EARLY DETECTION OF CHRONIC KIDNEY DISEASE (CKD) USING OPTIMIZED MACHINE LEARNING MODELS

BY

Md. Shahoriar Rahaman Shohan
ID: 181-15-1760

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

Supervised By

Md. Aynul Hasan Nahid
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Mahimul Islam Nadim
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2025

APPROVAL

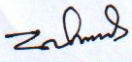
This Project titled "Early detection of chronic kidney disease (CKD) using optimized machine learning models", submitted by Md. Shahoriar Rahaman Shohan to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment 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 13-January-2025.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque — **Chairman**
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Abbas Ali Khan — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Md. Aynul Hasan Nahid — **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

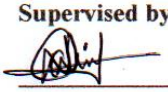


Dr. Md. Zulfiker Mahmud — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

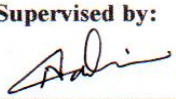
I hereby declare that, this project has been done by me under the supervision of **Md. Aynul Hasan Nahid, Lecturer, Department of CSE,** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Md. Aynul Hasan Nahid
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



Mr. Mahimul Islam Nadim
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Shohan

Md. Shahoriar Rahaman Shohan
ID: 181-15-1760
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project/internship successfully.

I'm really grateful and wish my profound indebtedness to **Md. Aynul Hasan Nahid**, **Lecturer**, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "*Machine Learning, Artificial Intelligence*" 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.

I would like to express my heartiest gratitude to Md. Aynul Hasan Nahid and Head**,** Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Chronic kidney disease is sometimes abbreviated to "CKD." The term "CKD" generally refers to this ailment. The kidneys are affected by this disorder, which is also known as chronic renal disease. The immense progress made in the area of machine learning and artificial intelligence is what has ignited the interest that has been produced as a result of these developments. Thus, any doctor with access to the dialysis report has the capacity to determine when the illness first manifested itself. This approach can also be used to identify the primary etiological component of the illness, which can be deduced from the study's findings. Our dataset was collected from the “Popular Diagnostic Center – Savar branch”, and UCI databases. “Random Forest, Naive Bayes, Decision Tree, K-Nearest Neighbor (KNN), XGBoost, AdaBoost”, and many other complex and adaptable algorithms are required to optimize the performance of this system. XGBoost was chosen as the most accurate algorithm, with an accuracy of 99.1 %, according to the results. The overall performance of this method is excellent for both negative and positive values, as well as for the Macro and Weighted Average variables.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

CKD is commonly held that the natural process of aging itself is one of the most important contributors that can lead to the development of this widespread disease. It has come tolight those persons of color, in particular those with ancestry in South Asia and the Caribbean, have a significantly greater incidence rate of the disorder in comparison to white people. People of color have been found to have a much higher incidence rate of the disorder, despite the fact that the disorder has the potential to affect a large number of different people. Despite the fact that end-stage renal disease (ESRD) is an extremely unlikely outcome. The term ESRD describes the very last stage of this disease. Even though they have CKD, a sizeable percentage of people are nonetheless able to enjoy long and healthy lives. This remains the true in spite of the fact that they have CKD. Researchers from a wide array of countries and regions of the world collaborated in an effort to determine the factors that contributed to the development of this kidney condition.

There are a number of irregularities that have the ability to develop in the kidneys, and each of these problems has the potential to lead to a condition that may pose a threat to the individual's life who is impacted by it. Some of the more severe conditions that can have an effect on the kidneys are “CKD, kidney stones, glomerulonephritis, polycystic kidney disease, and urinary tract infections”. These conditions can all be caused by a variety of underlying medical conditions. When all of these different entities are taken into consideration, hypertension emerges as a prevalent etiological factor that, if left untreated, might result in the development of chronic renal disease. Elevated blood pressure is a potential hazard to renal function due to its propensity to impose an additional burden on the glomerular filtration apparatus, responsible for the elimination of metabolic byproducts. The glomeruli, which are tiny blood vessels within the

kidneys, function to filter and purify the blood. Prolonged exposure to elevated pressure levels exerts adverse effects on the renal vasculature, ultimately leading to a decline in renal function. The renal function will eventually deteriorate to the extent of renal failure. In the eventuality of such a scenario, an individual would be required to undergo dialysis. The process of dialysis entails the elimination of surplus fluid and waste products from the bloodstream. The utilization of dialysis has been found to be beneficial in the management of renal disease; however, it does not possess the ability to provide a complete cure. Depending on the patient's medical condition, a renal transplantation may be a viable option. Diabetes is frequently identified as a leading cause of CKD. Elevated levels of glucose in the bloodstream are indicative of a range of medical conditions, including but not limited to diabetes mellitus. Prolonged hyperglycemia can lead to detrimental effects on the renal vasculature. This suggests that the renal system is incapable of executing its typical task of purifying the bloodstream. It is possible that the accumulation of toxins in the body is one of the factors that contribute to the progression toward renal failure. The actual mortality rate of chronic renal disease was provided in their study by researchers hailing from a number of academic subfields and specializations. The death rate of male patients who have been diagnosed with chronic renal sickness is shown to be substantially greater than that of their female counterparts, according to the findings of a number of different pieces of study. The number of deaths in Bangladesh that may be directly attributed to renal illness has reached a shockingly high level. This is because kidney disease is the leading cause of mortality in the country.

According to the most recent statistics that were made available to the public by the World Health Organization (WHO), the number of people who lost their lives as a consequence of the aforementioned cause in the year 2018 was 16,948, which equals 2.18 percent of the total number of fatalities that occurred all over the world. According to a credible source, Bangladesh has been rated 94th in the world in terms of its age-adjusted Death Rate, which now stands at 14.83 per 100,000 individuals. As a consequence of this, Bangladesh is currently positioned in the bottom half of the rankings for all countries in the globe. At this moment, the pandemic thatwas caused by COVID-19 is having an effect across the entirety of Bangladesh. The likelihood ofdeveloping a severe medical condition in individuals with CKD is heightened by exposure to COVID-19, thereby exacerbating the difficulties associated with obtaining dialysis and other medical care services [2].

Machine learning algorithms can be broken down into a wide variety of subcategories, each of which has its own specific application in the modern world. Machine learning can be classified into a number of distinct subfields, the most well-known of which are supervised machine learning, unsupervised machine learning, semi-supervised machine learning, and reinforcement machine learning. The current study makes use of a wide range of supervised machine learning algorithms in order to attempt to create accurate forecasts regarding CKD. Some examples of these algorithms include the "K-Nearest Neighbor algorithm, the Decision Tree algorithm, the Random Forest algorithm, the Perceptron algorithm, the AdaBoost program, the Gaussian Naive Bayes algorithm, and the XGBoost algorithm". When all that has been said up to this point is taken into consideration, it is feasible to come to the conclusion that the algorithms that are in question are of the utmost significance. By utilizing this method, individuals are able todetermine how likely it is that they may, at some point in the future, have a major kidney condition.It is highly likely that prove to be of great assistance not only to departments of pathology but also to patients who are looking for information regarding the possible implications of chronic renal illness. This is because it is highlylikely that patients will look for information regarding the potential impacts of chronic renal disease.

The main objective of my investigation is to apply models that has been trained on a sufficient dataset, determine which algorithm is the most effective one, and then forecast CKD at any stage. It is projected that in a not-too-distant future, the current epoch will evolve into a time period that is more technologically advanced and evolved as a result of technical advancements. The goal of this study is to raise people's understanding of the causes that contributeto renal illness and provide them the opportunity to take the required actions to limit the effects ofthe condition by providing them with the knowledge and information gained from this research. This will be performed through the use of prospective web-based reporting, and the overall purpose of the project is to meet both of these goals.

1.2 Motivation

In Bangladesh, a sizeable portion of the population is affected by a broad spectrum of persistent ailments. These conditions include diabetes and hypertension, amongst a variety of others. This category accounts for a sizeable proportion of the total population in the nation. It is anticipated that the current trend of diagnosing an increasing number of individuals with chronic diseases will continue in today's modern culture, where an increasing number of people are being diagnosed with these illnesses. It is extremely important for persons who are afflicted with chronic illnesses to continue making routine visits to medical institutions. Due to the fact that Bangladesh is located in a region that is blessed with an abundance of water resources as a result of its geographical qualities, the country of Bangladesh possesses a large quantity of these resources. This is because the countryis surrounded on all sides by water because it is surrounded on all sides by rivers that flow into theBay of Bengal. As a result, the country is surrounded by water on all sides.

It is believed that thosewho possess these specific characteristics have an increased risk of getting waterborne infections. This is owing to the fact that it is believed that these characteristics make peoplemore prone to getting diseases that are transmitted by water. It is of the utmost importance to devise a strategy that makes use of ML in order to acquire an understanding of the progression of this disease and to uncover methods by which its effects might be mitigated.

1.3 Rationale of the Study

Early diagnosis of CKD can facilitate prompt treatment, leading to a swift cure. In practical settings, individuals are required to conduct numerous laboratory teststo confirm the presence of CKD, a process that is known to be quite time- intensive. For the purpose of accurately predicting the presence of CKD throughout all of its phases, a predictive model has been designed and trained with the assistanceof a suitable dataset. This was accomplished. This was done with the objective of providing an accurate forecast regarding the course of the condition. The ultimate goal of this study is to makepeople's lives easier by improving their organizational abilities and their capacity to better manage their time. This will be accomplished through the course of this study. Patients now have the ability to input data and carry out a self-assessment without having to leave the convenience of their own homes. Patients who have been given a diagnosis of CKD, in addition to their treating physicians, may discover that this intervention is beneficial to them.

1.4 Research Questions

When it comes to this inquiry, the researchers might be asked a wide range of questions, and the specifics of those questions will depend on what exactly is being looked at. A number of questions, each of which was taken from a different source, have been prepared in order to collect the findings of this investigation into a format that is simpler to grasp and work with. This was done in order to make it possible to do so in a manner that is both efficient and effective.

A) Why did you choose to focus on predicting CKD?

CKD affects an alarmingly high percentage of people all over the world. The passage of time will not result in an improvement in the state of affairs; quite the contrary, in fact; things are only going to become worse. When it reaches the final stage, which is the completion of the process, it has already progressed through a succession of stages, with full renal failure serving as the culmination of the process. Because of the nature of the situation, the only rational conclusion that can be drawn is that an individual will eventually expire. If CKD is detected in its early stages, it is possible to manage it successfullywith the appropriate medicine and by closely adhering to a variety of various criteria. However, in order to accomplish this, the illness must initially be diagnosed at an early stage. As a direct consequence of this, chronic renal illness was settled upon as the primary focus of the investigation.

B) Does the use of a machine learning technique justify the decision made, and is it widely considered reliable?

Machine learning is a method that is frequently utilized across many purposes of making predictions, and the term "machine learning" was coined to describe this phenomenon. This is possible as a result of the ability of ML to learn from previously collected data. When a model has access to a larger volume of data, it has a stronger capacity to take part in self-training and, subsequently, to generate predictions for any event that can possibly be dreamed of. This capacity increases in directproportion to the amount of the datasets. It is feasible to produce more precise forecasts regarding CKD when a method of machine learning is applied to medical datasets. CKD stands for CKD. a time period in which there were enormous developments taking place all across the earth all at the same time.

The era during which the contemporary way of thinking and doing things began to become more widespread. Imagine an event that took place around ten years ago, before there had been substantial developments made in the area of AI or ML. This is an example of a hypothetical circumstance, which may be created by using one's imagination. During this time period in history, the primary foundational conceptions that supported these fields were theoretical creations that were founded on the principles of mathematics. Currently, a significant portion of global technology relies on Artificial Intelligence, accounting for approximately fifty percent of the world's technological advancements. Hence, the implementation of appropriate methodologies and increased accuracy in this domain can enhance its dependability, notwithstanding its current level of reliability.

C) What are the rationales behind utilizing eleven distinct algorithms?

Eleven different algorithms were used in the analysis of the CKD datasets in order to narrow down the search for an appropriate algorithm to use with the data. This was done in order to save time during the process of finding an algorithm to apply. It was decided to do this so that the search could be carried out in the most effective manner possible. It is impossible to determine which one is the best because it is impossible to guess our what algorithm will be the perfect for the dataset. It is not possible to establish which algorithm is the most effective if only one algorithm is employed in the process.

1.5 Expected Outcome

There have been a number of alterations in the direction that this investigation is taking. These shifts are the result of a variety of distinct forces interacting to bring about their occurrence. It is beneficial to explicate the precise result of this investigation. The present study holds promise in terms of mitigating the onset of CKD during its nascent phases, while concurrently elucidating the fundamental etiology of its propagation. The onset of CKD can be ascertained by medical professionals or analysts in relation to the age of individuals.

This will allow for the presentation of the findings. Having someone else do this for you can serve the objective of providing information. It is conceivable to take out this action with the goal of releasing the relevant information, as this is something that is possible. It is not out of the question for a group to send out prior warning to riverine villages and localities on the study that is now being conducted while the inquiry into CKD is being carried out. This would be done as part of the investigation. This endeavor is going to be carried out within the parameters of the study that is now being conducted.

1.6 Report Layout

Chapter 1 gives a comprehensive presentation to the inquire about venture that is presently being carried out. When examining kidney illness and the particulars of the affliction, this is something that needs to be taken into intellect as an imperative figure. This chapter presents the investigate inspiration, the defense for the examination, major inquire about questions, anticipated results, and considerable administration data, counting money related angles of the organization.

Chapter 2 provides a full explanation of the contextual framework under which this research was conducted. This research project is concentrating its efforts on the study of machine learning systems, information classification, and the work that is associated with these topics. This chapter presents a comparative analysis, defines the scope of the issue statement, and identifies the perceived challenges that need to be addressed. Additionally, the chapter outlines the steps that need to be taken. In addition to this, the chapter provides an explanation of the issues that require resolution.

Chapter 3 gives a point-by-point account of the technique utilized and the proposed framework for the inquire about ponder. The algorithmic complexities of each utilized calculation are clarified upon from their scientific establishments to their show state.

Chapter 4 offers a total investigation into the comes about accomplished at each consequent organize of the prepare. The discoveries of the examination are summed up by selecting the perfect procedure that gives the best exactness score and analyzing the viability of the calculation by making utilize of a number of distinctive measurements.

Chapter 5 elucidates the ethical implications that this study has on society, which is an essential component of any research project that strives to have a significant influence and ought to be included in all studies of this kind. a discussion of the ethical repercussions that this research has on the community. The chapter comes to a close with a discussion on the long-term repercussions of the research that was carried out, which was the primary focus of the chapter.

Chapter 6 provides an explanation of the potential future routes that this research project could go, as it is briefly elaborated as an extension of the work that has already been done. As a consequence of the fact that it offers a summary in a condensed form of the findings that are the most pertinent as a result of the research, the current chapter plays an important role in serving as a conclusion.

CHAPTER 2

LITERATURE REVIEW

2.1 Related Works

Here, they suggest using an IoMT setting to implement the HMANN architecture for “CKI detection, and classification”. The aforementioned HMANN model incorporates the MLP architecture, Backpropagation (BP), and Support Vector Machine (SVM) techniques. First, a segmented ultrasound image of the kidney is used to locate the target region. The HMANN approach, which is recommended for kidney segmentation, not only significantly reduces the time needed to create the contour, but also exhibits amazing accuracy. Using facial image processing, the authors of this research create a dual-stack network (DsNet) that can identify healthy people. They made sure the network could be trained in stages. The stack algorithm's first iteration does a respectable job of recognizing key high-level facial characteristics in pictures. It is important to collect the high-level features from the first stack subset and then assess them in the second stack network in order to categorize both diseases together in otherwise healthy persons. When compared to the accepted, non-invasive methods of diagnosis, this is a significant departure [3].

Ali et al.'s study, which investigates the use of automated decision systems. This research introduces a novel method for collaborative feature selection by means of an effective ensemble feature ranking algorithm. The method incorporates the desire for more cost realism. This research is remarkable since it is one of the earliest examples of cost-effective ensemble methods. Non- cutting groups should be ranked if you want to save money and receive reliable outcomes. According to the findings, automated CKD systems can benefit from taking cost into account inside the objective space of solution formulations. In order to detect CKD, researchers have proposed a screening technique [4].

Chebyshev's distance criterion is used to determine the winners. The proposed function selection strategy for data sets relevant to CKD resulted in a reduction of 36% of features, compared to the existing TLBO methodology, which resulted in a reduction of 25%. Compared to the original TLBO methodology described in Balakrishnan et al.'s publication [5], experimental results show that all three strategies improve classification accuracy for the intended feature subset.

Lambert et al. used numerical and nominal variables to improve their ability to predict the course of chronic renal disease. In this investigation, they use the CFS approach to classify personality features as indicative of CKD or not. Classification and prediction of features are key to the research. The CFS approach can be used to discover features in a dataset that is comprised entirely of nominal values, entirely of numeric values, or both. Three distinct techniques to select a ranking function—information gain, gain ratio, and the Relief method—are used in the comparison.

When applied to the classification of nominal, numerical, and combined nominal and numerical data achieved remarkable high rates of accuracy (98.5%, 95.25%, and 98.5%, respectively). The SMO algorithm also showed promising results when applied to renal illness. That's how you define a satisfactory situation. Medical professionals may benefit from using the Chronic Fatigue Syndrome Symptom Measurement Instrument (CFS-SMO) for more precise diagnosis of renal illness [6]. Nusinovici et al. also incorporate and assess information from simpler clinical prognoses in our prospective cohort analysis. Five machine learning models—including the one-hidden-layer-neural-network model, the vector support machine model, the random forest model. According to the aforementioned source [7], the best outcomes were achieved by combining a neural network with logistic regression and gradient boosting.

Segal and colleagues analyzed data consisting of millions claim out of 550,000 patient records. Participants were required to be adults with a diagnosis of Stage 1-4 CKD. They have collected over 240 potential predictors and organized them into 6 feature sets. Using the feature embedding approach of Word2Vec, information about diagnosis, procedures, and drugs is obtained together with associated timestamps. Use the XGBoost version of the gradient boosting technique for statistical testing [8].

Sealfon et al.'s study aims to determine if automated methodologies can be utilized to learn more about renal disease, and more precisely the relationship between genotype and phenotype. The current renal data set is very comprehensive and detailed. Therefore, machine-based teaching methods are improving in their ability to explain kidney biology and answer numerous issues having practical, biological and translational implications [9].

Luo et al. investigate the links between four metals found in the blood and renal function. Researchers interviewed 1,435 U.S. residents to compile data for the 2015-2016 NHANES. Albuminuria is often diagnosed through the use of the - eGFR. Renal failure is known as CKD or acute kidney injury (AKI). They used BKMR to investigate the possibility of effect shifts across correlations and to discover the existence of non-linear relationships between combinations and outcomes. The R programming language's interaction package allowed them to discover associations between continuous measures. All four renal action tests showed a dose-response curve, which correlated strongly with an elevated metal combination. Sambyal et al. [10] conducted an extensive analysis of existing statistical and ML algorithms for "retinopathy, neuropathy, and nephropathy" outcomes. Many machines learning techniques, including closest neighbor, deep neural networks, and naive Bayes, can improve the accuracy of predictive analyses. Both trials found that early treatment of glucose control dysfunction reduced the risk of developing Type 2 Diabetes and associated consequences. By promoting healthier lifestyles and so reducing healthcare costs, new analytical methodologies could be included into medical practice. This has been anticipated in the past [11].

This research suggests a complete model for predicting the spread of illnesses. Harimoorthy et al.'s method was developed with only a subset of the characteristics found in C.K.D., diabetes, and heart disease data. Using a modified version of the SVM with a racial bias kernel in R studio, I compared the "SVM-Linear, SVM-Polynomial, Random Forests, and Decision Tree approaches". AHDCNN was presented by Chen et al. to aid in the early identification of renal sickness. Experiments showed that when this method was applied to data sets dealing with chronic renal illness, diabetes, and heart disease, the results were quite accurate (98.3%, 98.7%, and 89.9%, respectively). The AHDCNN was validated as a reliable and efficient method of evaluating alternative deep-learning strategies [12].

2.2 Comparative Analysis

An accuracy rate of 96.51% was achieved on average by a CNN model that adhered to the recommendations. After conducting an exhaustive examination, it was discovered that the performance of the suggested system in data classification is superior when compared to the performance of other systems that have already been established. However, the authors of the current work used XGBoost Classifier to achieve an impressive level of accuracy of98.55%, with a score of 96.14% on the Jaccard scale, 98.17% on the Cross Validated scale, and 98.21% on the AUC scale. The results of these investigations are far more compelling than those of earlier research [13-17].

The machine learning repository at UCI, which was found to be filled with gaps in its contents, was the source of all of the information that was utilized in this inquiry. Imputation by means of KNN was the strategy used to address the previously described issue. During the course of this inquiry, prediction models were constructed utilizing a total of six distinct machine learning techniques. The diagnostic accuracy of the random forest model came in at a whopping 99.75%, making it the most accurate of all the machine learning models. After analyzing the shortcomings of the previously used models, a brand-new integrated framework was suggested as a solution [15, 16].

Over the course of ten separate simulations, this particular model, which makes use of logistic regression, random forest, and perceptron, demonstrated an accuracy level that averaged out to 99.83% overall. In the current study, data from UCI are combined with supplementary information gathered from a number of hospitals in Bangladesh. There was a total of 10,321 different pieces of information that were obtained. In this particular research project, eleven different algorithms were utilized to determine which one would prove to be the most successful in terms of accuracy for the particular data set that was analyzed [18].

For the purposes of this think about, the guess of CKD serves as a case ponder of healthcare administrations that are given through the utilize of cloud computing. In this consider, a single cleverly demonstrate for the forecast of CKD was made by combining straight relapse (LR) and neural organize (NN) strategies. Cloud computing and the Web of Things are basic to the working of the demonstrate. The discoveries show that the half breed show is able of anticipating CKD with a tall degree of exactness (precisely 97.8%), as appeared by the comes about. The authors consider incorporates the introduction of around eleven distinctive calculations. Calculating Taking after examination into each of the eleven potential approaches, the one decided to be the most successful has been recommended. The most viable strategy was chosen for utilize in the development of a demonstrate, which was taken after by the improvement of a user-friendly interface for it. This client interface gives you the capacity to put in the fitting realities, which makes it simpler to foresee the comes about [19].

An approach that places a more prominent accentuation on arriving at an exact conclusion or maybe than deciding the most viable course of treatment had been recommended in prior investigate. The essential objective is to assess and differentiate the two potential calculations for the forecast of CKD, with the conclusion objective of deciding which one is more exact. Amid the course of this request, a few diverse approaches to information mining were utilized, counting Arbitrary Woodland and Back Proliferation Neural Organize. It has been demonstrated that the Irregular Timberland Calculation has a precision score of 88.7 percent, while the Back Proliferation Calculation has an exactness score of 98.40 percent. Eleven diverse calculations were connected to the issue amid the course of the consider. Concurring to the cited inquire about, four unmistakable calculations were effective in accomplishing a level of exactness that was at slightest 95% exact. It has been decided that the XGBoost Classifier has an exactness of 98.55%, the Choice Tree and AdaBoost Classifier has an exactness of 98.11%, and the Irregular Timberland strategy has a precision of 97.09% [20][21].

2.3 Scope of the Problem

The South Asian nation of Bangladesh is sometimes referred to as the "River Kingdom" on account of the vast network of rivers and canals that can be found across the country. Bangladesh is well known for its extensive river network, which is comprised of the approximately 700 rivers that can be found snaking over the country's geography. As a direct result of this fact, the proximity of any potential future colonies to any current rivers will be the deciding factor in selecting their sites. The level to which the quality of the water has declined as a consequence of the Industrial

Revolution is substantial. People residing in this country are susceptible to a wide range of ailments as a direct consequence of the aforementioned factors. CKD is a significant problem that must be addressed in terms of its impact on the public's health in the United States. The most recent data that has been compiled and made accessible by the World Health Organization suggests that the death rate that may be attributed to renal disease has increased overthe past few years.

In the year 2018, the total number of fatalities that were registered in Bangladesh was 16,948, which corresponds to a percentage that is equivalent to 2.18% of all fatalities that have been recorded. Because of this, there is a risk that the individual will become permanently disabled or even die as a consequence. As a direct consequence of this, fast action is essential to assist in mitigating the effects of the scenario. It is feasible to detect CKD in its earlier stages if one monitors unusual health conditions and maintains thorough medical records. This allows for earlier diagnosis and treatment of the condition. If appropriate precautions are followed, the progression of chronic renal disease beyond the disease's current stage may be slowed down, and in some cases, it may even be stopped altogether.

2.4 Challenges

Intricate and time-consuming processes involving the excretion and reabsorption of many substances must first be completed. The time commitment needed to complete this process could be substantial. Maintaining a steadychemical balance in the body relies on this system working as it should at all times. Therefore, it is crucial that this system always works as intended. Depending on the stage of their disease, those with CKD will experience a complete or partial delay of this process. Thepresence or absence of renal disease may be difficult to diagnose if the patient's fluid intake is notaccurately measured.

Eliminating the presence ofnull values in the obtained dataset is one of the most difficult aspects of data analysis. One possibleoutcome is that a long period of time will be needed to finish this procedure. Creating a straightforward interfacefor the system that allows for data input and CKD prediction at any time proved to be a difficult problem for the group. The project's objective was to improve the system's usability.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The selection of a research approach is a crucial step in initiating a research endeavor. Initially, it is imperative to identify a problem, followed by the implementation of appropriate methodologies to address the identified issue. This chapter provides an explanation of the research topic. Subsequently, the algorithms employed to address the problem were elucidated, followed by a detailed exposition of the methodology, which was accompanied by a visual diagram to enhance comprehension.

3.2 Proposed System

After thoroughly examining all of the preceding methods, the proposed system may be presented. Figure 3.1 is a system diagram, which is chosen since it describes the specific method of the system.

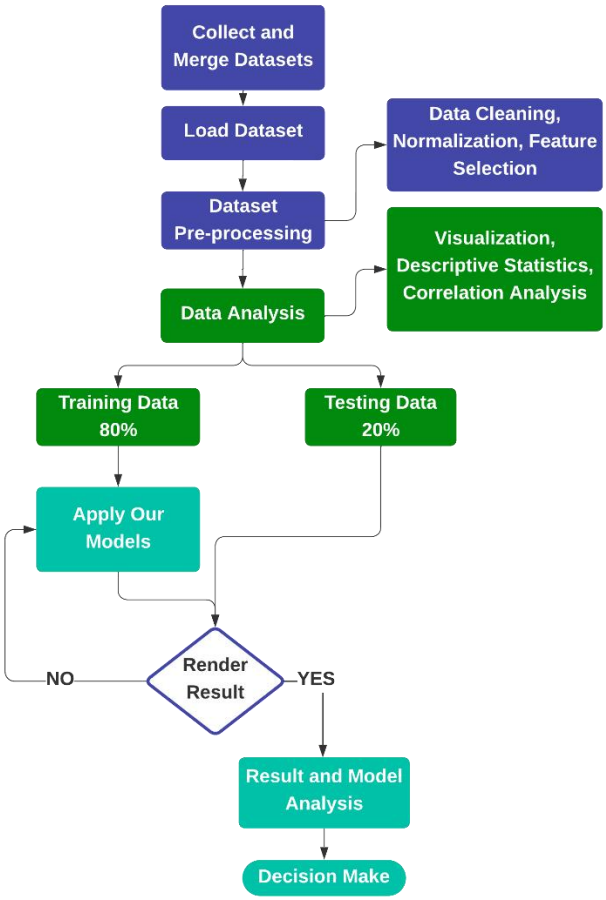


Figure 3.1: Overview of the Proposed System Methodology

3.2.1 Data Collection

To further our understanding of CKD, I need to gather data from real-world patients. A number of hospitals Popular Diagnostic Center, Savar Branch and the University of California, Irvine (UCI) in addition to the Kidney Foundation in Bangladesh, provided data for this research.

3.2.2 Dataset

Our first research step involves gathering information from numerous medical facilities. Next, I merged datasets into a single CSV file for more convenient reading and analysis. The requested dataset pertains to patient CKD blood sample data has 25 attributes. The accuracy of ML algorithms prediction is proportional to the completeness and quality of the data used to train them. With 10,321 rows and 25 columns, the raw data collection is incomplete.

Table 3.2.2: Dataset Description

Attributes Name	Count	Mean	Std	Min	25%	50%	75%	Max
Age	10321	51.51	16.95	2	42	54	64	90
Blood Pressure	10321	79.62	70.39	0	70	76	80	1400
Specific Gravity	10321	1.02	0.01	1.01	1.02	1.02	1.02	1.03
Albumin	10321	1.02	1.27	0	0	1	2	5
Sugar	10321	0.40	1.03	0	0	0	0	5
Red Blood Cells	10321	0.88	0.32	0	1	1	1	1
Pus Cell	10321	0.76	0.39	0	0.76	1	1	1
Pus Cell Clumps	10321	0.12	0.32	0	0	0	0	1
Bacteria	10321	0.06	0.23	0	0	0	0	1
Blood Glucose Random	10321	148.4	74.87	22	101	127	150	490
Blood Urea	10321	57.73	49.63	1.5	27	4.4	64	391
Serum Creatinine	10321	3.04	5.31	0.4	0.9	1.4	3.07	76
Sodium	10321	144.03	87.07	104	135	137.53	141	1436
Potassium	10321	4.43	0.73	1.4	3.9	4.63	4.8	7.6
Hemoglobin	10321	12.46	2.83	3.1	10.8	12.53	14.6	17.8
Packed Cell Volume	10321	38.75	8.09	9	34	38.75	44	54

White Blood Cell Count	10321	8403.41	2534.8	2200	700	8406	9400	26400
Red Blood Cell Count	10321	4.85	2.81	2.1	4.5	4.71	5.1	58
Hypertension	10321	0.37	0.48	0	0	0	1	1
Diabetes Mellitus	10321	0.35	0.48	0	0	0	1	1
Coronary Artery Disease	10321	0.09	0.28	0	0	0	0	1
Appetite	10321	0.79	0.40	0	1	1	1	1
Pedal Edema	10321	0.19	0.39	0	0	0	0	1
Anemia	10321	0.15	0.36	0	0	0	0	1
Class	10321	0.62	0.48	0	0	1	1	1

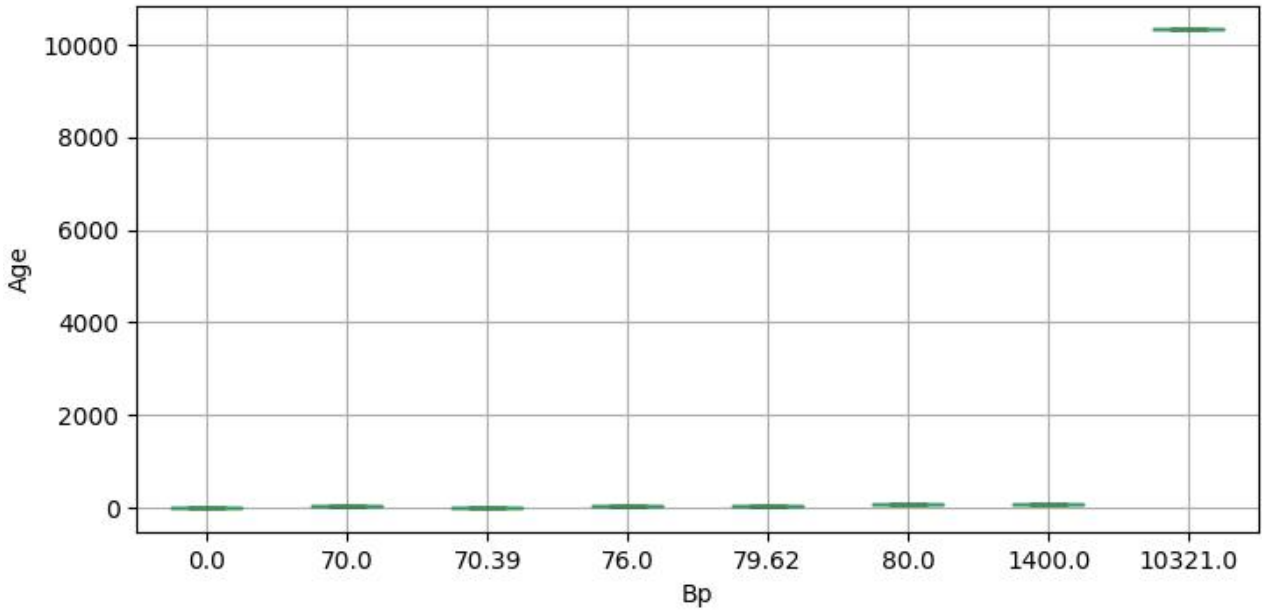


Fig 3.2.2: Box plot of 'Age' values, grouped by 'Bp' (blood pressure)

Box plot displays the distribution of 'Age' values, grouped by 'Bp' (blood pressure). It shows how the age distribution varies across different blood pressure levels. The y-axis represents age, and the x-axis represents different blood pressure categories.

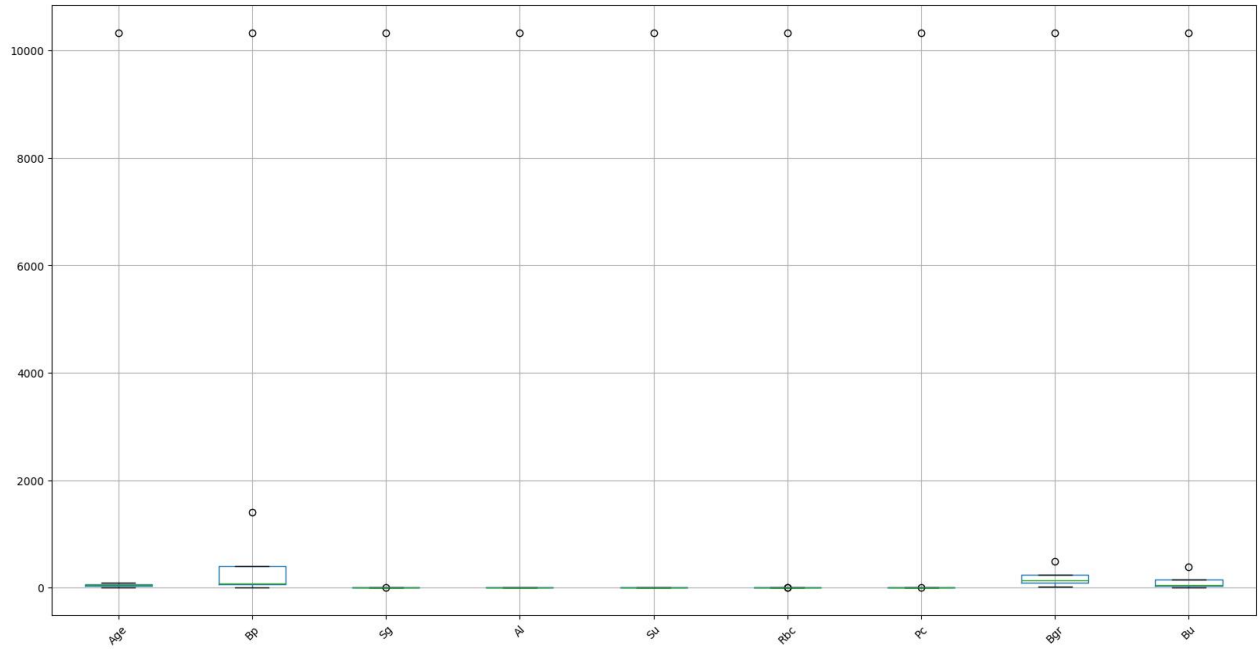


Fig 3.2.3: Box plot of Dataset

The second box plot shows the distribution of all numerical features in the dataframe `df`. Each box represents a different column (attribute) in the dataset. This visualization helps to identify potential outliers and understand the spread and central tendency of each feature. The plot is sized for better readability, and the x-axis labels (feature names) are rotated for better visibility.

3.2.3 Data Pre-processing

Quantitative and qualitative information that was lacking from the original dataset required conversion. The qualitative data was first transformed into quantitative form. That's why it's crucial to find a solution to the issue of missing values. To account for any gaps in the data, I simply took the mean. When conducting the study, variables were labeled as either independent (X) or dependent (Y). Then I standardized X, our independent variable, to provide more precise findings. Validation was performed on 20% of the whole dataset, whereas model training was performed on 80%. A model may be evaluated on the Testing subset of the dataset after it has been trained on the Training subset to see how well it predicts.

3.2.4 Imply Algorithms

Eleven distinct algorithms were evaluated and tested in search of the highest accuracy before the optimal method was selected. These are the eleven algorithms and their respective names: This ranges from modern techniques like Adaptive Boosting (AdaBoost) and eXtreme Gradient Boosting (XGBoost) to more conventional approaches like “Support Vector Machines (SVMs) andStochastic Gradient Descent (SGD)”.

Some common examples of these machine learning methods are the “Perceptron, Naive Bayes, K- Nearest Neighbors, Logistic Regression, and Linear Support Vector Classification”. Using these techniques, researchers found a broad variety of insights.

3.2.5 Model Analysis

Algorithm performance was measured using the “Confusion Matrix, Accuracy Score, Jaccard Score, Cross Validated Score, Area Under the Curve (AUC), Misclassification, Mean Absolute Error (MAE), and Mean Squared Error (MSE)”. I collected the information for future study. A quick summary of the data's predictive accuracy may be seen in the confusion matrix. Data projection accuracy may be evaluated using tools like the Accuracy Score metrics that may be used to evaluate the efficacy of the algorithms.

CHAPTER 4

EXPERIMENTAL RESULTS & DISCUSSION

4.1 Introduction

Obtaining a favorable outcome is crucial to the success of any investigation or endeavor. Success or failure is revealed in the final product. The results are provided in tabular form in this section. This chapter provides background information on the CKD dataset, including its purpose, methods of data collection, and results of feature significance analysis. The algorithms' outputs were then shown in a confusion matrix. The pertinent information has been presented in a tabular format to enhance comprehension.

4.2 Data Acquisition

There is one "target" (class) variable, eleven "measurement" variables, and thirteen "nominal" (referential) variables in this analysis. The categories are denoted by the two nominal values CKD (sample). The study included both people with CKD and healthy people (control group). The data set is mostly devoid of information because of several gaps. Table 4.2 provides a concise overview of the dataset.

Table 4.2: Data Acquisition

Attribute	Scale	Data Type	Missing Values (%)
Age	age in years	Numerical	2.27
Blood Pressure	in mm/Hg	Numerical	0
Specific Gravity	(1.005,1.010,1.015,1.020,1.02)	Nominal	0
Albumi n	(0,1,2,3,4,5)	Nominal	0
Sugar	(0,1,2,3,4,5)	Nominal	0
Red Blood Cells	(normal, abnormal)	Nominal	0
Pus Cell	(normal, abnormal)	Nominal	16.2
Pus Cell Clumps	(present, notpresent)	Nominal	0.97
Bacteria	(present, notpresent)	Nominal	0.97
Blood Glucose Random	in mgs/dl	Numerical	11.06
Blood Urea	in mgs/dl	Numerical	0
Serum Creatinine	in mgs/dl	Numerical	0
Sodium	in mEq/L	Numerical	0.72
Potassium	in mEq/L	Numerical	0
Hemoglobin	in gms	Numerical	0
Packed Cell Volume	in gms	Numerical	17.88

White Blood Cell Count	in cells/cumm	Numerical	0
Red Blood Cell Count	in millions/cmm	Numerical	0
Hypertension	(yes, no)	Nominal	0
Diabetes Mellitus	(yes, no)	Nominal	0
Coronary Artery Disease	(yes, no)	Nominal	0.5
Appetite	(good, poor)	Nominal	0.25
Pedal Edema	(yes, no)	Nominal	0.25
Anemia	(yes, no)	Nominal	0.25
Class	(ckd, notckd)	Nominal	0

Data entry and manipulation in a computer system was simplified by uniquely encoding each category variable (nominal variable). The values of rbc and pc were recorded as 1 and 0 respectively whether they were within or outside of the normal range. A binary scheme was used to classify the presence or absence of pcc and ba; 1 was assigned to the former and 0 to the latter. Therefore, 1 was assigned to affirmative answers and 0 to negative ones for the binary responses. One point was assigned for exceptional appet and zero points for subpar appet in a binary scoring system that measured its worth. The values of sg, al, and su were calculated using numerical correlations, which is an important fact given that these variables are categorical. In order to modify the category variables, factorization was used. To make it easier to tell them apart, I assigned each sample a random number between 1 and 10321. There were several gaps in the data due to incomplete surveys. Patient reluctance to implement pre-diagnosis safety measures might stem from a variety of factors. The data are incomplete since I don't know the diagnostic categories of the samples. In such a situation, it is essential to use a reliable imputation technique. After the categorical variables were encoded in the core CKD dataset, the missing values were processed and filled.

Methods known as "feature importance" techniques attempt to assess input attributes in terms of their predictive value. One manner in which experts convey the relevance of the features in the input characteristics of a prediction model is by assigning scores to those features. It's possible that the feature significance score, when applied to a dataset or a model, may provide some interesting insights. The accuracy of a prediction model may be enhanced by using it as well.

4.3 Result & Discussion

In this study, the CKD variable was assigned a positive value and the not CKD variable was allocated a negative value. Confusion matrices have been successfully used to analyze the effectiveness ofmachine learning systems and to show accurate findings.

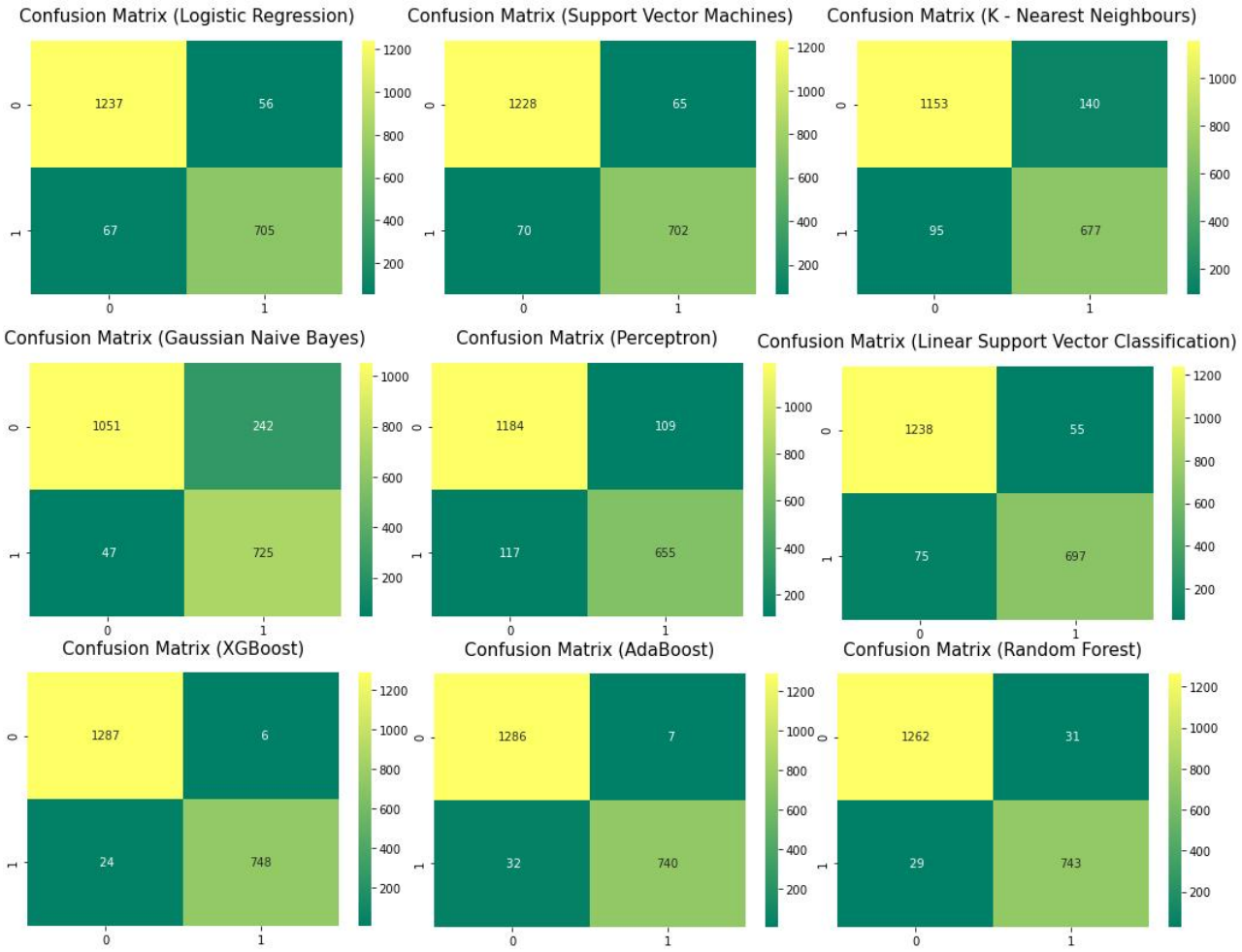


Fig 4.3: Confusion Matrix for Algorithms

4.4 Result Analysis

Once data has been gathered using various metrics including Precision, Recall, F1-Measure, and Accuracy, it may be analyzed. This analysis aims to determine the optimal algorithm among a set of algorithms, as well as identify any algorithms that may be underperforming in comparison to the others.

4.4.1 Accuracy

An algorithm's accuracy is a quantitative indicator of how well it performs. The quality of data sent into the system is what determines how well it performs. Performance may be evaluated probabilistically with the use of accuracy metrics. While eXtreme Gradient Boosting was shown to be the most accurate of the studied methods, Gaussian Nave Bayes was found to be the least accurate. When applied to gradient boosting machines, the eXtreme Gradient Boosting technique has shown to be very efficient and extensible. As an added bonus, it has showed promise in performing at the cutting edge of what boosted trees algorithms are capable of computationally. The original goal of its development was to make data analysis and model execution on computers more efficient. Figure 4.1 shows a chart showing how well the model predicts, details the relative contribution of each prediction technique.

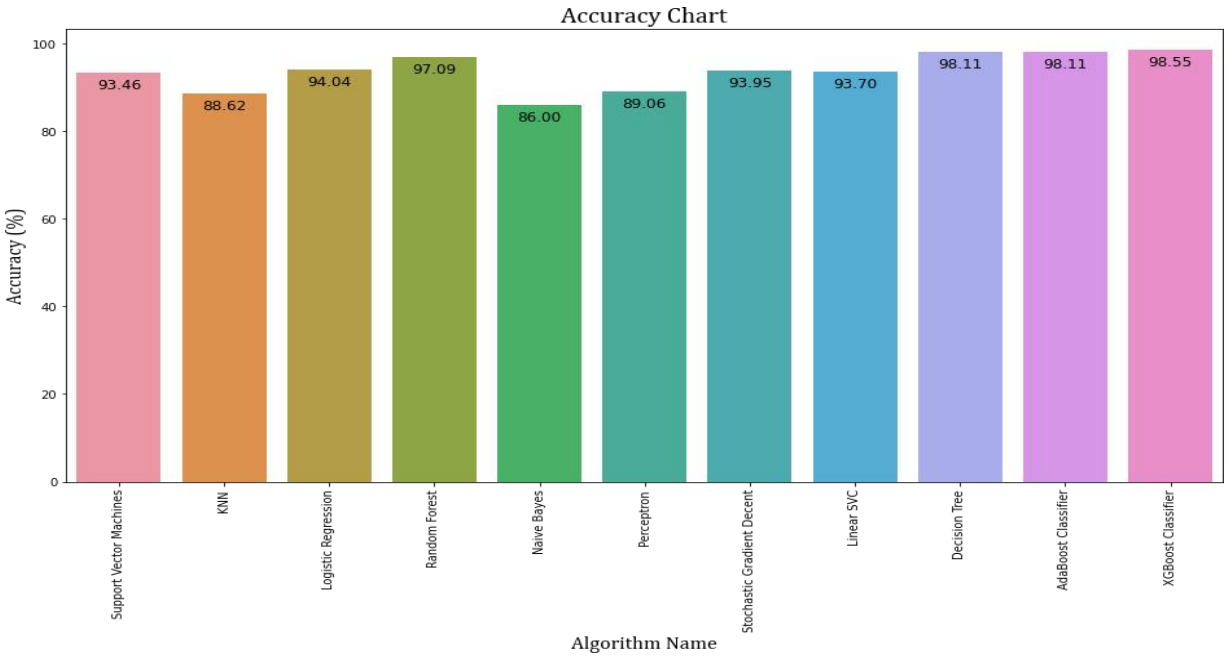


Figure 4.1: Accuracy Chart

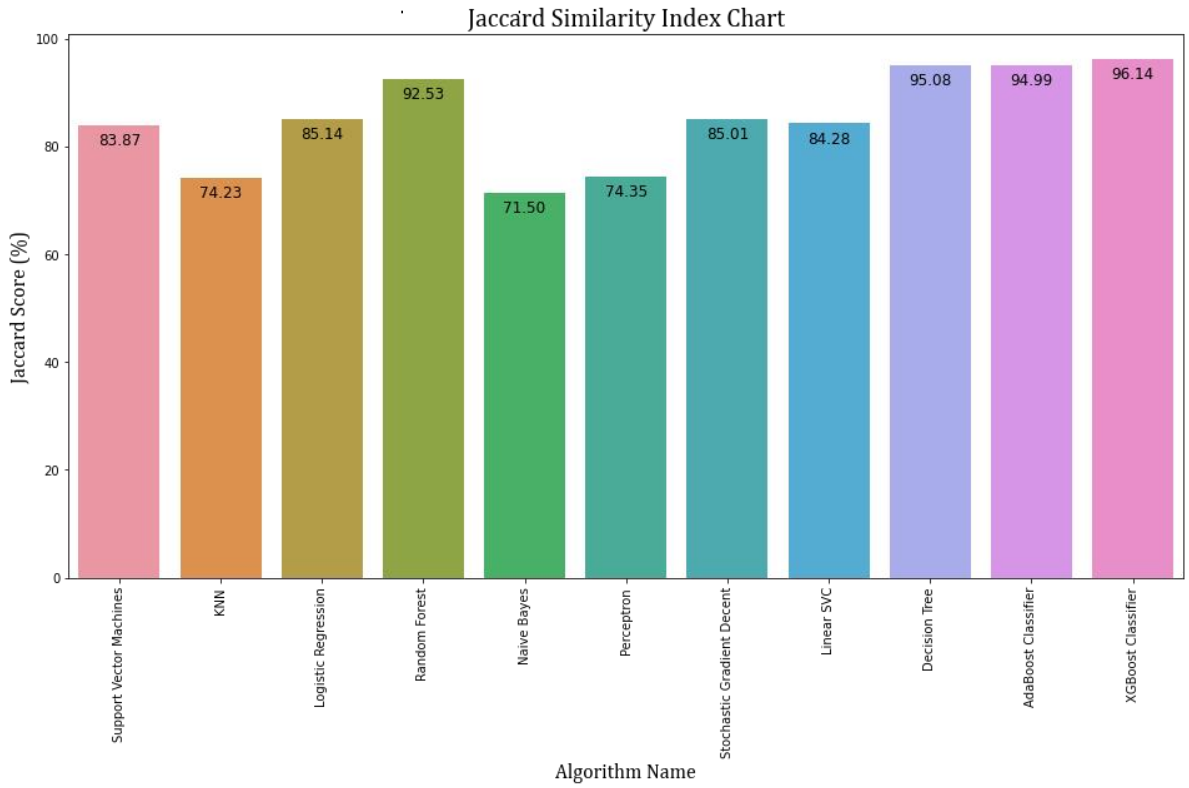


Figure 4.2: Jaccard Score Chart

4.4.2 Jaccard Score

In statistics, the "Jaccard index" measures the degree of similarity or dissimilarity between two samples. Equal emphasis is placed on both Intersection and Union. A mathematical measure of the degree of similarity between two collections of discrete data is called the Jaccard coefficient. Size of union is determined by dividing size of intersection.

4.4.3 Cross Validated Score

Statisticians use the cross-validation approach to assess the performance of machine learning models. Step one of the Cross Validation procedures involves arbitrarily splitting the dataset into k subsets. Here's what happens next: k models are trained on k-1 subsets of the data, and the remaining k-1 subsets are utilized for evaluation. The final evaluation score is arrived at by adding together all of the individual scores. The following phase involves putting the model to use by adjusting it to the whole dataset. The cross-validated score chart may be shown in Figure 4.3, while the percentages of matching algorithms used for prediction are shown in Table 4.1.

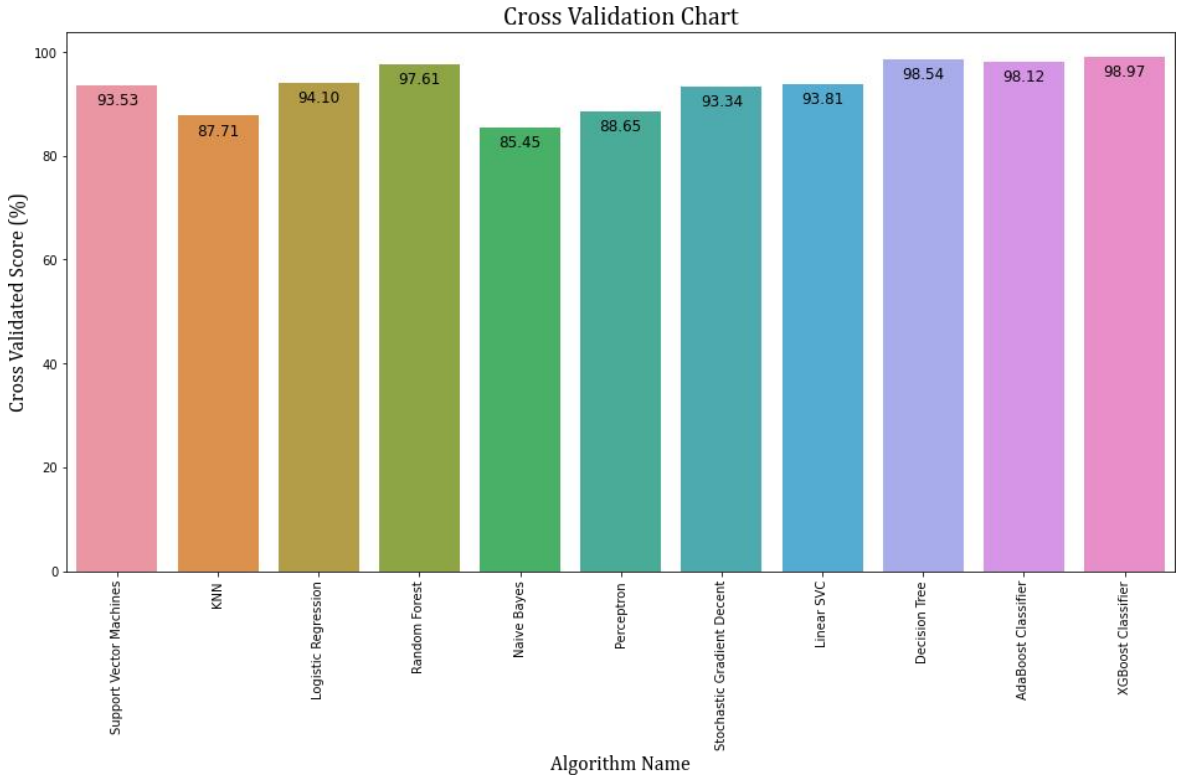


Figure 4.3: Cross Validated Score

4.4.4 AUC Score

When used with machine learning methods, this metric may help assess the breadth of viable system classifications, which can have applications in a number of contexts. The AUC may be determined by comparing the model's performance in the percentile of randomly selected positive instances, where it shows a big improvement, to its performance in the % of randomly selected negative cases, where it shows a considerable decline. This number may be zero, one, two, or three, with one being the highest possible value. There is a continuous range beginning at zero and ending at one, with zero being the absolute low. As can be seen in Figure 4.4 and Table 4.4.5, modelswith a perfect success rate get an accuracy value of 1, while those with a perfect failure rate receivean accuracy score of 0.

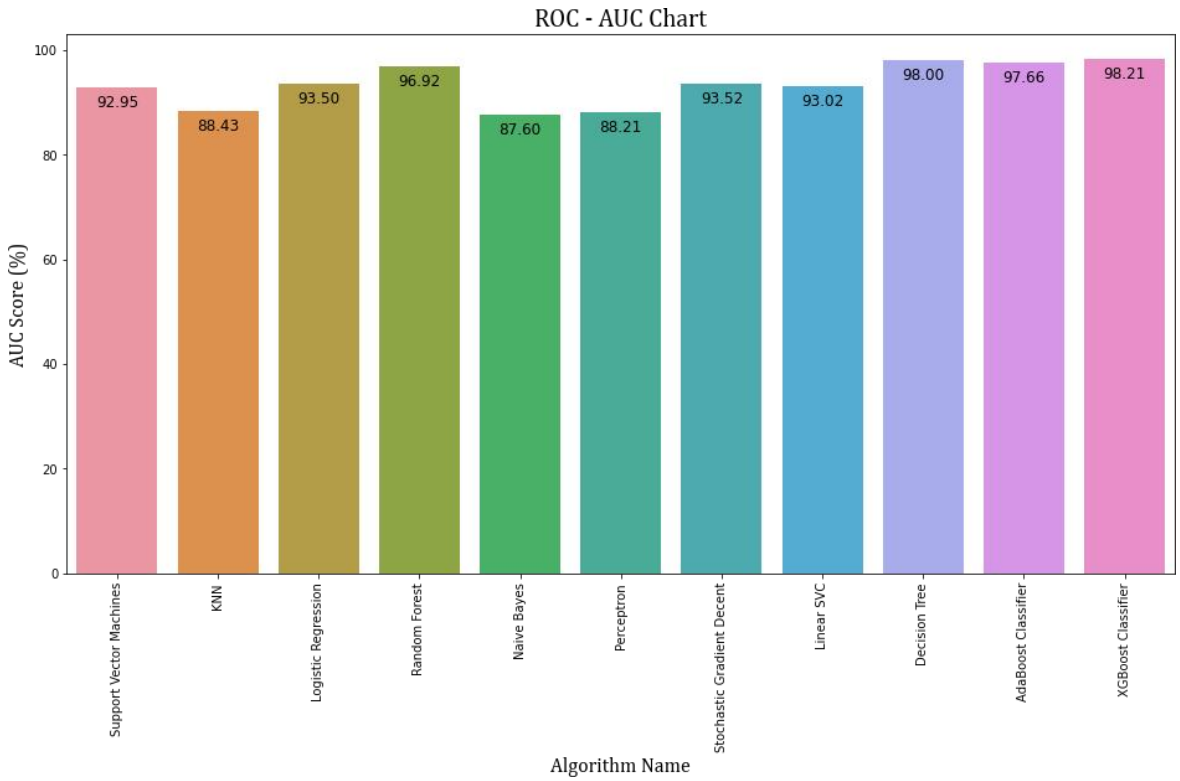


Figure 4.4: AUC Score Chart

4.4.5 ROC Curve

It is hard to overstate the relevance of using ROC analysis for establishing the accuracy of diagnostic tests and the amount of precision attained by statistical models when categorizing people as either healthy or unwell. This is because they are both very difficult tasks. One of the most helpful apps available today is called ROC curve analysis, and it is mostly used in the field of medicine. This analysis provides a graphical representation of how accurate a diagnostic test is, and it is widely considered to be one of the most valuable applications available. The significance curve score for your investigation is shown in figure 4.5 below.

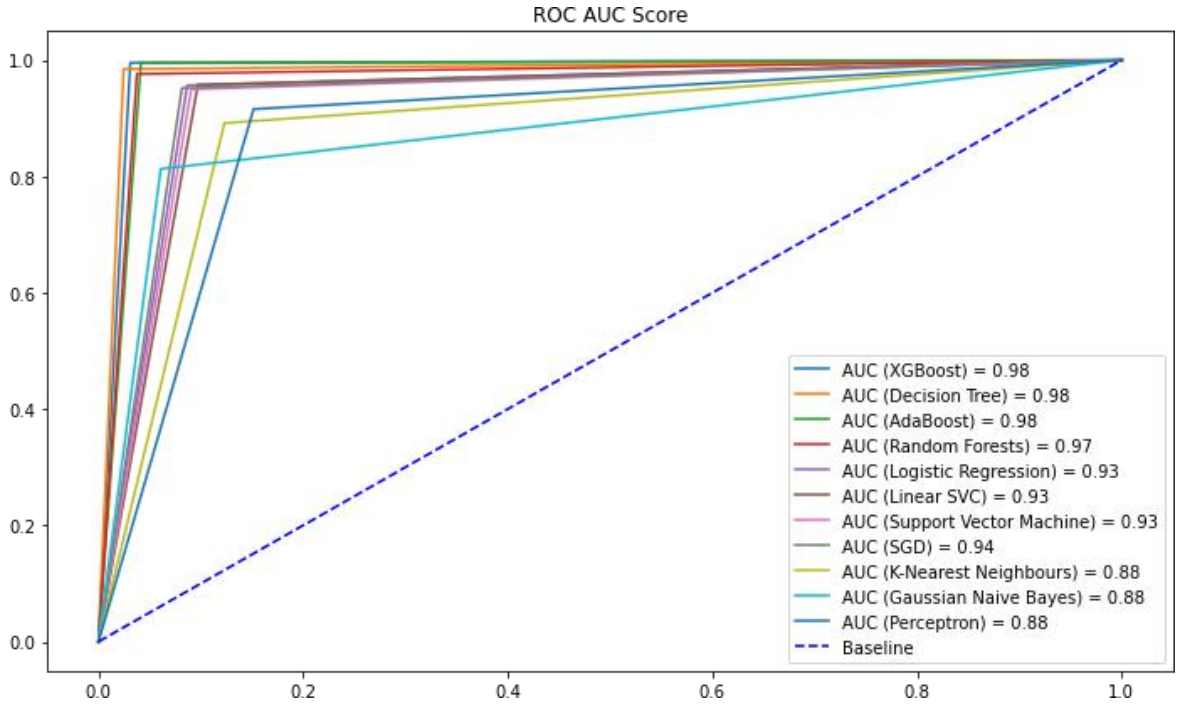


Figure 4.5: ROC Curve

Research showed that the XGBoost Classifier achieved the highest level of accuracy. Exact scores for this model's accuracy, Jaccard, Cross Validated, and Area Under Curve are as follows: 98.55%; 96.14%; 98.97%; 98.21%. Table 4.4.5 provides a concise and understandable breakdown of the degree of accuracy.

Table 4.4.5: Accuracy, Jaccard, Cross Validated and AUC Score

Algorithm Name	**Accuracy Score (%)**	**Jaccard Score (%)**	**Cross Validated Score (%)**	**AUC Score (%)**
XGBoost Classifier	98.55	96.14	98.97	98.21
Decision Tree	98.11	95.08	98.54	98
AdaBoost Classifier	98.11	94.99	98.12	97.66
Random Forest	97.09	92.53	97.61	96.92
Logistic Regression	94.04	85.14	94.1	93.5
StochasticGradient Decent	93.95	85.01	93.34	93.52
Linear SVC	93.7	84.28	93.81	93.02
Support Vector Machines	93.46	83.87	93.53	92.95
Perceptron	89.06	74.35	88.65	88.21
KNN	88.62	74.23	87.71	88.43
Naive Bayes	86	71.5	85.45	87.6

CHAPTER 5

IMPACT ON SOCIETY & SUSTAINABILITY

5.1 Introduction

It's vital to have a thorough discussion and analysis of the project's after-effects on society. This portion of the report delves into the study's findings about the three subcategories of participants with CKD. The initiative's positive social impacts are dissected in the first part. Then, the moral implications of the circumstance must be considered. The ethical considerations were thoroughly examined to comprehend the potential benefits of this project for the patients. Ultimately, the sustainability of the project was deliberated upon. The potential for future growth of the project and its capacity to benefit a larger population is being deliberated.

5.2 Impact on Society

The findings of this research have significant implications for public policy. considerable amount of time and effort. Due to the hectic schedules of individuals, visiting a medical facility for the purpose of ensuring one's health status is a time-consuming task. A significant duration has elapsed. A user-friendly interface has been developed to facilitate the input of data for the purpose of predicting the occurrence of kidney disease, yielding prompt results. Due to the global COVID-19 pandemic, there is an increased level of risk associated with undergoing medical testing at hospitals. It would be highly beneficial if individuals could conduct an analysis of their report from the comfort of their own residence. An ongoing survey is being conducted to gather data from both patients and doctors regarding the efficacy of this research. The poll results will undoubtedly reflect well on society as large and the relationship between patients and physicians.

5.3 Sustainability

It is encouraging that the research is taking a fresh approach by having participants use a website to monitor their CKD. Future opportunities to enhance the reliability of studieswill arise from the use of deep learning, artificial intelligence, and the Internet of Things. As a result, the findings of this research may open the door to a variety of sustainable development projects in the future. A mobile-friendly website or application may be developed. This study is focused only on CKD prognostication. Predicting kidney and other organ problems using machine learning is the focus of this research. Soon, an image of the skin might be uploaded over a web interface and utilized to identify skin conditions.

CHAPTER 6

FUTURE SCOPE & CONCLUSION

6.1 Implication for Further Study

The above concept may serve as the framework for a website created for any facility worldwide that treats patients with kidney disease, regardless of how advanced it is or where it is located. In the not-too-distant future, it's feasible that individuals will be able to access a website that employs the Internet of Things. CKD may be diagnosed earlier and preventive treatment can be given to the patient in the convenience of their own home if a website is used to collect the necessary information. The individual will subsequently have the capacity to determine whether or not they are afflicted with CKD. The website's Internet of Things architecture necessitates the storage of user-generated data, which is expected to be updated frequently. Subsequently, the model will acquire knowledge from every novel data point, thereby enhancing its precision as timeprogresses. Utilizing Deep Learning methodology and Neural Network algorithms may prove to be increasingly efficacious over time, with the potential for integration of Artificial Intelligence.

6.2 Recommendations

In diagnostic procedures, a standard level is established and any deviation from this normative level is indicative of an abnormal state. This system is capable of detecting such specific conditions. Serum Creatinine, Creatinine Clearance, Urine Albumin, and Hemoglobin are the primary attributes that play a significant role in predicting CKD in patients.

The content provided by the user is not suitable for academic purposes. Please provide academic content to be rewritten. In academic writing, abbreviations like "hgb" are popular. Men should have serum creatinine levels between 0.7 and 1.3 mg/dL, while women's levels should be between 0.6 and 1.1 mg/dL. For reference, a healthy female has a creatinine clearance of 88 mL/min, whereas a healthy guy has a clearance of 97 mL/min, on average. An increase in the amount of protein in the urine may be an indicator of renal impairment. Testing for the presence of protein in the urine is an important diagnostic step. Typically, this metric's values will hover between 0 and 8 mg/dL. A high protein urine level might be a sign of renal trouble. Testing for the presence of protein in the urine is an important diagnostic step. The normal range for a man's Hgb is 14–18 g/dL, whereas a woman's Hgb is 12–16 g/dL when she is healthy. Chronic renal illness may be detected early on if values deviate from norms. Hence, it is advisable to implement certain preventive measures during the initial phase. Modifying dietary patterns, engaging in physical activity, and increasing water intake are among the recommended lifestyle changes. It is imperative to adhere to the advice of a medical professional during such circumstances.

6.4 Conclusion

When CKD is detected and treated early, it is easier to control the condition and, in some situations, the patient may recover more quickly. Many different lab tests are needed to confirm the presence of CKD . The duration of these examinations may be rather lengthy. Researchers determined that a model trained with sufficient data might potentially foretell the onset of CKD at any point in time. To help patients acquire their CKD report, the authors of this study created a user interface in the form of a short online form. Eleven distinct machine learning algorithms were trained, and their accuracy, Jaccard score, Cross Validated score, and Area Under the Curve (AUC) were used to determine which one was the best for the given dataset. A total of three metrics are calculated, including a score for accuracy, a score for Jaccard, and a score for the Cross Validation. XGBoost's efficiency is high in comparison to other classifiers. Going with this option is often the wisest course of action (in the vast majority of cases, about 87%). When complete, the project website will be accessible from any healthcare facility that cares for individuals with renal disease. Individuals will be able to obtain CKD reports without the need to physically depart from their residences.

REFERENCES

[1] 'Kidney Disease in Bangladesh'. World Life Expectancy, https://www.worldlifeexpectancy.com/bangladesh-kidney-disease. Accessed 15 Jan. 2025.

[2] CDC. 'People with Certain Medical Conditions'. Centers for Disease Control and Prevention, 11 May 2023, https://archive.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

[3] Ma, Fuzhe, et al. 'Detection and Diagnosis of Chronic Kidney Disease Using Deep Learning-Based Heterogeneous Modified Artificial Neural Network'. Future Generation Computer Systems, vol. 111, Oct. 2020, pp. 17–26. DOI.org (Crossref), https://doi.org/10.1016/j.future.2020.04.036.

[4] Ali, Syed Imran, et al. 'Ensemble Feature Ranking for Cost-Based Non-Overlapping Groups: A Case Study of Chronic Kidney Disease Diagnosis in Developing Countries'. IEEE Access, vol. 8, 2020, pp. 215623–48. IEEE Xplore, https://doi.org/10.1109/ACCESS.2020.3040650.

[5] M, Manonmani., and Sarojini Balakrishnan. 'Feature Selection Using Improved Teaching Learning Based Algorithm on Chronic Kidney Disease Dataset'. Procedia Computer Science, vol. 171, Jan. 2020, pp. 1660–69. ScienceDirect, https://doi.org/10.1016/j.procs.2020.04.178.

[6] Lambert, Jerlin Rubini, et al. 'Identification of Nominal Attributes for Intelligent Classification of Chronic Kidney Disease Using Optimization Algorithm'. 2020 International Conference on Communication and Signal Processing (ICCSP), July 2020, pp. 0119–25. Semantic Scholar, https://doi.org/10.1109/ICCSP48568.2020.9182206.

[7] Nusinovici, Simon, et al. 'Logistic Regression Was as Good as Machine Learning for Predicting Major Chronic Diseases'. Journal of Clinical Epidemiology, vol. 122, June 2020, pp. 56–69. PubMed, https://doi.org/10.1016/j.jclinepi.2020.03.002.

[8] Segal, Zvi, et al. 'Machine Learning Algorithm for Early Detection of End-Stage Renal Disease'. BMC Nephrology, vol. 21, no. 1, Nov. 2020, p. 518. BioMed Central, https://doi.org/10.1186/s12882-020-02093-0.

[9] Sealfon, Rachel S. G., et al. 'Machine Learning, the Kidney, and Genotype-Phenotype Analysis'. Kidney International, vol. 97, no. 6, June 2020, pp. 1141–49. PubMed, https://doi.org/10.1016/j.kint.2020.02.028.

[10] Luo, Juhua, and Michael Hendryx. 'Metal Mixtures and Kidney Function: An Application of Machine Learning to NHANES Data'. Environmental Research, vol. 191, Dec. 2020, p. 110126. PubMed, https://doi.org/10.1016/j.envres.2020.110126.

[11] Harimoorthy, Karthikeyan, and Menakadevi Thangavelu. 'RETRACTED ARTICLE: Multi-Disease Prediction Model Using Improved SVM-Radial Bias Technique in Healthcare Monitoring System'. Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 3, Mar. 2021, pp. 3715–23. Springer Link, https://doi.org/10.1007/s12652-019-01652-0.

[12] Sambyal, Nitigya, et al. 'A Review of Statistical and Machine Learning Techniques for Microvascular Complications in Type 2 Diabetes'. Current Diabetes Reviews, vol. 17, no. 2, 2021, pp. 143–55. PubMed, https://doi.org/10.2174/1573399816666200511003357.

[13] Chen, Guozhen, et al. 'Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform'. IEEE Access, vol. 8, 2020, pp. 100497–508. IEEE Xplore, https://doi.org/10.1109/ACCESS.2020.2995310.

[14] Jashwanth Reddy, D., et al. 'WITHDRAWN: Predictive Machine Learning Model for Early Detection and Analysis of Diabetes'. Materials Today: Proceedings, Oct. 2020, p. S2214785320372382. DOI.org (Crossref), https://doi.org/10.1016/j.matpr.2020.09.522.

[15] Fauvel, Charles, et al. 'Prognostic Importance of Kidney, Heart and Interstitial Lung Diseases (KHI Triad) in PH: A Machine Learning Study'. Archives of Cardiovascular Diseases, vol. 113, no. 10, Oct. 2020, pp. 630–41. PubMed, https://doi.org/10.1016/j.acvd.2020.05.011.

[16] Yashfi, Shanila Yunus, et al. 'Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms'. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1–5. IEEE Xplore, https://doi.org/10.1109/ICCCNT49239.2020.9225548.

[17] Navaneeth, Bhaskar, and M. Suchetha. 'A Dynamic Pooling Based Convolutional Neural Network Approach to Detect Chronic Kidney Disease'. Biomedical Signal Processing and Control, vol. 62, Sept. 2020, p. 102068. ScienceDirect, https://doi.org/10.1016/j.bspc.2020.102068.

[18] Qin, Jiongming, et al. 'A Machine Learning Methodology for Diagnosing Chronic Kidney Disease'. IEEE Access, vol. 8, 2020, pp. 20991–1002. IEEE Xplore, https://doi.org/10.1109/ACCESS.2019.2963053.

[19] Abdelaziz, Ahmed, et al. 'A Machine Learning Model for Predicting of Chronic Kidney Disease Based Internet of Things and Cloud Computing in Smart Cities'. Security

in Smart Cities: Models, Applications, and Challenges, edited by Aboul Ella Hassanien et al., Springer International Publishing, 2019, pp. 93–114. Springer Link, https://doi.org/10.1007/978-3-030-01560-2_5.

[20] Snegha, J., et al. 'Chronic Kidney Disease Prediction Using Data Mining'. 2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE), 2020, pp. 1–5. IEEE Xplore, https://doi.org/10.1109/ic-ETITE47903.2020.482.

EARLY DETECTION OF CHRONIC KIDNEY DISEASE (CKD) USING OPTIMIZED MACHINE LEARNING MODELS

ORIGINALITY REPORT

18%	17%	3%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	9%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	7%
3	Submitted to Caleb University Student Paper	1%
4	123dok.com Internet Source	<1%
5	www.upgrad.com Internet Source	<1%
6	"Proceedings of the Fifth International Conference on Trends in Computational and Cognitive Engineering", Springer Science and Business Media LLC, 2024 Publication	<1%
7	namibian-studies.com Internet Source	<1%
8	web.archive.org Internet Source	

		<1%
9	www.banktrack.org Internet Source	<1%
10	www.mdpi.com Internet Source	<1%
11	Dinesh Goyal, Bhanu Pratap, Sandeep Gupta, Saurabh Raj, Rekha Rani Agrawal, Indra Kishor. "Recent Advances in Sciences, Engineering, Information Technology & Management - Proceedings of the 6th International Conference "Convergence2024" Recent Advances in Sciences, Engineering, Information Technology & Management, April 24–25, 2024, Jaipur, India", CRC Press, 2025 Publication	<1%
12	export.arxiv.org Internet Source	<1%
13	ijarcce.com Internet Source	<1%
14	www.federalregister.gov Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		