good, no, no, notckd 318, 61, 701, 1.025, 0, 0, normal, normal, notpresent, notpresent, 102, 48, 11, 1219, 4., 315, 51, 52, 700, 100, no, no, no, good, no, no, notckd 318, 61, 701, 1.025, 0, 0, normal, normal, notpresent, notpresent, 120, 29, 0, 7, 137, 3.5, 17, 4, 52, 7000, 5, 3, no, no, no, good, no, no, notckd 317, 600, 102, 0, 0, normal, normal, notpresent, n
```

### DATA AFTER PREPROCESSING

Data is cleaned and scaled using drop axis method, then min-max scaler.

```
Su rbc pc pcc ba ...

9 0.0 1.0 0.0 1.0 0.0 ...

5 0.0 0.0 0.0 1.0 0.0 ...

75 0.4 1.0 0.0 1.0 1.0 ...

50 0.0 0.0 0.0 0.0 0.0 ...

1.0 1.0 0.0 0.0 ...

1.0 1.0 0.0 0.0 ...

9 0.0 0.0 0.0 0.0 0.0 ...

1.0 1.0 0.0 0.0 ...

9 0.0 0.0 0.0 0.0 0.0 0.0 ...
In [17]: print(X)

        id
        age
        bp
        sg
        al

        0.000000
        0.545455
        0.333333
        0.00
        1.00

        0.066369
        0.616930
        0.666667
        0.75
        0.50

        0.012739
        0.740260
        0.333333
        0.25
        0.75

        0.019108
        0.805195
        0.500000
        0.25
        0.75

        0.025478
        0.714280
        0.500000
        0.50
        0.50

                                         14
20
                                                                                                                                                          0.500000
0.333333
0.500000
0.166667
0.500000
                                         0.75 0.00
1.00 0.00
0.75 0.00
1.00 0.00
1.00 0.00

        hemo
        pcv

        0.551020
        0.333333

        0.435374
        0.242424

        0.523810
        0.333333

        0.170068
        0.000000

        0.312925
        0.121212

                                                                                                                                                                                                                                                                                                                      ppet pe ane
1.0 1.0 1.0
1.0 0.0 1.0
1.0 1.0 0.0
1.0 1.0 0.0
1.0 1.0 1.0
                                                                                                                                                                                                                                                                                                              appet
1.0
1.0
1.0
                                                                                                                                                          wc rc htn dm cad
0.600000 0.350 1.0 0.0 0.0
0.157143 0.300 1.0 1.0 0.0
0.557143 0.325 1.0 1.0 0.0
0.114286 0.050 1.0 1.0 1.0
0.900000 0.175 1.0 1.0 1.0
                                                                                                                                                          0.600000 0.575 0.0 0.0 0.0 0.0 0.742857 0.900 0.0 0.0 0.0 0.0 0.55714 0.700 0.0 0.0 0.0 0.671429 0.825 0.0 0.0 0.0 0.0 0.614286 0.875 0.0 0.0 0.0 0.0
                                          .. ... ... ... ... 395 0.857143 0.757576
                                                                                                                                                                                                                                                                                                                      0.857143 0.757576
0.911565 0.969697
0.863946 0.818182
0.755102 0.878788
                                         396
397
                                          398
                                          399 0.863946 0.939394
                                         [158 rows x 25 columns]
```

### CONCLUSION, LIMITATIONS AND SCOPE FOR FUTURE WORK

The first time in the process of diagnosing CKD using machine learning techniques. \
Then, four classification algorithms were explored, namely: ANN, SVM, Naïve
Bayes, and k-NN. The performance of each of these classifiers was examined by the
classification accuracy, precision, recall, and f-measure achieved by the classifier.

ANN, SVM, and NB all achieved an accuracy of 98% while k-NN achieved an
accuracy of 93.9%. Further research can be done to exceed the classification accuracy
currently achieved, by using different classifiers or feature selection methods.

This project can be used as a prototype to develop a healthcare system for CKD
patients. To further assess the performance of the model, testing the model with large
number of data will help to analyze the accuracy levels of the current model with
more accuracy.

### **REFERENCES**

- [1] I. Veerappan and G. Abraham, "Chronic kidney disease: current status, challenges and management in India.," Nephrology, no. Table 1, pp. 593–7, 2008.
- [2] H. Almutary, A. Bonner, and C. Douglas, "Chronic kidney disease in Saudi Arabia: a nursing perspective," Middle East J Nurs., vol. 7, no. 6, pp. 17–26, 2013.
- [3] IHME, "Saudi Arabia | IHME." [Online]. Available: http://www.healthdata.org/saudi-arabia. [Accessed: 31-Oct-2017].
- [4] A. a Al-Sayyari and F. a Shaheen, "End stage chronic kidney disease in Saudi Arabia. A rapidly changing scene.," Saudi Med. J., vol. 32, pp. 339–346, 2011.
- [5] G. D. Magoulas and A. Prentza, "Machine Learning in Medical Applications," Springer, Berlin, Heidelberg, 2001, pp. 300–307.
- [6] L.Jerlin Rubini, "UCI Machine Learning Repository: Chronic\_Kidney\_Disease Data Set," 2015. .
- [7] L. Jena and N. Ku Kamila, "Distributed Data Mining Classification Algorithms for Prediction of ChronicKidney-Disease," Int. J. Emerg. Res. Manag. &Technology, vol. 9359, no. 4, pp. 2278–9359, 2015.
- [8] M. Kumar, "Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm," Int. J. Comput. Sci. Mob. Comput., vol. 5, no. 2, pp. 24–33, 2016.
- [9] E. Celik, M. Atalay, and A. Kondiloglu, "The Diagnosis and Estimate of Chronic Kidney Disease Using the Machine Learning Methods," no. 4, pp. 27–31, 2016. [10] A. Salekin and J. Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes," Proc. 2016 IEEE Int. Conf. Healthc. Informatics, ICHI 2016, pp. 262–270, 2016.
- [11] "R: What is R?" [Online]. Available: https://www.rproject.org/about.html. [Accessed: 20-Dec-2017].
- [12] S.Ramya and N.Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," Res. J. Pharm. Biol. Chem. Sci., vol. 7, no. 3, pp. 365–373, 2016.
- [13] A. S. Nur, N. H. Mohd Radzi, and A. O. Ibrahim, "Artificial Neural Network Weight Optimization: A Review," TELKOMNIKA Indones. J. Electr. Eng., vol. 12, no. 9, pp. 6897–6902, 2014.

- [14] S. O. Olatunji, M. S. Al-Ahmadi, M. Elshafei, and Y. A. Fallatah, "Forecasting the Saudi Arabia stock prices based on artificial neural networks model," Int. J. Intell. Inf. Syst., vol. 2, no. 5, pp. 77–86, 2013.
- [15] D. Kriesel, "A Brief Introduction to Neural Networks," Retrieved August, p. 244, 2005.
- [16] D. Ciobanu, "Mathematical and Quantative Methods Using SVM for Classification 1 Introduction," vol. 8, no. 13, pp. 207–222.
- [17] J. M. Moguerza and A. Muñoz, "Support Vector Machines with Applications," Stat. Sci., vol. 21, no. 3, pp. 322–336, 2006.
- [18] M. Rajnoha, R. Burget, and M. K. Dutta, "Offline handwritten text recognition using support vector machines," 2017 4th Int. Conf. Signal Process. Integr. Networks, SPIN 2017, pp. 132–136, 2017.
- [19] S. O. Olatunji, "Extreme Learning machines and Support Vector Machines models for email spam detection," Can. Conf. Electr. Comput. Eng., pp. 1–6, 2017.
- [20] A. Kampouraki, D. Vassis, P. Belsis, and C. Skourlas, "e-Doctor: A Web based Support Vector Machine for Automatic Medical Diagnosis," Procedia Soc. Behav. Sci., vol. 73, pp. 467–474, 2013.
- [21] R. Amami, D. Ben Ayed, and N. Ellouze, "An Empirical Comparison of SVM and Some Supervised Learning Algorithms for Vowel recognition," Int. J. Intell. Inf. Process., vol. 3, no. 1, pp. 63–70, 2012.
- [22] Z. Gavrilov, "SVM tutorial," Slides, p. 33, 2003.
- [23] E. Krishnaveni, K;Radhamani, "Diagnosis and evaluation of ADHD using Naïve Bayes and J48 classifiers," in Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on, 2016, pp. 1809–1814.
- [24] T. I. Cannings, "New Approaches to Modern Statistical Classification Problems," no. August, 2015.
- [25] H. Sethi, A. Goraya, and V. Sharma, "Artificial Intelligence based Ensemble Model for Diagnosis of Diabetes," Int. J. Adv. Res. Comput. Sci., vol. 8, no. 4, pp. 1540–1549, 2017.
- [26] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA Workbench," Morgan Kaufmann, Fourth Ed., pp. 553–571, 2016.
- [27] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly,

- B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," pp. 1–15, 2013.
- [28] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization," in Advances in Neural Information Processing Systems (NIPS), 2011, pp. 2546–2554.
- [29] K. . Vijayarekha, "Activation Functions," NPTEL Electron. Commun. Eng. Pattern Recognit., pp. 1–6.
- [30] Lars Hulstaert, "Understanding Activation Functions in Neural Networks," pp. 1–10, 2017.
- [31] A. Karpathy, "CS231n Convolutional Neural Networks for visual recognition.," 2017. .
- [32] P. Moritz, R. Nishihara, and M. I. Jordan, "A Linearly-Convergent Stochastic L-BFGS Algorithm," vol. 41, 2015.
- [33] K. Hechenbichler and K. Schliep, "Weighted kNearest-Neighbor Techniques and Ordinal Classification," Mol. Ecol., vol. 399, p. 17, 2004.
- [34] M. M. Roche and P. P. Wang, "Factors associated with a diabetes diagnosis and late diabetes diagnosis for males and females," J. Clin. Transl. Endocrinol., vol. 1, no. 3, pp. 77–84, 2014.