Design and Development of Chronic Kidney Disease Management System Using Machine Learning

BY

Nafiz Ahmed Emon
ID: 191-15-12368
&
Sharif Ahamed
ID: 191-15-12406

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. Md Zahid Hasan
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/internship titled **"Design and Development of Chronic Kidney Disease Management System Using Machine Learning"**, submitted by Nafiz Ahmed Emon, ID No: 191-15-12368 and Sharif Ahamed, ID No: 191-15-12406 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 26-01-2023.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Subhenur Latif
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mohammad Monirul Islam
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

26-01-23

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

I hereby declare that, this project has been done by **Nafiz Ahmed Emon** and **Sharif Ahamed** under the supervision of **Md Zahid Hasan, Assistant professor, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Md Zahid Hasan
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By:

Mr. Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Nafiz Ahmed Emon
ID: 191-15-12368
Department of CSE
Daffodil International University

Sharif Ahamed
ID: 191-15-12406
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully. We really grateful and wish our profound our indebtedness to **Dr. Md Zahid Hasan, Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of machine learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan**, **Professor & Head,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Kidney disease is a chronic condition because it causes a progressive loss of kidney function over time. When kidney function is impaired, the body can no longer perform all of these essential functions, leading to a range of symptoms, including swelling, fatigue, and shortness of breath. That's why early detection of chronic kidney disease is important because it allows for early intervention, which can slow the progression of the disease and prevent or delay the development of end-stage renal failure. Early detection can also help identify treatments to reduce or prevent further damage to the kidneys. The study aims to develop an automated system to predict CKD, identify the stage of CKD, find out the probability of CKD, predict CKD progression, and manage chronic kidney disease. The study proposes a machine-learning-based system using supervised learning algorithms to predict the onset of CKD accurately. The system is designed to use a combination of patient data, such as age, gender, ethnicity, family history, lifestyle, and laboratory test results, to predict and determine the probability of the onset of CKD. In addition, the system is designed to provide personalized recommendations for lifestyle modifications and all the relevant information related to CKD, such as symptoms, risk factors, causes, nutrition, diagnosis and testing, and treatments, and to create awareness about CKD. The study also proposes using interactive visualizations to help patients understand their condition and make informed decisions about their health. The study also outlines a procedure for validating the accuracy of the system and provides recommendations for further development. To predict the CKD, we have applied different machine learning algorithms, including Decision Tree, Ada Boost, KNN, Random Forest, Gradient Boosting, Stochastic Gradient Boosting, XGBoost, and Extra Tree. Of these, Decision Tree and AdaBoost have higher accuracy, at 96.25% each.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

Chronic kidney disease (CKD) is a long-term kidney disease that progresses slowly over several years. It is characterized by a gradual loss of kidney function and can lead to complete kidney failure. Common causes of CKD include diabetes, high blood pressure, and long-term use of certain medications. People with CKD may experience symptoms such as fatigue, swelling, and high blood pressure [1][2][11]. Treatment options for CKD can include lifestyle changes, medications, dialysis, and kidney transplantation. Early detection and treatment of CKD can help to prevent further complications [3].

Kidney disease is a chronic condition because it is a progressive loss of kidney function over time. The kidneys play a vital role in the body by filtering waste from the bloodstream, regulating electrolytes, and balancing the body’s fluids. When kidney function is impaired, the body can no longer perform all of these essential functions, leading to a range of symptoms, including swelling, fatigue, and shortness of breath [11]. Over time, kidney disease can lead to serious complications, including kidney failure and death. As such, it is considered a chronic condition that requires ongoing management and monitoring. That's why early detection of Chronic Kidney Disease (CKD) is important because it allows for early intervention, which can slow the progression of the disease and prevent or delay the development of end-stage renal failure. Early detection can also help to identify treatments to reduce or prevent further damage to the kidneys, including lifestyle changes, medications, and other treatments [9]. Early detection can also lead to earlier diagnosis of other medical conditions that are associated with CKD, such as diabetes, hypertension, and anemia. With early detection and treatment, patients can often manage their condition and lead healthier life [4][6][10].

Chronic Kidney Disease is evaluated through physical examination, laboratory testing and imaging studies. A physical exam may include blood pressure measurements, evaluation of the heart, lungs and abdomen, and urinalysis [7]. Laboratory tests may include a complete blood count, serum creatinine and blood urea nitrogen, electrolytes, uric acid, albumin, and calcium. Imaging studies such as ultrasounds, CT scans, and MRI scans are used to detect any changes in the size and shape of the kidneys. Once diagnosed, CKD will be staged according to the severity of kidney damage and the glomerular filtration rate (GFR) [5].

Table 1.1: Different Stages of CKD

Different Stages of CKD		
Stage	**GFR ($ml/min/1.73m^2$)**	**Qualitative Description**
1	>90	Normal Kidney Function - This is when the kidneys are functioning normally and all of the tests show that they are working properly
2	60-89	Mildly Impaired Kidney Function - This is when the kidney function is slightly impaired and certain tests may indicate the beginning of a decline in kidney performance
3	30-59	Moderately Impaired Kidney Function - This is a more serious form of impaired kidney function, where test results may show a decline in performance compared to stage 2
4	15-29	Severely Impaired Kidney Function - At this stage, the kidneys are significantly impaired and are often unable to filter toxins out of the body effectively
5	<15	Kidney Failure - This is a critical stage where the kidneys are no longer able to filter toxins and waste products out of the body. This stage requires dialysis or a kidney transplant to keep the patient alive

The aim of this study is to develop a machine learning-based model to predict CKD and to develop a CKD management system that can aid in the treatment and management of CKD. The model should accurately identify the risk factors associated with CKD and be able to predict the development of CKD. Additionally, the management system should provide personalized guidance and advice to individuals with CKD to help them better manage their condition by providing all the information related to CKD.

1.2 Motivation

There is a prevalence of CKD (Chronic Kidney Disease) patients in recent years. Several factors have contributed to the increasing prevalence of CKD patients. These include an aging population, rising rates of obesity, an increase in sedentary lifestyles, rising rates of hypertension, diabetes, and other chronic conditions, increasing use of medications and environmental toxins, and an increase in the use of long-term dialysis treatments. In addition, the rate of CKD has been rising in many parts of the world due to a lack of access to healthcare and preventive measures. Therefore, there is a need for CKD management system because this progressive decline in kidney function can cause many complications and increase the risk of mortality if not monitored and managed in a timely manner. Our proposed system would allow to predict and track the progression of CKD in a patient, giving them the opportunity to take necessary steps to manage and minimize the complications associated with the disease. Additionally, predictive analytics could allow for early detection and intervention of declining kidney function, allowing for improved outcomes and better quality of life for those living with CKD.

1.3 Rationale of the study

The study of chronic kidney disease management system is important for improving the quality of life of individuals who suffer from the condition. Chronic kidney disease (CKD) is a progressive condition that can lead to end-stage renal failure, requiring dialysis or a kidney transplant. Early detection and intervention are essential for reducing the risk of complications associated with CKD. The development of a system that can accurately predict and manage CKD could help to identify those at risk before they experience serious symptoms and provide them with necessary interventions to slow the progression of the disease. Such a system could also provide health care providers with the ability to monitor their patients' conditions and provide personalized care.

1.4 Research Questions:

- How accurately this CKD Management System predict CKD?
- How does the CKD Management System help improve patient care?
- What benefits does the CKD Management System provide to patients?
- How effective is the CKD Management System in helping to identify and address potential health risks?
- What are the cost savings associated with using the CKD Management System?
- How does the CKD Management System help to reduce healthcare costs?
- What are the most common challenges associated with implementing the CKD Management System?

1.5 Research Objective:

- To build an effective CKD management system that can accurately predict CKD, and detect CKD stages, allowing for earlier diagnosis and intervention.
- To provide all the information related to CKD.
- To improve patient outcomes by predicting potential risks of chronic kidney disease (CKD) and suggesting personalized treatment plans.
- To integrate machine learning with current standards of practice to enable optimized and effective CKD management.

1.6 Expected Outcome:

Prediction: Based on the patient's symptoms and laboratory results, the system should predict that the patient is likely to have CKD.

Management: The system should provide a personalized management plan for the patient which includes lifestyle changes, medications, and other treatments which are tailored to their individual needs. Additionally, the system will provide resources for the patient and their family to help them stay informed about CKD and its treatment.

Treatment plan: Patient should receive lifestyle modification counseling and should consider lifestyle changes such as reducing sodium and protein intake, increasing physical activity, monitoring blood pressure and blood glucose, limiting alcohol intake and avoiding smoking. In addition, patient should take any medications that may be prescribed to help manage symptoms and control disease progression.

Education: Patient should receive education about CKD and its potential long-term implications such as different staging of CKD, kidney failure, dialysis, and potential need for kidney transplant. Patient should also understand importance of medication adherence, healthy diet and lifestyle changes, and importance of regular checkups.

1.7 Project Management and Finance:

Project management phases of CKD management system:

1. Project Initiation

- ❖ Establish project objectives
- ❖ Develop project plan
- ❖ Define project scope

2. Requirements Analysis

- ❖ Collect requirements
- ❖ Analyze requirements
- ❖ Define system architecture
- ❖ Create use cases
- ❖ Identify data sources

3. Design

- ❖ Design user interface
- ❖ Design system architecture
- ❖ Develop database schema
- ❖ Create data models
- ❖ Develop algorithms

4. Implementation

- ❖ Develop code
- ❖ Test code
- ❖ Integrate system components
- ❖ Deploy system

5. Testing

- ❖ Develop a test plan
- ❖ Execute test plan
- ❖ Identify and fix bugs
- ❖ Verify the accuracy of results

6. Deployment

- ❖ Train users
- ❖ Document system
- ❖ Monitor system performance

- ❖ Monitor user feedback
- ❖ Maintain system

Finance:

CKD management system requires:

- ❖ Server cost
- ❖ publishing cost in Google's Play Store
- ❖ publishing cost in Apple's App Store

1.8 Report Layout

Chapter 1: This section includes the research introduction, motivation, rationale of the the study, research questions, and report layout.

Chapter 2: This section is about terminologies, related work, comparative analysis and summary, the scope of the problem, and challenges.

Chapter 3: This section includes the research subject and instrumentation, implementation requirements, data preprocessing, static analysis, applied algorithms, and data collection.

Chapter 4: This section includes experimental setup, experimental results, and analysis and result discussion.

Chapter 5: This section is about the impact on society and the environment, ethical aspects, and sustainable plan.

Chapter 6: This section includes the outline of the study, conclusions, suggestions for advance consider.

CHAPTER 2
Background

2.1 Terminologies

Chronic kidney disease is a major health concern as the ratio of CKD patient are increasing day by day. Doctors and researchers are working hard to find the remedy and prevention of CKD. They conducted different studies on this subject and implemented different Machine Learning and Deep Learning algorithm. Many of them have achieved significant success. All of this research has made a markable change in the detection and treatment of CKD patients. In our study, we employed machine learning techniques.

2.2 Related Works

Due to the high payment of an expert specialist for the manual process of diagnosing a disease, it is costly and needs more time. So automatic disease detection and diagnosis system can be beneficial as it is cheap and fast. Many machine learning and deep learning methods like KNN, SVM, XGBoost, ANN, DT, NB, and LR have come a long way and applied to the diagnosis of diseases with success. One of the focus areas of this trend in AI-based diagnosis is Chronic Kidney Disease. [12] When the kidney is damaged and can't perform blood filtration at its normal rate, is known as Chronic Kidney Disease [13]. The prevalence of chronic kidney disease (has been identified as a major public health issue on a global scale. There are between 5 and 7 million people with kidney failure who require renal replacement treatment, and the estimated global prevalence of CKD is 13.4% [14]. Chronic kidney disease lowers the ability of filtering wastes from the body. As a result, the ratio of waste in blood increases. CKD patients could have issues like anemia, weak bones, nerve damage, poor nutritional health, and high blood pressure [15]. CKD patients need to monitor and take care of on regular basis. Controlling diabetics, medication intervals, a healthy diet, and regular checks up are the main concerns for a CKD patient [16]. CKD is classified as G1, G2, G3, G4, and G5 which are five stages determined by GFR and other parameters [17]. Many have made different models to detect and diagnose CKD with machine learning algorithms.

Models with Decision Tree (DT), Artificial Neural Network (ANN), and Naive Bayes (NB) were produced by Al-Hyari et al. [18]. The model's precisions for the DT, NB, and ANN are in order of 92.2%, 88.2%, and 82.4%. Although 102 data points were utilized, over-fitting might make them too few.

The association between eleven chronic diseases including CKD was established by Gupta et al. [19]. The AdaBoost algorithm performed with the highest accuracy of 88.66% when many algorithms were employed to identify chronic kidney disease.

K-NN, Random Forest, and neural networks were used in the models' derivation by Salekin and Stankovic [20]. For the KNN model and the Neural Network, they used values from the IBK method to fill in the dataset's missing data. To build a simplified model for a quick calculation, the features of their model were further cut back.

Using a proper diet plan, M. P. N. M. Wickramasinghe et al [21] provide an approach to control the disease. In this study, various methods, including Multiclass Decision Jungle, Multiclass Decision Forest, Multiclass Neural Network, and Multiclass Logistic Regression, are used to build classifiers.

Kernel-based Extreme Learning Machine (ELM) to detect Chronic Kidney Disease was suggested and tested by H. A. Wibawa et al [22]. Four kernel-based ELMs' performance, including RBF-ELM, Linear-ELM, Polynomial-ELM, and Wavelet-ELM, is compared to that of regular ELM. Better forecasting rates were obtained using Radial Basis Function - Extreme Learning Machine (RBF-ELM).

S. Revathy et al [23] presented a prediction algorithm that can predict CKD at an early stage using a Decision tree, Random Forest, and Support vector machine (SVM) machine learning models. Among them, Random Forest has the best accuracy of 99.16%.

Parul Sinha et al [24] predicted CKD using SVM and KNN machine learning algorithms. KNN performed better than SVM in their study with an accuracy of 78.75%.

2.3 Comparative Analysis and Summary

We have evaluated some earlier works which can relate to our study. We compared the outcomes to know which algorithms give us a better result. In those papers SVM, KNN, Random Forest, Decision Tree, ELM, Multiclass Decision Jungle, Multiclass Decision Forest, Multiclass Neural Network, Multiclass Logistic Regression, Artificial Neural Network, and Naive Bayes are used to build their model. In this section, we'll compare previous attempts side by side. The comparison is shown in Table 2.1.

Table 2.1: Comparative analysis of existing Articles

Work Title	Best Algorithm	Best Accuracy
XGBoost Model for Chronic Kidney Disease Diagnosis	XGBoost	98.7%
Clinical decision support system for diagnosis and management of Chronic Renal Failure	Decision Tree	92.2%
A method to predict diagnostic codes for chronic diseases using machine learning techniques	AdaBoost	88.66%
Chronic Kidney Disease Prediction using Machine Learning Models	Random Forest	99.16%.
Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM	Support vector machine	78.75%
Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease	Radial Basis Function - Extreme Learning Machine	99.38%

2.4 Scope of the Problem

After evaluating and studying other works and sources we have found that some of the models are effectively able to predict CKD's early stages. Some of the studies provide various information about the diagnosis of CKD patients. S. W. Ong et al [5] integrated a Smartphone–Based Self–Management System into the Usual Care of Advanced CKD. But it was not implemented. So, there is no whole system on which CKD patients can rely. There is no system where the user can predict whether someone has CKD or not, the stage of CKD, diagnosis module, meal recommendation, or AI-based inquiry system.

2.5 Challenges

- **Complexity of the disease:** CKD is a complex disease and it is difficult to predict its course and management.
- **User-friendliness**: The system must be easy to use so that patients and healthcare providers can use it effectively.
- **Accessibility:** Accessibility to the system must be ensured to ensure its effective use.
- **Technology:** In order for the CKD management system to be effective, it must be equipped with the latest technology for analysis, prediction, and management. This requires substantial investment and expertise to ensure that the system remains up-to-date and effective.
- **Regulatory challenges:** Privacy, data security, and ethical considerations are all important challenges when implementing a system of this scale, and require thoughtful consideration of applicable regulations and laws.
- **Patient engagement:** Collecting and analyzing healthcare data on a national scale is a daunting task, especially when it comes to patient engagement. It is important to ensure that patients are aware and engaging with the system while still protecting their privacy and security.

- ❖ **Interoperability:** Interoperability between different systems and components of the CKD prediction and management system is essential for its successful functioning. However, achieving interoperability between different systems is a major challenge.
- ❖ **User Interface:** Developing a user-friendly interface for the CKD prediction and management system is a challenge. The interface must be simple and intuitive so that users can easily access the system and understand its features.

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation

In our study, we intend to create a model that can predict whether a person has chronic kidney disease or not. We used a dataset from Kaggle to create this model. In machine learning algorithms, there are two ways to create a model. Machine learning is classified into two types: Supervised learning and unsupervised learning. We have used supervised learning to create our model. We have provided input as well as output data to train our model. Various classification models are used to tackle classification problems in supervised learning. Our study is concerned with the binary classification problem. As a result, we shall employ certain classification algorithms that work well and provide high accuracy with text data. We will employ XGBoost, Extra Trees Classifier, KNN, Ada Boost Classifier, Random Forest Classifier, Decision Tree Classifier, Gradient Boosting Classifier, and Stochastic Gradient Boosting models in our study. In the upcoming proposed methodology section, we will discuss all of the algorithms, how it works and which algorithms perform very well in our dataset. After that, we need to evaluate our model based on some criteria like precision, recall, and f1-score. We will briefly discuss all terms in the proposed methodology section.

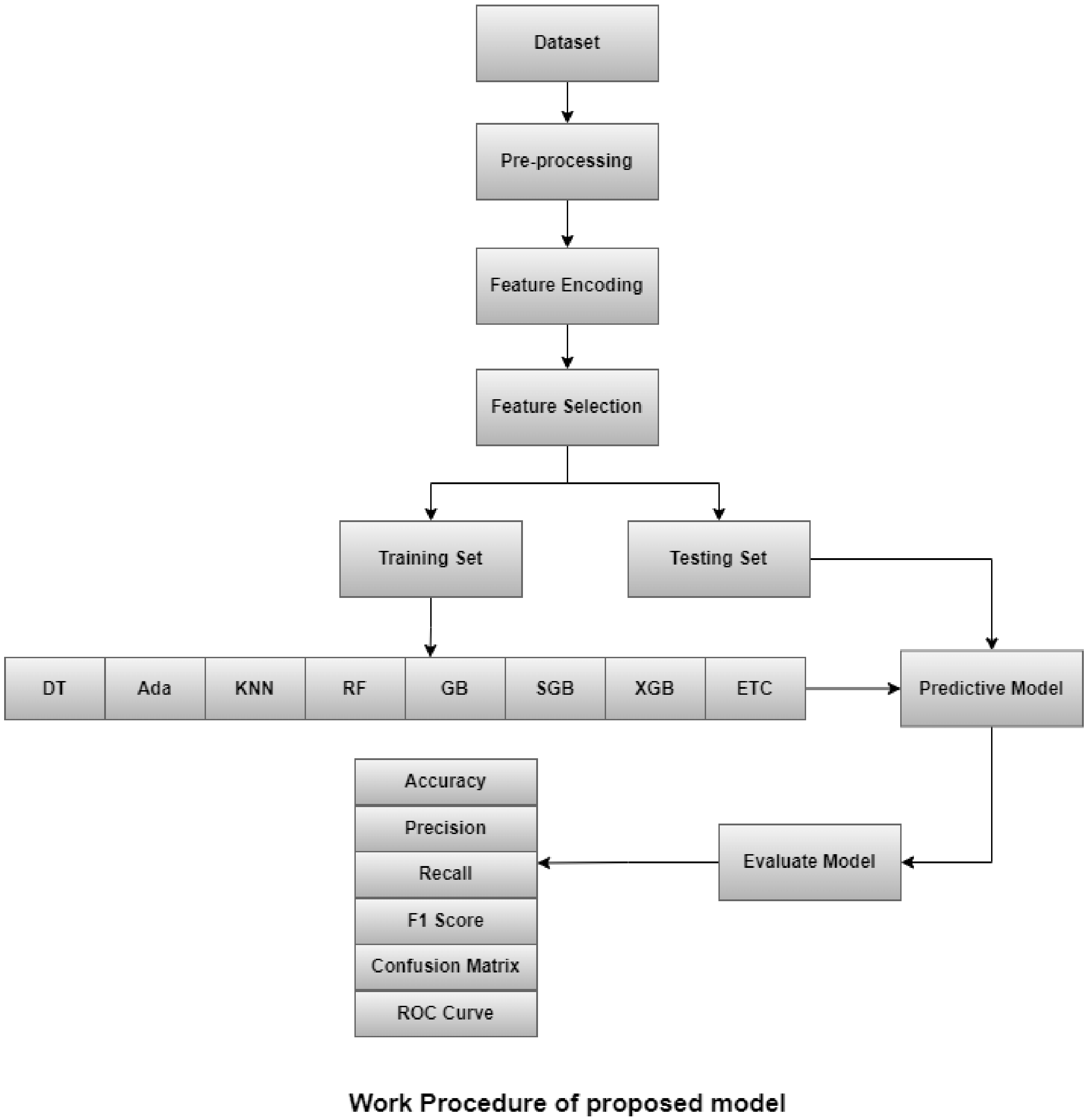


Figure 3.1: Work Procedure of Proposed Model

3.2 Data Collection Procedure

We have collected our data set from an online platform named Kaggle. In this dataset, there is a total of 400 entries. This dataset contains two specification classes, ckd (indicate by 0 in figure 3.2) and notckd (indicate by 1 in figure 3.2). There are 250 entries with chronic kidney disease.

Another 150 people aren’t affected by chronic kidney disease. The ratio of our specification class is given below.

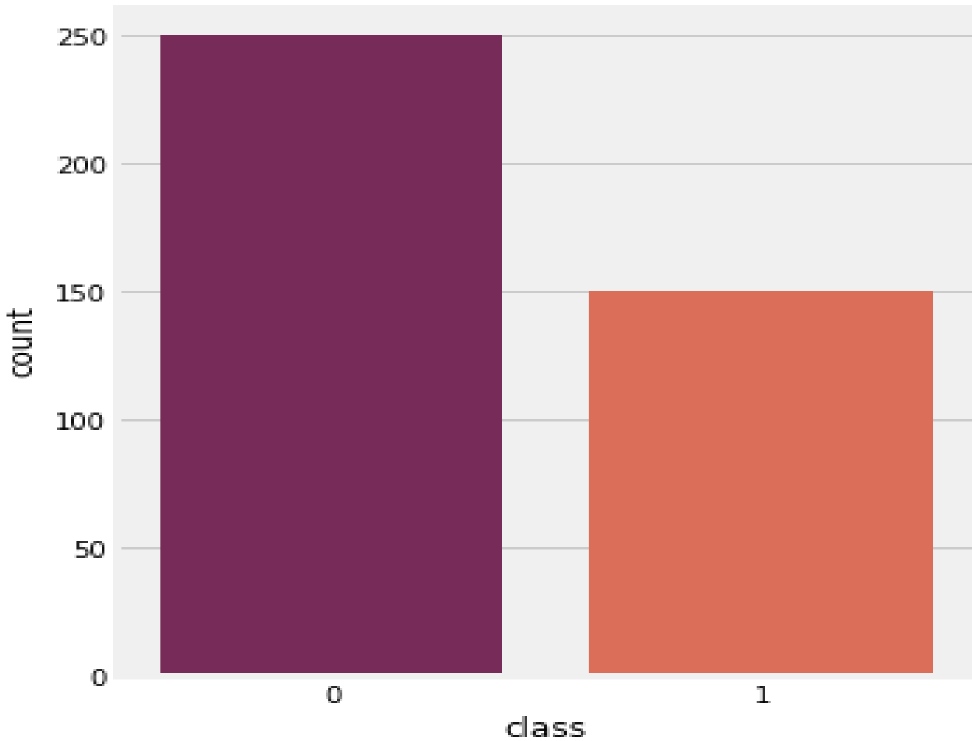


Figure 3.2: Class variables of the Dataset

3.3 Data Pre-processing

We can’t use raw text input in our classification model. The raw text file contains letters and symbols which is not appropriate to create a classification model from the dataset. These inappropriate letters or characters can downgrade the accuracy of our classification model. Besides in datasets, empty or dummy entries can be seen. We also find several empty values in our dataset. So, this dataset is useable to create a classification model. Even if we manage to construct the model, the result will be much below our expectations. Either we have to fill up the empty values with fitting values or we have to delete the entire row. Here we have used the first method. We have filled up the empty values with appropriate values.

If we talk more specifically, after analyzing the dataset we have found some columns are in object type. We have changed them into numerical data types. Then we searched for empty values. As this dataset contains empty values, we have to fill up them with values. To do this we have to do some steps. We have categorized the values of the categorical columns. After categorizing, we found incorrect values. We have replaced them with correct values and tried to keep the variety of values as minimum as possible.

In our dataset, excluding class variables, there are 25 variables. For better shape of our dataset, we have dropped some variables from the dataset based on dependencies. We have figured out the correlation between the variables by generating a heatmap. The dropped variables have less impact on the classification model compared to other variables. The less column we have the more user-friendly the dataset is.

After sorting the dataset, we filled in the null values. We have used two methods to fill up the null values. For higher null values we have used random sampling. On the other hand, for lower null values, we have used mean-mode sampling.

After handling all the missing values, we have done categorical feature encoding. We encode categorical data into numerical because math is generally performed using numbers. For this reason, our algorithms cannot run and process data if data are not in numerical form. That's why had to encode categorical data into numbers. As all of our categorical columns has only 2 categorical values, we have used label encoder to convert categorical data to numerical data.

At this point, we are set to feature selection and model building. Feature selection method is implemented to reduce the number of input variables by eliminating redundant and irrelevant features. We have used sklearn for feature selection.

3.4 Statistical Analysis

We have used a dataset of CKD patients from Kaggle which contains 400 records and 26 variables. Among 400 entries, 250 people have Chronic Kidney Disease. Another 150 doesn't contain Chronic Kidney Disease. So, our class variable contains two categorical values. Excluding class variable and id among other 24 variables, 14 variables are numerical (figure – 3.4.1) and other 10 are categorical (figure – 3.4.2).

Numerical variables contain only numbers. This number can be whole number or any fractional numbers.

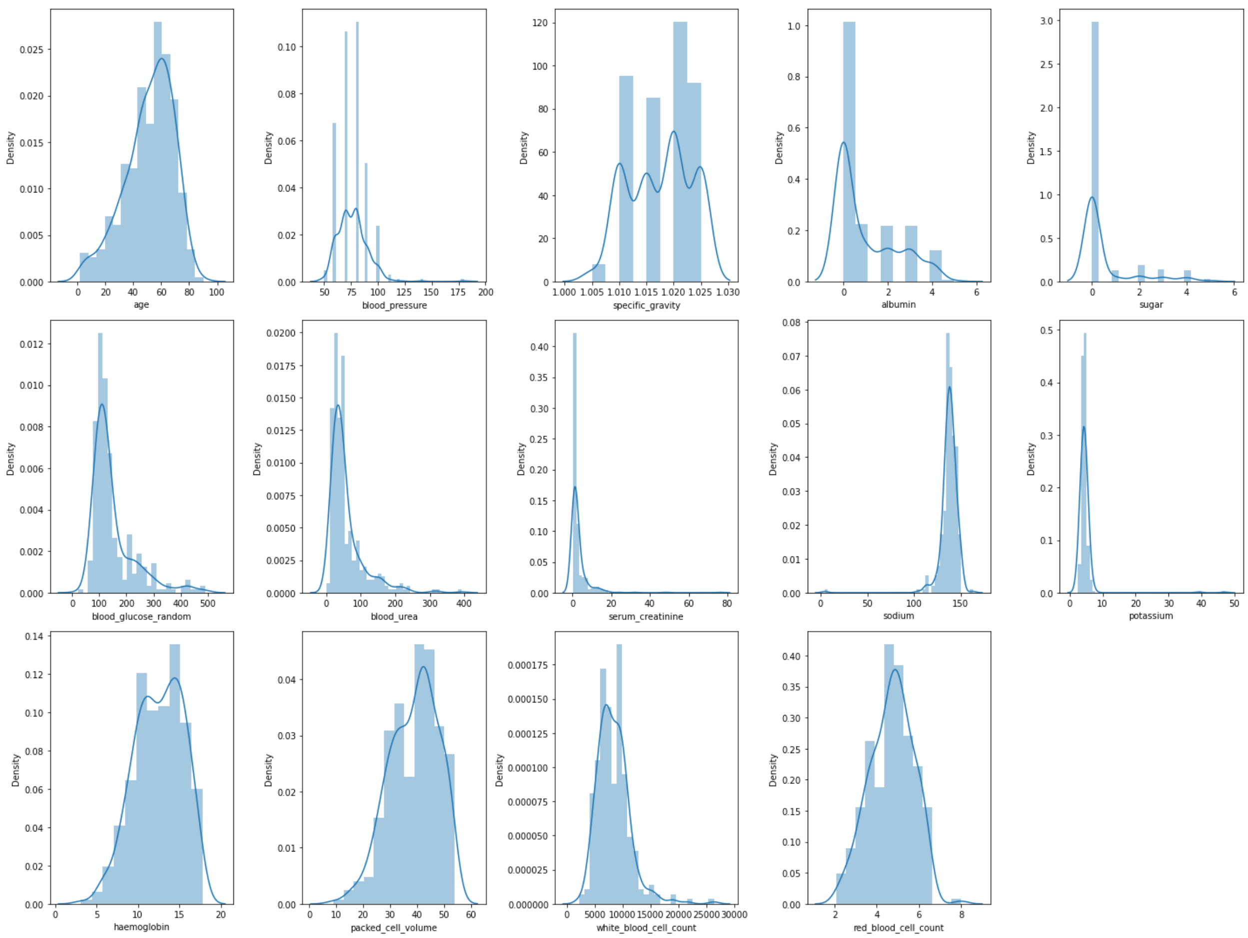


Figure – 3.4.1: Distribution of Numerical Variables

Unlike numerical variables, categorical variables don't contain number values. Instead of numbers categorical variables contain text data.

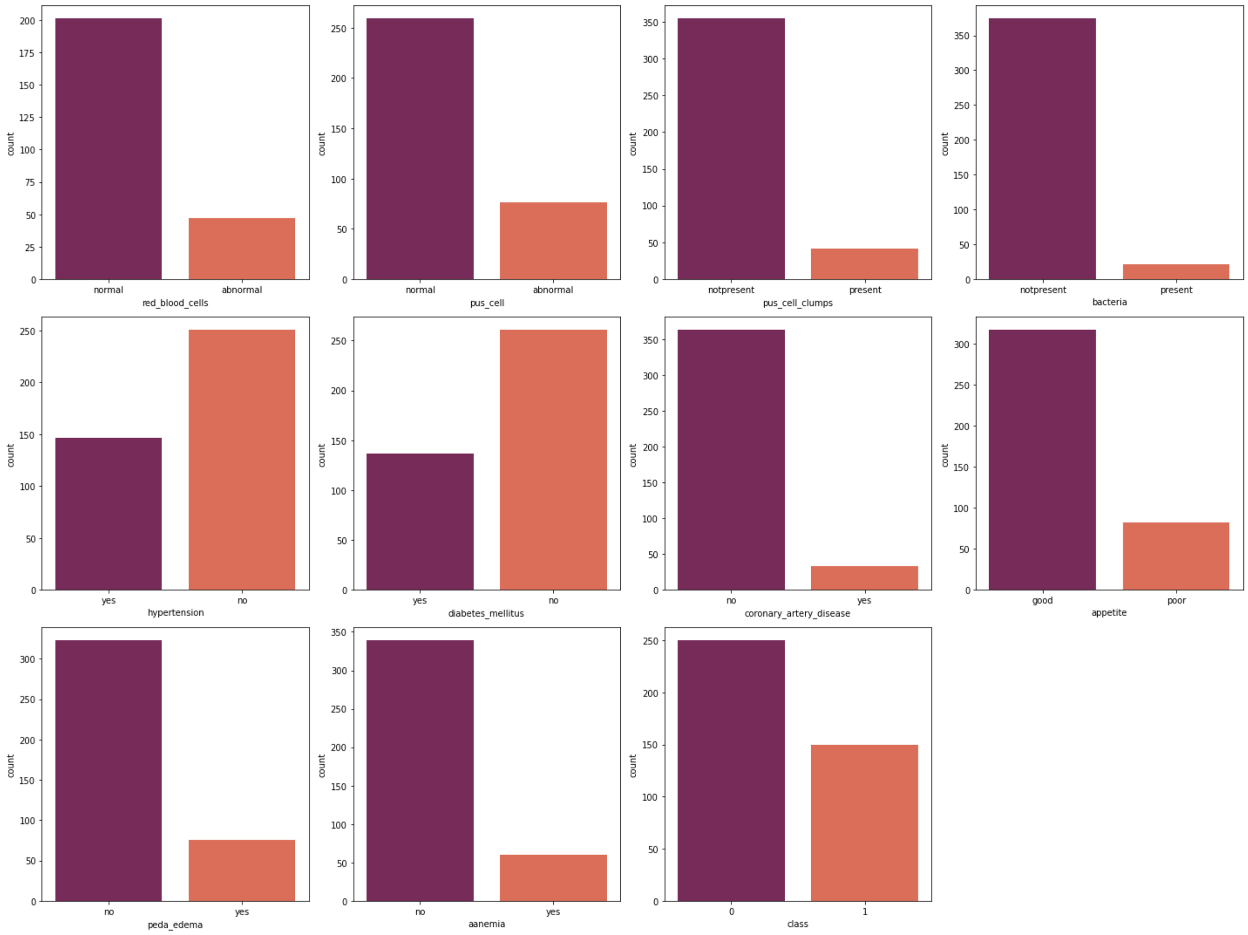


Figure – 3.4.2: Categorical Variables

We have generated heatmap of our dataset (figure – 3.4.3). Heatmap allow us to visualize the correlation between variables. Which helps us to analyze dataset and remove variables which has less impact on classification.

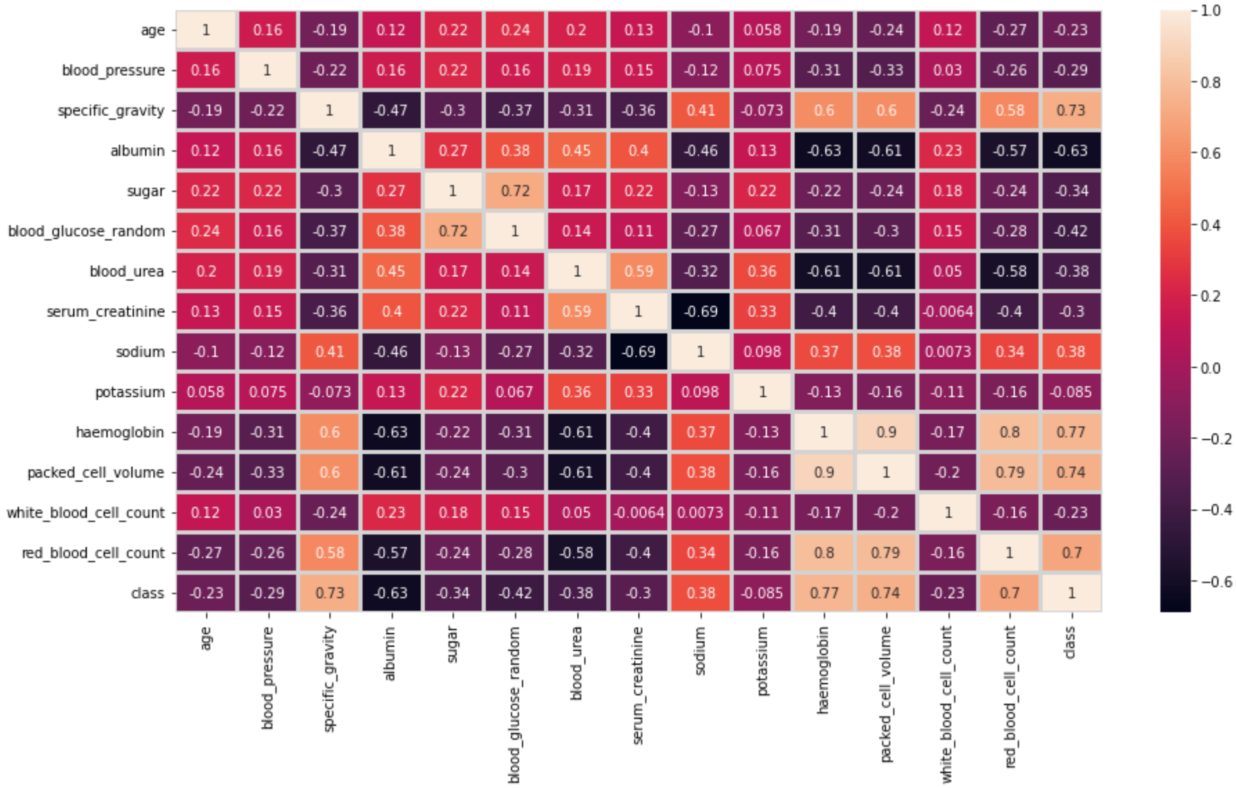


Figure – 3.4.3: Heatmap

After preprocessing of data, we have performed feature selection. Then we analyzed the score of feature selection.

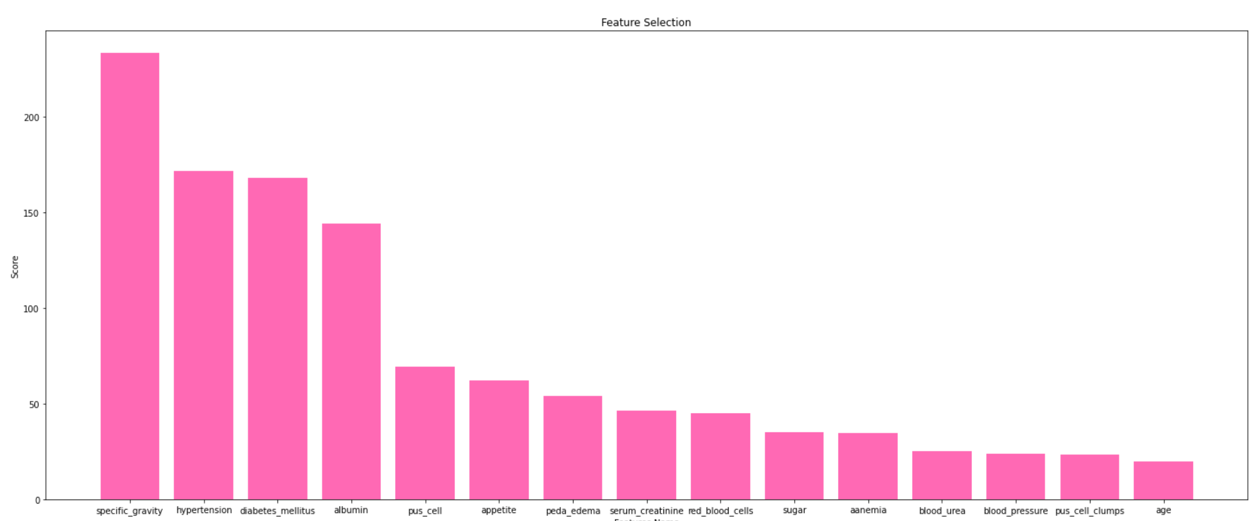


Figure – 3.4.4: Feature Selection Score

3.5 Applied Algorithm

In the next part, we built our model. As our work contains both input and output information to create a classifier model, we must employ a supervised approach. To build our model we have used 8 different machine-learning classification algorithms. XGBoost, Extra Trees Classifier, KNN, Ada Boost Classifier, Random Forest Classifier, Decision Tree Classifier, Gradient Boosting Classifier, and Stochastic Gradient Boosting are used to prepare our classification model. We separated our data into two parts: training samples and test samples. We took 80% data for training and 20% for testing.

3.5.1 Extreme Gradient Boosting

XGBoost is a machine learning system based on decision trees that use gradient boosting techniques. This technique can accurately predate data when the data is unstructured, such as text or image data [25]. This approach is commonly utilized to solve regression and classification problems. In our categorization task, our method successfully anticipated the outcome. This classifier predicted with a high degree of accuracy 95%.

3.5.2 Extra Tree

Extra Tree Classifier is a decision tree-based ensemble learning algorithm that is used for both classification and regression. It works by randomly selecting a set of features to be used for splitting the data, and then using those features to create a decision tree. Extra Tree Classifier is considered an improvement on the traditional decision tree model as it is more accurate and robust than the standard decision tree approach. The algorithm is also less prone to overfitting, which is one of the main issues with decision trees. It predicted the outcome with 95% accuracy for our dataset.

3.5.3 Gradient Boosting

A greedy method that may quickly overcome training data is gradient enhancement. Gradient Boosting may be utilized as a continuous prediction model as well as a categorical prediction model [26]. The model accurately predicts the classes with an accuracy of 95%.

3.5.4 K-Nearest Neighbor (KNN)

The K-nearest neighbor approach is a popular classification algorithm. It's also utilized to solve regression difficulties. It operates by computing the distance between one or more independent variables and the dependent variables (our predicted result) (our features). This method forms a group by using comparable data points, which signifies which data point is closest to the predicted conclusion [29]. It determines how much data it takes to build a group based on the value of k (neighbors' numbers). In this instance, k has a value of 3. This method does not anticipate the outcome; instead, it memorizes the formed groups and compares the test data to those categories to determine an outcome. As a result, displaying the desired conclusion takes time. As a result, this technique is also known as a non-parametric and sluggish algorithm. However, for our dataset, it works admirably, with an accuracy of 95%.

3.5.5 Ada Boost

AdaBoost (Adaptive Boosting) is a boosting technique that is used to combine several weak classifiers into a strong classifier. The output of the weak classifiers is combined into a weighted sum to produce the final output. AdaBoost iteratively creates weak classifiers which are then combined to form a strong classifier. Each classifier is assigned a weight which is used to update the weights of the observation. AdaBoost is best used for binary classification problems, however, there are multi-class versions [30]. It is one of the most popular boosting techniques and is used for many applications including face recognition, computer vision, and classifying people using genetic markers. The accuracy of the Ada Boost Classifier in our datasets is 96.25% which is the maximum among all.

3.5.6 Random Forest

Random forests are a machine-learning approach for solving regression and classification issues. These methods divide the dataset into numerous pieces and generate several decision trees [28]. It makes a judgment or predicts the output based on the conclusion of the decision trees with the highest chance of occurrences. It predicted the outcome with 95% accuracy for our dataset.

3.5.7 Decision Tree

Both regression and classification issues are solved using decision trees. Decision Tree also works with categorical and continuous I/O (input/output) variables. If the goal of this method is to create a tree, this is the technique to take [27]. Decision Tree is a tree in which each internal node corresponds to a characteristic. value and the leaf node indicate a decision. For classification problems, the Decision Tree produces high-accuracy output. It works well in our dataset and has a maximum accuracy of 96.25% which is the same as the Ada Boost Classifier.

3.5.8 Stochastic Gradient Boosting

Friedman suggested a small change to the algorithm shortly after gradient boosting was introduced. He was inspired by Breiman's bootstrap aggregation "bagging" technique. He specifically suggested that a base learner be fitted on a subsample of the training set chosen at random without replacement at each iteration of the algorithm. Friedman found that this change significantly increased the accuracy of gradient boosting. The subsample size is some fixed percentage f of the training set's size. The algorithm is deterministic and identical to the one previously described for f = 1. Smaller values serve as a type of regularization by introducing randomness into the process and assisting in preventing overfitting. Because regression trees must be fitted to fewer datasets with each iteration, the process also becomes faster. Friedman obtained that 0.5Subsampling also enables one to establish an out-of-bag error of the prediction performance increase by evaluating predictions on those observations which were not used in the construction of the subsequent base learner. Out-of-bag estimates assist in avoiding the requirement for a separate validation dataset, but frequently overestimate real performance improvement and the ideal number of iterations [31]. Stochastic Gradient Boosting works in our dataset with accuracy of 95%.

3.6 Implementation Requirements

To perform the complete job, we need a high-end PC with high GPU, processor and RAM. As we mount our task on Google Colab, it will provide some extra RAM and Space. So, an average PC can perform our task with the help of Google Colab. All required tools are given bellow:

3.6.1 Hardware & Accessories

- Intel Core i3 8th gen or higher
- 8 GB (+4GB by Colab) RAM or Higher
- 512 GB HDD
- High-Speed Internet connection

3.6.2 Software, Language & Tools

- Windows 11
- Python 3.9
- Google Colab
- Browser (Edge/Chrome)

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Experimental Setup

The proposed methodology was used as the reference to set up the experiment. Data collection, data preprocessing, feature selection, Machine Learning models for classification, all steps were completed.

For this task, we have used a personal computer that is built with AMD Ryzen 5 2400G processor, 8 GB RAM, 1 TB storage, 1 GB integrated GPU and Windows 11 Pro 64-bit operating system.

4.2 Experimental Results & Analysis

We have employed different methods in this research and analyzed them in great depth. Now in this section, we will discuss about the outcomes of various Machine Learning Algorithms.

Performance of any algorithm measured by some parameters. Accuracy is one of them. Accuracy is defined by the proportion of correct predictions divided by total prediction.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{Correct\ Prediction}}{\textbf{Total Prediction}}\ \boldsymbol{X}\ \mathbf{100}\%$$

Different algorithm uses different methods. As a result, we can see differences between one's accuracy to another. Different algorithms deliver different accuracy on different datasets. Algorithms used on our dataset and their accuracy are given below:

Table 4.1: Accuracy of Algorithms

Model Name	Accuracy
Decision Tree	96.25%
Ada Boost	96.25%

KNN	95%
Random Forest	95%
Gradient Boosting	95%
Stochastic Gradient Boosting	95%
XGBoost	95%
Extra Tree	95%

The proportion of true positives to the total number of genuine positive pulse false positives is the definition of precision. the share of actual results among all favorable outcomes that the model accurately predicted. It is typically employed to retrieve data.

$$\boldsymbol{Precision} = \frac{\textbf{True Positive}}{\textbf{True Positive} + \textbf{True Negative}}\ \boldsymbol{X}\ \mathbf{100}\%$$

Recall is determined by comparing the proportion of true positives to true positives to false negatives. It serves as a gauge for the model's thoroughness.

$$\boldsymbol{Recall} = \frac{\textbf{True Positive}}{\textbf{True Positive} + \textbf{False Negative}}\ \boldsymbol{X}\ \mathbf{100}\%$$

F-score, also known as the F1-score, is a measure of a model's accuracy that combines precision and recall. It is calculated as the harmonic mean of precision and recall, where precision is the proportion of true positives from all positive predictions, and recall is the proportion of true positives from all actual positives. The F-score ranges from 0 to 1, with a higher score indicating better performance [32].

$$\boldsymbol{F1\ Score} = \frac{\textbf{2 X Precision X Recall}}{\textbf{Precision} + \textbf{Recall}}\ \boldsymbol{X}\ \mathbf{100}\%$$

Table 4.2: Performance measure of proposed model

Model Name	Classification	Precision (%)	Recall (%)	F1 Score (%)
Decision Tree	True	96	98	97
	False	96	93	95
Ada Boost	True	95	100	97
	False	100	89	94
KNN	True	96	96	96
	False	93	93	93
Random Forest	True	93	100	96
	False	100	86	92
Gradient Boosting	True	93	100	96
	False	100	86	92
Stochastic Gradient Boosting	True	93	100	96
	False	100	86	92
XGBoost	True	93	100	96
	False	100	86	92
Extra Tree	True	94	98	96
	False	96	89	93

4.2.3 Macro and Weighted Average

When calculating loss, macro approximation assigns identical values to every projected result; however, if data are inequitably allocated and users want to give a few predictions weight age, they employ a "weighted" average.

Table 4.3: Macro Average

Model Name	Precision (%)	Recall (%)	F1 Score (%)
Decision Tree	97	95	96
Ada Boost	97	95	96
KNN	95	95	95
Random Forest	96	93	94
Gradient Boosting	96	93	94
Stochastic Gradient Boosting	96	93	94
XGBoost	96	93	94
Extra Tree	95	94	94

Table 4.4: Weighted Average

Model Name	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
Decision Tree	96	96	96
Ada Boost	96	96	96
KNN	95	95	95
Random Forest	95	95	95
Gradient Boosting	95	95	95
Stochastic Gradient Boosting	95	95	95
XGBoost	95	95	95
Extra Tree	95	95	95

4.2.4 Confusion Matrix

A confusion matrix is a table that is often used to evaluate the performance of a classification model (or "classifier") where the input is a set of predictions and the true values, and it displays the number of correct and incorrect classifications. It is a way of summarizing the performance of a classifier and understanding where it makes mistakes.

Figure 4.3.1: Confusion Matrix of Ada Boost

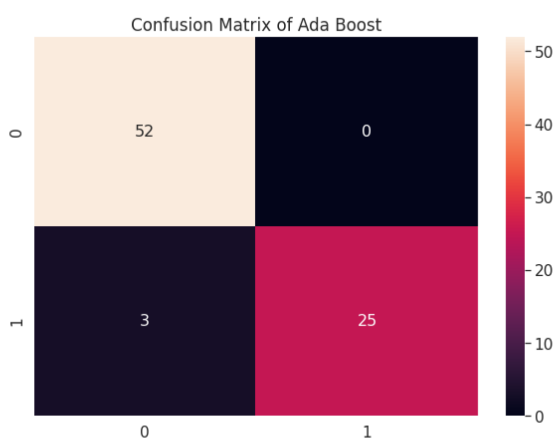
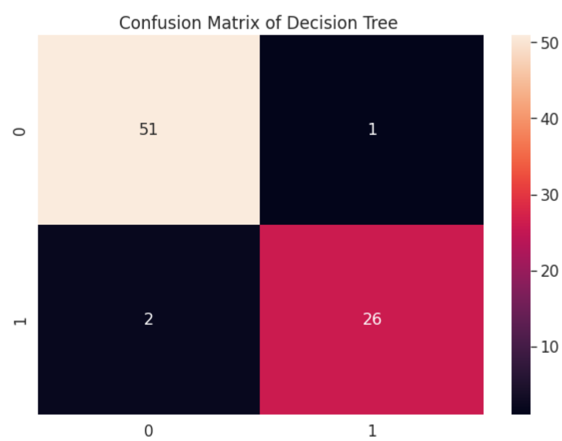


Figure 4.3.2: Confusion Matrix of Decision Tree



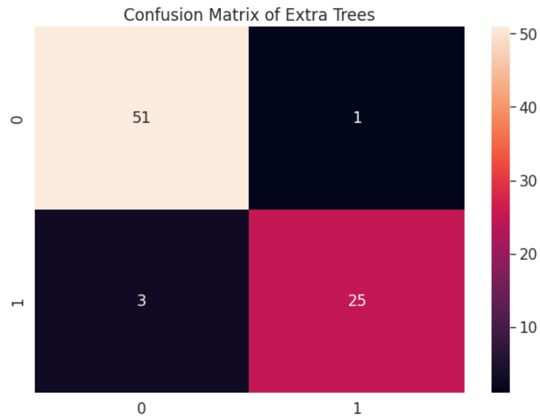


Figure 4.3.3: Confusion Matrix of Extra Trees

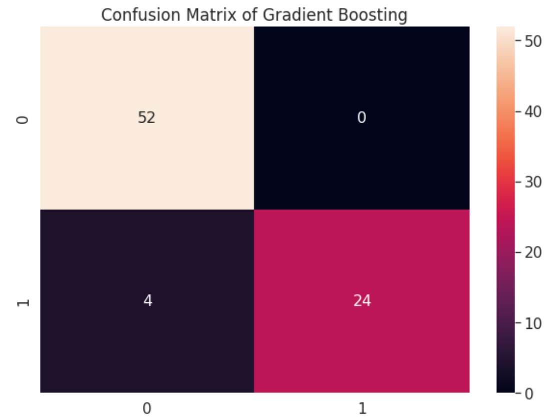


Figure 4.3.4: Confusion Matrix of Gradient Boosting

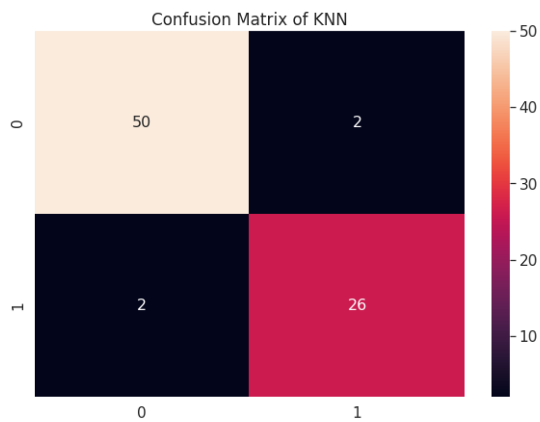


Figure 4.3.5: Confusion Matrix of KNN

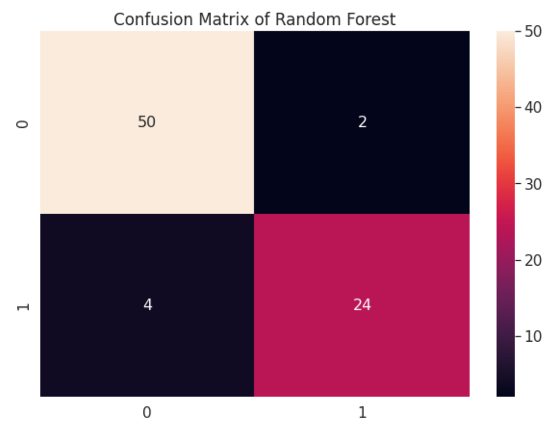


Figure 4.3.6: Confusion Matrix of Random Forest

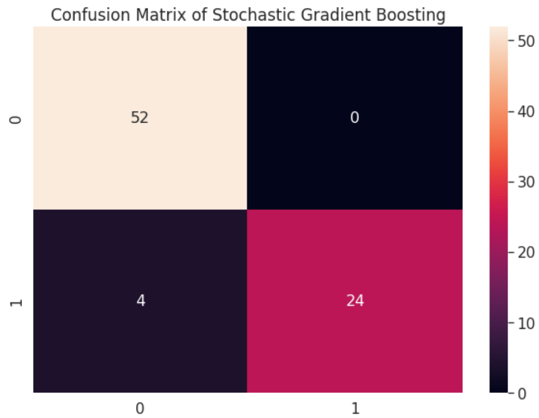


Figure 4.3.7: Confusion Matrix of Stochastic Gradient Boosting

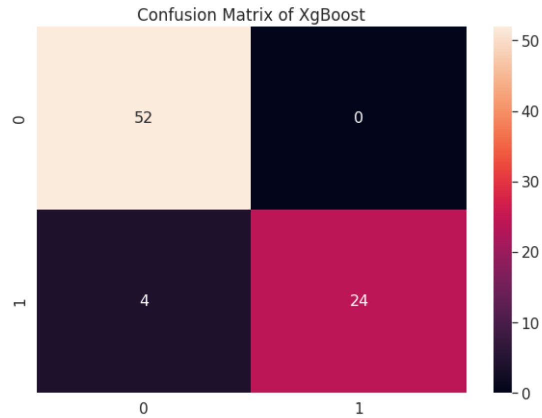


Figure 4.3.8: Confusion Matrix of XgBoost

4.2.5 ROC Curve

The ROC curve is a diagram that displays how well a classification method performed across all categorization levels. Two parameters are shown on this curve, one is TPR and other is FPR. ROC curve is generated by plotting True Positive Rate (TPR) against False Positive Rate (FPR) at various threshold setting. Below we show graphical views of all algorithms ROC curve.

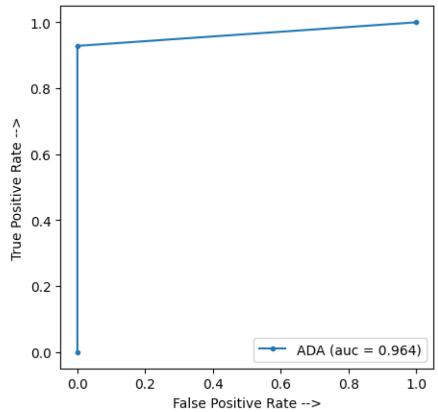


Figure 4.3.1: ROC Curve of Ada Boost

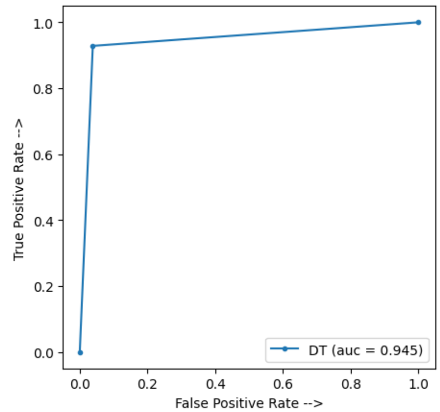


Figure 4.3.2 ROC Curve of Decision Tree

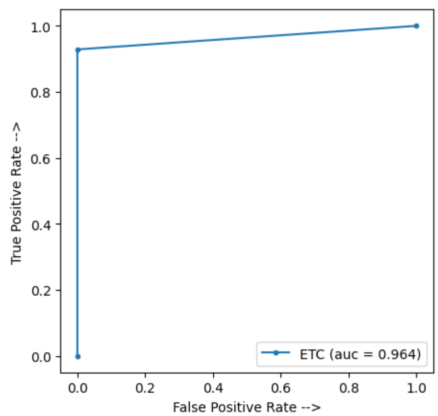


Figure 4.3.3: ROC Curve of Extra Trees

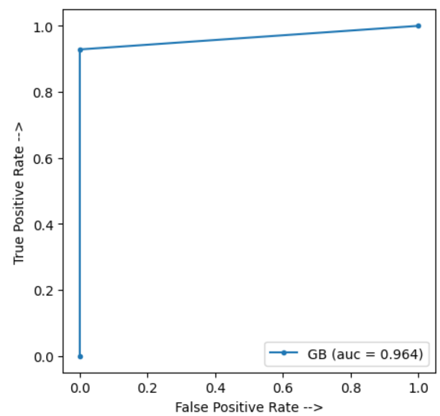


Figure 4.3.4: ROC Curve of Gradient Boosting

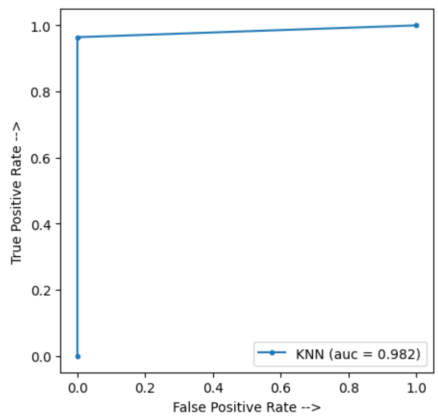


Figure 4.3.5: ROC Curve of KNN

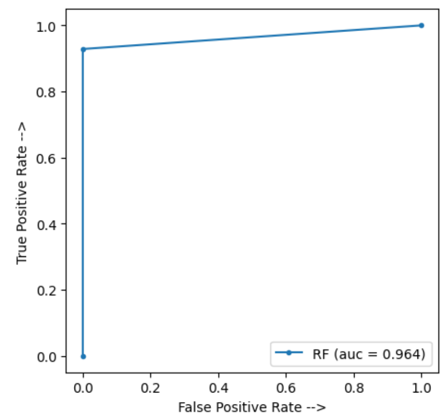


Figure 4.3.6: ROC Curve of Random Forest

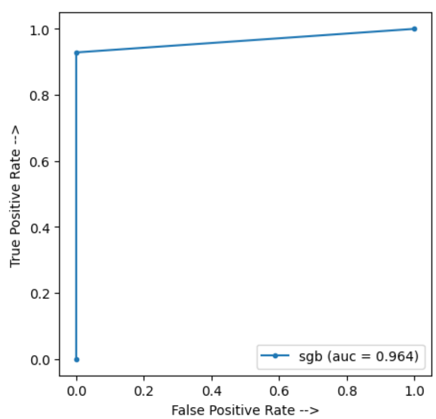


Figure 4.3.7: ROC Curve of Stochastic Gradient Boosting

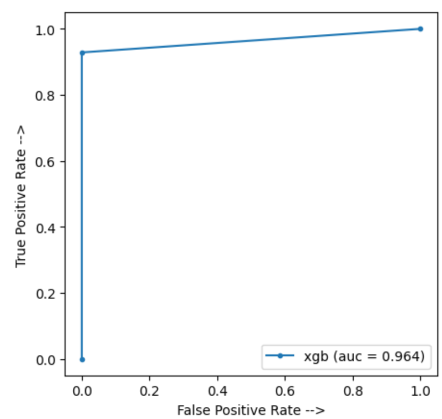


Figure 4.3.8: ROC Curve of XgBoost

4.3 Discussion

In this work, we have used several machine-learning classifiers to predict Chronic Kidney Disease. By preprocessing technique, we have cleaned our data so that the model can work more efficiently and gives us a better result. We have visualsized Confusion matrix, heatmap, and ROC curve for better understanding. We have measured accuracy, precision, recall, f1 score, and also macro and weighted average scores of all the models. After all, we have tested some CKD and Non-CKD data for enquiring whether the algorithms are working accurately. After all, it can be said that the result is at a satisfactory level.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

The potential impact of Chronic Kidney Disease (CKD) Management System on society is immense. Firstly, it can increase access to care and help reduce the burden of this disease. With the early detection and proper treatments, the quality of life for those affected by CKD would improve significantly. This could result in reducing the mortality rate for those suffering from CKD, potentially saving many lives.

Additionally, prevention and early diagnosis of CKD could potentially save billions of dollars in healthcare costs associated with treating this condition. Early diagnosis of CKD could reduce or eliminate the need for dialysis and associated costs, which would have an impact not only on the individual and their families but also on society in general.

Finally, the CKD Management System has the potential to reach out to vulnerable communities that lack access to healthcare services. By identifying those at risk, and reaching out to them with care, education, and awareness, we can address the detrimental impact of CKD in communities that are disproportionally affected by it.

5.2 Impact on Environment

The impact of a Chronic Kidney Disease Management System on the environment is likely to be minimal. If implemented correctly, the system should help to reduce water and energy consumption, reduce the number of medications and other treatments necessary, and ultimately improve the health of those with the condition. In addition, the system should help to reduce the amount of waste generated by medical facilities and labs associated with chronic kidney disease identification and treatment. These reductions could result in lower levels of water and air pollution which in turn would help to preserve the environment.

5.3 Ethical Aspects

The ethical aspects of a chronic kidney disease management system have to do with ensuring the privacy and confidentiality of patient data, ensuring that the decisions made by the system are based on reliable and valid data, providing resources so that patients can make informed decisions, protecting the health and safety of patients, and respecting the autonomy of patients. Furthermore, the ethical considerations associated with such a system should include the equitable distribution of resources, which is the equitable and responsible use of available resources for the benefit of all patients. Additionally, ethical considerations should be taken into account when designing and implementing the system, for example, focusing on reducing medical errors and protecting patients from potential harm. Furthermore, any potential conflicts between the interests of healthcare providers and those of the patient should be taken into account in order to protect the patient's best interests. Finally, ethical considerations should also be taken into account when selecting and evaluating treatment options, to ensure that they are consistent with established ethical principles.

5.4 Sustainability Plan

- Develop a freely available open-source environment for the chronic kidney disease management system to increase its usability and promote teamwork.
- The chronic kidney disease management system should be integrated with other existing healthcare systems to ensure seamless communication and data sharing between healthcare providers. This will ensure that all healthcare providers have access to the same information and can collaborate more effectively on the patient’s care.
- Utilize an automated testing platform to validate accuracy and reliability.
- Design data collection protocols to ensure the system is not collecting sensitive data without user consent.
- Ensure the system is secure and resilient to attack.
- Incorporate feedback loops to ensure the system is always learning and improving.
- Develop a comprehensive user interface to ensure ease of use for all users.
- Establish a maintenance plan to ensure the system is regularly updated.
- Regularly audit the system for any potential vulnerabilities or errors.

Chapter 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the study

The study aims to develop an automated system to predict CKD, identify the stage of CKD, find out the probability of CKD, predict CKD progression and manage Chronic Kidney Disease (CKD). The study proposes a machine learning-based system using supervised learning algorithms to predict the onset of CKD accurately. The system is designed to use a combination of patient data, such as age, gender, ethnicity, family history, lifestyle, and laboratory test results, to predict and find the probability of the onset of CKD. In addition, the system is designed to provide personalized recommendations for lifestyle modifications and all the relevant information related to CKD such as symptoms, risk factors, causes, nutrition, diagnosis and testing, and treatments and to create awareness about CKD. The study also proposes using interactive visualizations to help patients understand their condition and make informed decisions about their health. The study also outlines a procedure for validating the accuracy of the system and provides recommendations for further development.

6.2 Conclusion

This study of CKD management system using machine learning has demonstrated that it is possible to create an automated system for predicting, diagnosing, and managing Chronic Kidney Disease (CKD). The system is able to accurately detect stages of CKD, find the probability of CKD from a variety of laboratory tests, as well as predict the progression of the disease. Furthermore, the system is capable of providing personalized management plans to help slow the progression of CKD, including lifestyle changes and other treatments.

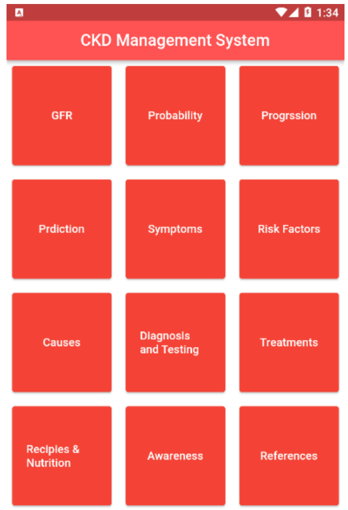


Figure 6.1: Homepage

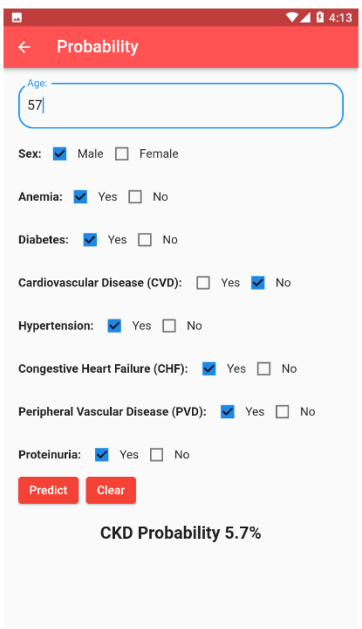


Figure 6.2: CKD Prediction(positive)

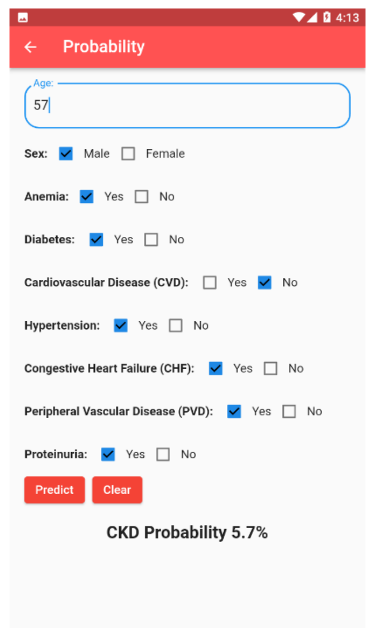


Figure 6.3: CKD Prediction(Negative)

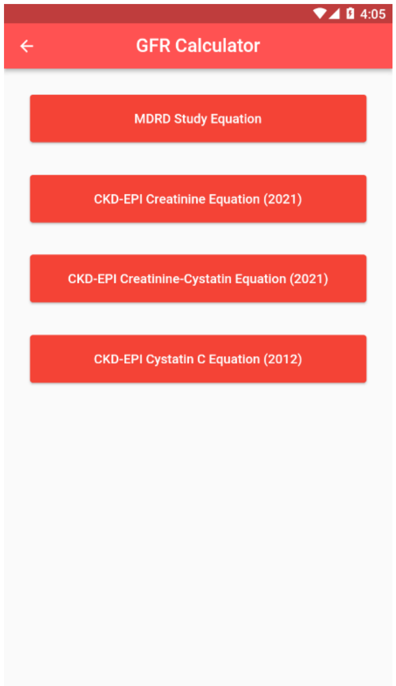


Figure 6.4: GFR Calculator Section

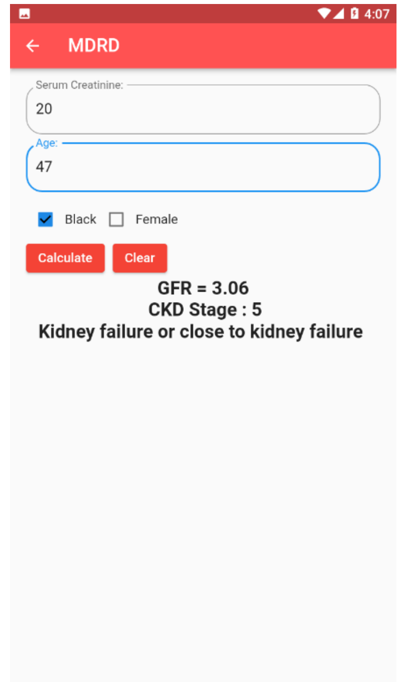


Figure 6.5: Finding GFR and CKD Stage Using MDRD Equation

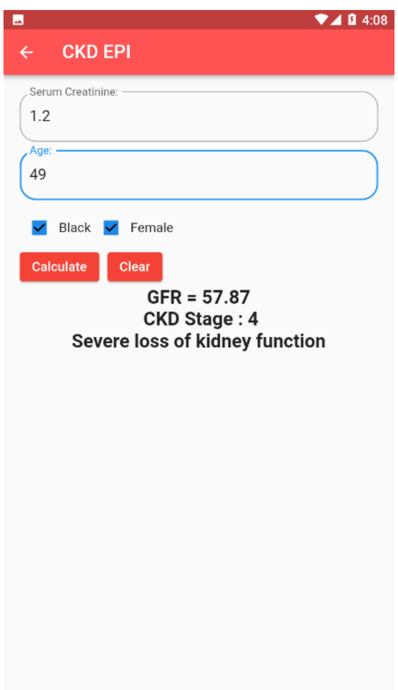


Figure 6.6: Finding GFR and CKD Stage Using CKD-EPI Equation

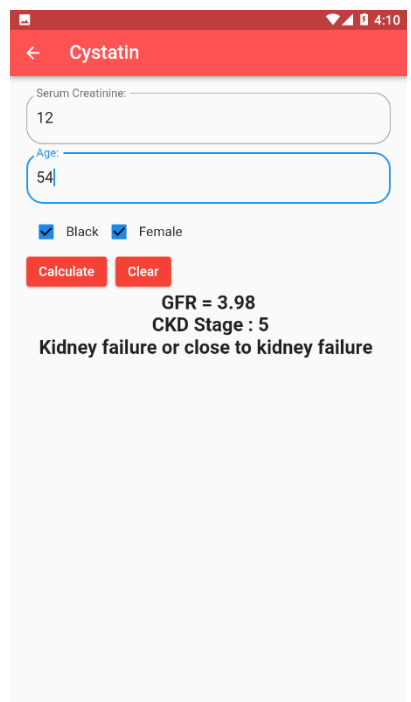


Figure 6.7: Find GFR and CKD Stage Using Cystatin Equation

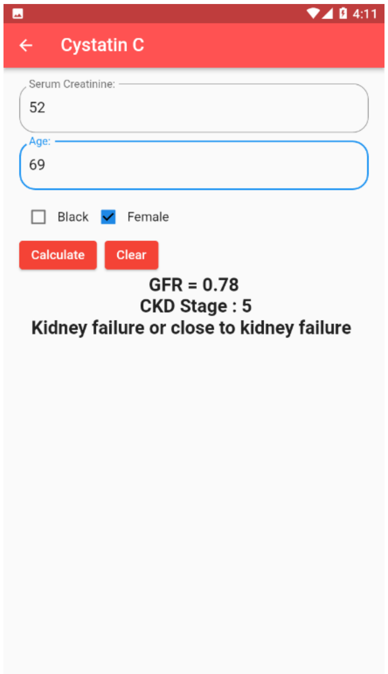


Figure 6.8: Find GFR and CKD Stage Using Cystatin Equation

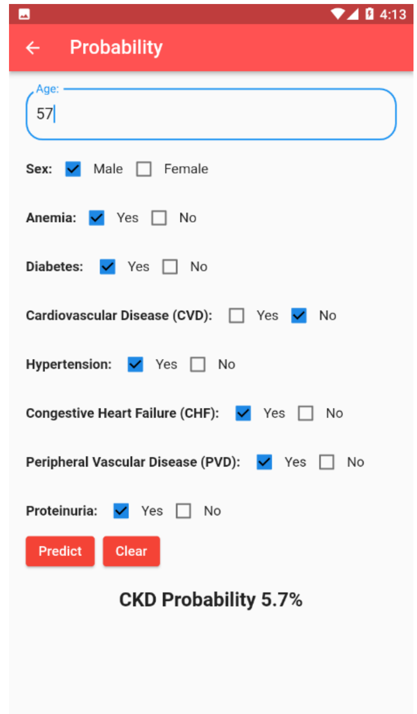


Figure 6.10: Probability of CKD

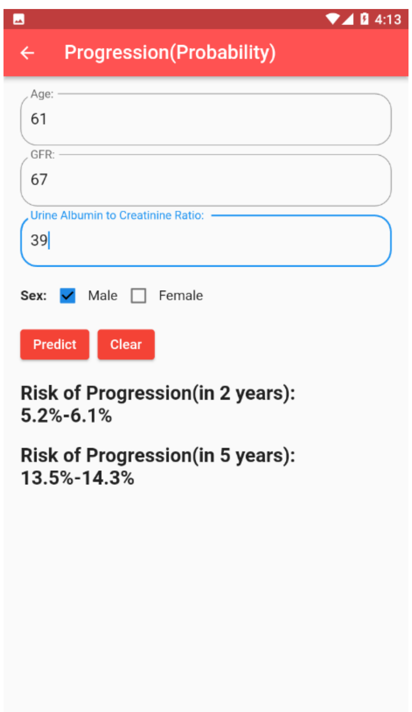


Figure 6.11: Progression of CKD

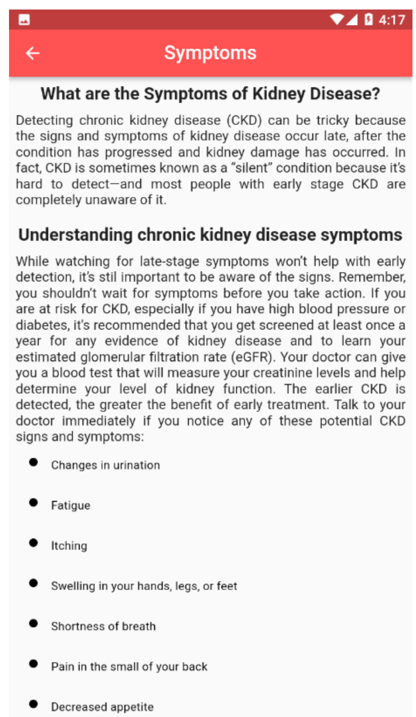


Figure 6.12: Symptoms of CKD (Information-based feature)

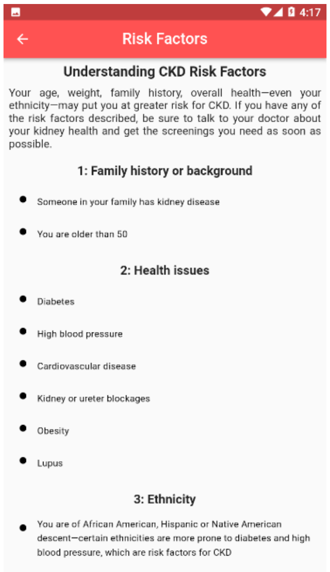


Figure 6.13: Risk Factors of CKD(Information-based feature)

Figure 6.14: Causes of CKD(Information-based feature)

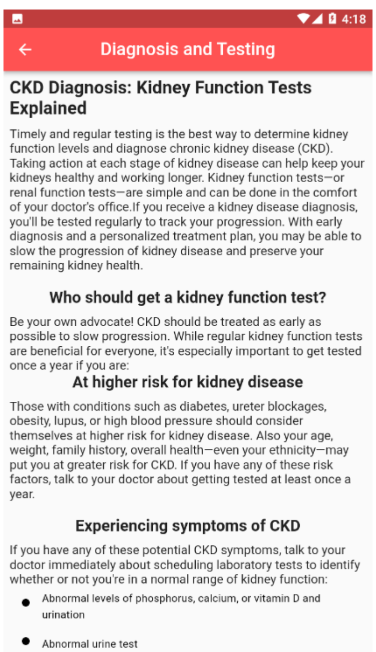


Figure 6.15: Diagnosis of CKD(Information-based feature)

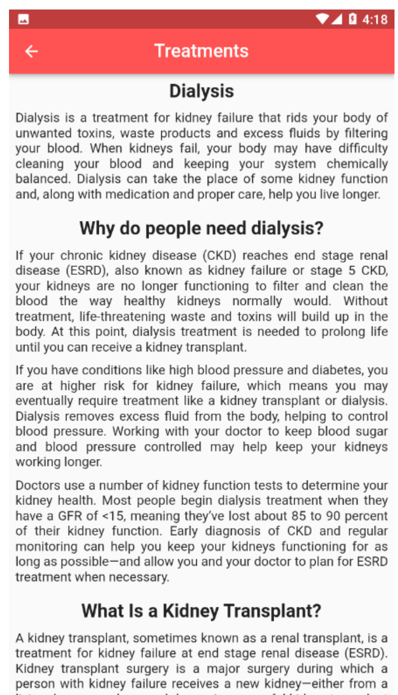


Figure 6.16: Treatments of CKD(Information-based feature)

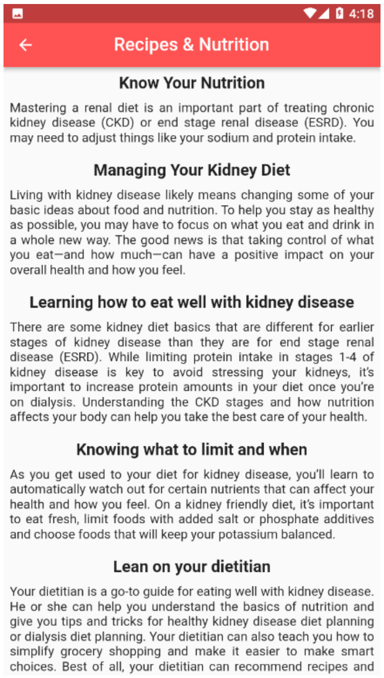


Figure 6.17: Recipies&Nutrition of CKD(Information-based feature)

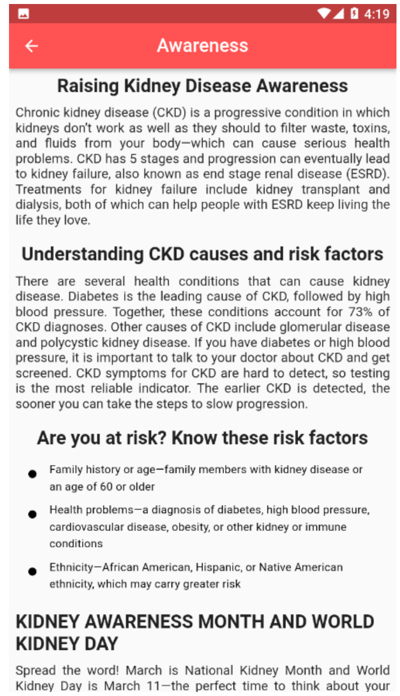


Figure 6.18: Awareness about CKD(Information-based feature)

With the successful development of this system, it is likely that CKD can be managed more effectively and efficiently in the future. However, it is essential to perform further research and validate the outcomes of this study in a larger and more diverse population.

6.3 Implication for Further Study

Further study is needed to develop a comprehensive and effective CKD management system that incorporates the latest evidence-based guidelines and technologies. Such a system should be designed to provide personalized care to patients, facilitate communication between healthcare providers, and enable patient self-management. Additionally, the system should be tailored to the needs of different clinical settings and patient populations. The system should also facilitate the integration of data from different sources, such as electronic health records, laboratory tests, and medical imaging, to provide a comprehensive view of patient health. Finally, further research should be conducted to evaluate the effectiveness of the proposed CKD management system in improving patient outcomes.

References:

[1] A. S. Levey and J. Coresh, "Chronic kidney disease," *The Lancet*, vol. 379, no. 9811, pp. 165–180, Jan. 2012, doi: 10.1016/s0140-6736(11)60178-5.

[2] J. Zhong, H.-C. Yang, and A. B. Fogo, "A perspective on chronic kidney disease progression," *American Journal of Physiology-Renal Physiology*, vol. 312, no. 3, pp. F375–F384, Mar. 2017, doi: 10.1152/ajprenal.00266.2016.

[3] O. J. Wouters, D. J. O'Donoghue, J. Ritchie, P. G. Kanavos, and A. S. Narva, "Early chronic kidney disease: diagnosis, management and models of care," *Nature Reviews Nephrology*, vol. 11, no. 8, pp. 491–502, Jun. 2015, doi: 10.1038/nrneph.2015.85.

[4] A. Ogunleye and Q.-G. Wang, "Enhanced XGBoost-Based Automatic Diagnosis System for Chronic Kidney Disease," *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, Jun. 2018, doi: 10.1109/icca.2018.8444167.

[5] P. Arulanthu and E. Perumal, "An intelligent IoT with cloud centric medical decision support system for chronic kidney disease prediction," *International Journal of Imaging Systems and Technology*, vol. 30, no. 3, pp. 815–827, Mar. 2020, doi: 10.1002/ima.22424.

[6] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021, doi: 10.1109/access.2021.3053763.

[7] K. R. Anantha Padmanaban and G. Parthiban, "Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease," *Indian Journal of Science and Technology*, vol. 9, no. 29, Aug. 2016, doi: 10.17485/ijst/2016/v9i29/93880.

[8] P. Ghosh, F. M. J. M. Shamrat, S. Shultana, S. Afrin, A. A. Anjum, and A. A. Khan, "Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm," *IEEE Xplore*, Nov. 01, 2020. https://ieeexplore.ieee.org/document/9376787 (accessed Mar. 23, 2021).

[9] W. H. S. D. Gunarathne, K. D. M. Perera, and K. A. D. C. P. Kahandawaarachchi, "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD)," *IEEE Xplore*, Oct. 01, 2017. https://ieeexplore.ieee.org/document/8251305 (accessed Oct. 09, 2020).

[10] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," *IEEE Access*, vol. 8, pp. 20991–21002, 2020, doi: 10.1109/ACCESS.2019.2963053.

[11] F. Kronenberg, "Emerging risk factors and markers of chronic kidney disease progression," *Nature Reviews Nephrology*, vol. 5, no. 12, pp. 677–689, Dec. 2009, doi: 10.1038/nrneph.2009.173.

[12] A. Ogunleye and Q.-G. Wang, "XGBoost Model for Chronic Kidney Disease Diagnosis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 6, pp. 2131–2140, Nov. 2020, doi: 10.1109/tcbb.2019.2911071.

[13] "Chronic Kidney Disease Basics | Chronic Kidney Disease Initiative | CDC," *www.cdc.gov*, Mar. 09, 2021. https://www.cdc.gov/kidneydisease/basics.html#:~:text=CKD%20is%20a%20condition%20in

[14] J.-C. Lv and L.-X. Zhang, "Prevalence and Disease Burden of Chronic Kidney Disease," *Advances in Experimental Medicine and Biology*, vol. 1165, pp. 3–15, 2019, doi: 10.1007/978-981-13-8871-2_1

[15] National Kidney Foundation, "About Chronic Kidney Disease," *National Kidney Foundation*, 2017. https://www.kidney.org/atoz/content/about-chronic-kidney-disease

[16] S. W. Ong *et al.*, "Integrating a Smartphone–Based Self–Management System into Usual Care of Advanced CKD," *Clinical Journal of the American Society of Nephrology*, vol. 11, no. 6, pp. 1054–1062, May 2016, doi: 10.2215/cjn.10681015.

[17] T. K. Chen, D. H. Knicely, and M. E. Grams, "Chronic Kidney Disease Diagnosis and Management," *JAMA*, vol. 322, no. 13, p. 1294, Oct. 2019, doi: 10.1001/jama.2019.14745.

[18] A. Y. Al-Hyari, A. M. Al-Taee, and M. A. Al-Taee, "Clinical decision support system for diagnosis and management of Chronic Renal Failure," 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), pp. 1–6, 2013.

[19] D. Gupta, S. Khare, and A. Aggarwal, "A method to predict diagnostic codes for chronic diseases using machine learning techniques," 2016 International Conference on Computing, Communication and Automation (ICCCA), pp. 281–287, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7813730/

[20] A. Salekin and J. Stankovic, “Detection of chronic kidney disease and selecting important predictive attributes,” IEEE International Conference on Healthcare Informatics, pp. 262–270, 2016.

[21] M. P. N. M. Wickramasinghe, D. M. Perera, and K. A. D. C. P. Kahandawaarachchi, “Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms,” *IEEE Xplore*, Dec. 01, 2017. https://ieeexplore.ieee.org/document/8268202 (accessed Nov. 23, 2021).

[22] H. A. Wibawa, I. Malik, and N. Bahtiar, “Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease,” *2018 2nd International Conference on Informatics and Computational Sciences (ICICoS)*, Oct. 2018, doi: 10.1109/icicos.2018.8621762.

[23] “Chronic Kidney Disease Prediction using Machine Learning Models,” *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 6364–6367, Oct. 2019, doi: 10.35940/ijeat.a2213.109119.

[24] Parul Sinha and Poonam Sinha, “Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM,” *International Journal of Engineering Research and*, vol. V4, no. 12, Dec. 2015, doi: 10.17577/ijertv4is120622.

[25] I. L. Cherif and A. Kortebi, "On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification," 2019 Wireless Days (WD), 2019, pp. 1-6, doi: 10.1109/WD.2019.8734193.

[26] D. Upadhyay, J. Manero, M. Zaman and S. Sampalli, "Gradient Boosting Feature Selection With Machine Learning Classifiers for Intrusion Detection on Power Grids," in IEEE Transactions on © Daffodil International University 33 Network and Service Management, vol. 18, no. 1, pp. 1104-1116, March 2021, doi: 10.1109/TNSM.2020.3032618.

[27] Peng Ye, "The decision tree classification and its application research in personnel management," Proceedings of 2011 International Conference on Electronics and Optoelectronics, 2011, pp. V1-372- V1375, doi: 10.1109/ICEOE.2011.6013123.

[28] Y. Xiao, W. Huang and J. Wang, "A Random Forest Classification Algorithm Based on Dichotomy Rule Fusion," 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC), 2020, pp. 182-185, doi: 10.1109/ICEIEC49280.2020.9152236.

[29] K. Taunk, S. De, S. Verma and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.

[30] R. E. Schapire, “Explaining AdaBoost,” *Empirical Inference*, pp. 37–52, 2013, doi: 10.1007/978-3-642-41136-6_5.

[31] J. H. Friedman, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, Feb. 2002, doi: 10.1016/s0167-9473(01)00065-2.

[32] Koo Ping Shung, “Accuracy, Precision, Recall or F1?,” *Towards Data Science*, Mar. 15, 2018. https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

ORIGINALITY REPORT

29%	24%	22%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	Afrin Jaman Bonny, Puja Bhowmik, Md. Shihab Mahmud, Abdus Sattar. "Detecting Fake News in Benchmark English News Dataset Using Machine Learning Classifiers", 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2022 Publication	2%
3	Adeola Azeez Ogunleye, Wang Qing-Guo. "XGBoost Model for Chronic Kidney Disease Diagnosis", IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019 Publication	1%
4	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1%
5	Submitted to University of Hertfordshire Student Paper	1%