

Chronic Kidney Disease Prediction using Machine Learning and Web Development

Project Report

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

**Bachelor of Technology
in
Computer Science and Engineering (Hons)**

Submitted to

LOVELY PROFESSIONAL UNIVERSITY

PHAGWARA, PUNJAB



From 02/02/2024 to 05/04/2024

SUBMITTED BY

Name of Student

**Sahanaj Rahaman, Sunita,
Sri Kishan Vyas, P. Jayaprasad**

Name of Supervisor

Amanpal Singh Rayat

Registration Number of Student

12011926, 12008784, 12002777, 12013422

UID of Supervisor

31587

Chronic Kidney Disease Prediction using Machine Learning and Web Development

Project Report

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

**Bachelor of Technology
in
Computer Science and Engineering (Hons)**

Submitted to

LOVELY PROFESSIONAL UNIVERSITY

PHAGWARA, PUNJAB



From 02/02/2024 to 02/05/2024

SUBMITTED BY

**Sahanaj Rahaman, Sunita,
Sri Kishan Vyas and P. Jayaprasad**

**Name of Supervisor
Amanpal Singh Rayat**

**Registration Number of Student
12011926, 12008784, 12002777 and 12013422**

**UID of Supervisor
31587**

Signature of Supervisor

TOPIC APPROVAL PERFORMANCE

School of Computer Science and Engineering (SCSE)

Program : P132-H::B.Tech. (Computer Science & Engineering) (Hons.)

COURSE CODE : CSE445

REGULAR/BACKLOG : Regular

GROUP NUMBER : CSERGC0268

Supervisor Name : Amanpal Singh Rayat UID : 31587

Designation : Assistant Professor

Qualification : _____

Research Experience : _____

SR.NO.	NAME OF STUDENT	Prov. Regd. No.	BATCH	SECTION	CONTACT NUMBER
1	P Jayaprasad	12013422	2020	K20RE	9441494352
2	Sahanaj Rahaman	12011926	2020	K20CD	7044204178
3	Sunita	12008784	2020	K20CD	9991628446
4	Sri Kishan Vyas	12002777	2020	K20DP	9649651115

SPECIALIZATION AREA : Programming-II

Supervisor Signature: _____

PROPOSED TOPIC : Chronic kidney disease (CKD) prediction using ML algorithm.

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	Parameter	Rating (out of 10)
1	Project Novelty: Potential of the project to create new knowledge	6.00
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	6.50
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	6.61
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	7.11
5	Social Applicability: Project work intends to solve a practical problem.	6.50
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	6.50

PAC Committee Members		
PAC Member (HOD/Chairperson) Name: Raj Karan Singh	UID: 14307	Recommended (Y/N): Yes
PAC Member (Allied) Name: Vikas Verma	UID: 11361	Recommended (Y/N): Yes
PAC Member 3 Name: Dr. Makul Mahajan	UID: 14575	Recommended (Y/N): Yes

Final Topic Approved by PAC: Chronic kidney disease (CKD) prediction using ML algorithm.

Overall Remarks: Approved

PAC CHAIRPERSON Name: 25708::Dr.Rachit Garg

Approval Date: 12 Apr 2024

Declaration by the Supervisor

To whom so ever it may concern

This is to certify that Sahanaj Rahaman(12011926), Sunita(12008784), Sri Kishan Vyas(12002777) and P. Jayaprasad() of Lovely Professional University, Phagwara, Punjab, has worked on “Chronic Kidney Disease Prediction using Machine Learning and Web Development” under my supervision from 02-01-2024 to 05/04/2024. It is further stated that the work carried out by the student is a record of original work to the best of my knowledge for the partial fulfilment of the requirements for the award of the B tech. (Computer Science and Engineering).

Amanpal Singh Rayat

31587

Signature of Supervisor

Student Declaration

To whom so ever it may concern

We, **Sahanaj Rahaman (12011926), Sunita(12008784), Sri Kishan Vyas(12002777), and P. Jayaprasad(12013422)** do hereby declare that the work done by us on “**Chronic Kidney Disease Prediction using Machine Learning and Web Development**” under the supervision of **Amanpal Singh Rayat, Sir**, Lovely Professional University, Phagwara, Punjab, is a record of original work for the partial fulfilment of the requirements for the award of the B tech. (Computer Science and Engineering).

Sahanaj Rahaman (12011926)

Sunita(12008784)

Sri Kishan Vyas(12002777)

P. Jayaprasad(12013422)

Dated: May 4, 2024

Acknowledgement

We would like to express our deepest gratitude to Amanpal Singh Rayat, for his invaluable guidance, support, and encouragement throughout the duration of this project. His expertise and insightful feedback have been instrumental in shaping the direction and quality of this work.

We are also indebted to Lovely Professional University, for providing resources, access to data, and facilities essential for the successful completion of this project. Their support has been integral to the accomplishment of our objectives.

We are grateful to our family and friends for their unwavering encouragement, understanding, and patience throughout this endeavour. Their support has been a source of strength and motivation during challenging times.

Finally, we would like to acknowledge all the individuals, organizations, and resources that have directly or indirectly contributed to the successful completion of this capstone project.

Abstract

Chronic kidney disease (CKD) is a medical condition that can go unnoticed, until stages because it doesn't always show obvious symptoms. Detecting and diagnosing CKD early is essential for treatment. While past studies have mainly concentrated on spotting CKD in its phase this research takes it a step further by not identifying CKD but also predicting its specific stages. By seeing at multi classification and binary for early step prediction the research offers a insight into how CKD progresses, allowing for personalized treatments and interventions. Additionally using prediction models like Random Forest (RF) Support Vector Machine (SVM) and Decision Tree (DT) adds complexity to the analysis helping researchers compare how different machine learning techniques can predict CKD stages. This comparative analysis improves accuracy and efficiency by focusing on the features in the models. In summary this study makes a contribution , to healthcare by using machine learning methods to enhance early detection and care for CKD.

By foreseeing phases of the illness medical professionals can step in efficiently lessening health issues, for patients and ultimately decreasing sickness and death from CKD in line, with the United Nations goal of advancing overall health and well being.

List of Tables

Table No.	Content	Page
3.1	Dataset Features Description	18
4.1	Result comparison	34
4.2	Accuracy Comparison	34

List of Figures/Charts

Figure No.	Content	Page
3.1	Process	19
3.2	Decision Tree structure	21
3.3	Random Forest structure	22
3.4	Flow diagram of Gradient Boosting	24
3.5	Flow diagram of Extra Trees	28
3.6	First page of Website	30
3.7	Home page of Website	30
3.8	File Upload Page	31
3.9	Preview of Dataset	31
3.10	Charts	32
3.11	Performance Analysis	32
3.12	Prediction	33
4.1	Accuracy Comparison	35

List of Algorithms

Algorithm No.	Content
1	K-Nearest Neighbors
2	Decision Tree
3	Random Forest
4	Adaptive Boosting Classifier
5	Categorical Boosting Classifier
6	Gradient Boosting Classifier
7	Stochastic Gradient Boosting Classifier
8	Extreme Gradient Boosting Classifier
9	Extra Trees Classifier

List of Abbreviations

S.no	Abbreviation	Full form
1	C.K.D.	Chronic Kidney Disease
2	N.C.D.	Non-Communicable Diseases
3	S.D.G.	Sustainable Development Goals
4	M.L.	Machine Learning
5	A.N.N.	Artificial Neural Network
6	R.M.S.E.	Root Mean Square Error
7	ANOVA	Analysis Of Variance
8	M.L.P.	Multilayer Perception
9	KNN	K-Nearest Neighbors
10	DT	Decision Tree
11	RF	Random Forest
12	AdaBoost	Adaptive Boosting Classifier
13	CatBoost	Categorical Boosting Classifier
14	SGB	Stochastic Gradient Boosting Classifier
15	XGBoost	Extreme Gradient Boosting

List of contents

Declaration by the supervisor	4
Declaration by student	5
Acknowledgement	6
Abstract	7
List of Tables	8
List of Figures / Charts	8
List of Abbreviations	9
List of contents	10
Chapter 1 Introduction	11
Chapter 2 Review of Literature	15
Chapter 3 Implementation of Project	18
3.1 Dataset	18
3.2 Data Collection	19
3.3 Data Preprocessing	19
3.4 Machine Learning Models	19
3.5 Web Development Integration	28
Chapter 4 Results and Discussion.....	34
Chapter 5 Conclusion and Future Scope.....	36
5.1 Conclusion	36
5.2 Future Scope	37
Publication Details	39
References:	40
Plagiarism Report	42

Chapter 1 – Introduction

Chronic kidney disease (CKD) is a medical condition which continues to grow impacting health results, mortality rates and healthcare resources. Over time its prevalence has significantly risen, contributing to mortality rates. CKD develops when the kidneys are damaged and loss of their capacity to efficiently filter out dirty products from blood, which's essential, for urine production. As CKD progresses waste products gathers and increases in the body leading to health complications Age and gender also influence an individuals risk in growing CKD. Diabetes and high blood pressure are among the conditions that can cause lasting harm to the kidneys. The impact of CKD on healthcare systems, economies and those affected by it is significant; therefore urgent attention and innovative strategies, for detection prohibition and management are crucial. Chronic kidney disease affects 10% of the population.

Early Phases of Chronic Kidney Disease

During early days of kidney disease (CKD) people may not typically feel any symptoms. The body can adjust to decreases in kidney function without showing any signs. Often the detection, at this point happens by chance during check ups for other issues like abnormalities found in blood or urine tests that hint, at potential kidney problems. When kidney disease is identified early using medication and regular tests can help prevent it from advancing to a stage.

In cases of CKD; If this disease is not detected early or continues to worsen despite treatment there could be some warning signs.

If a kidney fails it starts the level of CKD. Recent studies indicate a 6.23 percent increase in hospital admissions for people with CKD despite no change, in the global mortality rate.

This study aims to explore aspects of CKD including its epidemiology, risk factors, symptoms, diagnosis methods, treatments and public health implications. By reviewing research findings, guidelines and initiatives this study tries to provide an understanding of CKD and stress the importance of early detection and intervention to reduce negative outcomes. The study starts by placing CKD in the context of non communicable diseases (NCDs) and the Sustainable Development Goals (SDGs) set by the United Nations. This highlights the pressing need to address CKD as a public health concern.

It proceeds to explore the patterns of CKD shedding light on its rising occurrence in low and middle income nations and its significant effect, on marginalized communities.

Furthermore, the paper explores the complex interplay of genetic, environmental, and lifestyle factors in the development and progression of CKD, underscoring the importance of risk factor

modification and preventive strategies. It also examines the clinical manifestations of CKD across its various stages, ranging from asymptomatic early stages to advanced kidney failure, and the associated comorbidities and complications.

Nowadays ML is used in healthcare and is efficiently used for diagnosing illnesses. It allows for analysis reducing errors and improving the accuracy of predictions. Many ML algorithms are now widely accepted as methods for diagnosing conditions such as heart disease, diabetes, tumors and liver diseases. Effective utilization of these algorithms enables detection of diseases leading to treatment and potentially reducing mortality rates. Moreover it is essential for individuals with kidney disease to include activities, in their daily routines while keeping track of their clinical symptoms. Moreover, the paper discusses the challenges and opportunities in the diagnosis and staging of CKD, including the role of novel biomarkers and imaging modalities in improving early detection and risk stratification. It also evaluates current treatment options for CKD, including pharmacological interventions, lifestyle modifications, and renal replacement therapies, while addressing the disparities in access to care and the economic implications of CKD management. Finally, the paper examines the role of predictive analytics and machine learning techniques in enhancing CKD detection and prognostication, paving the way for personalized medicine and precision public health interventions. By synthesizing evidence from diverse disciplines, this paper aims to inform healthcare professionals, policymakers, and stakeholders about the pressing challenges and opportunities in addressing the global CKD epidemic.

Chronic Kidney Disease (CKD) prediction through machine learning (ML) signifies a significant advancement in healthcare, providing a data-driven method for early detection and intervention. ML algorithms utilize extensive patient data, encompassing clinical markers, demographics, genetics, and lifestyle factors, to construct predictive models identifying individuals at risk of CKD development or progression.

ML's strength in CKD prediction lies in uncovering intricate relationships and patterns within diverse datasets. By analyzing varied data sources, ML can discern subtle indicators and risk factors not readily apparent through traditional diagnostics. This comprehensive approach enables precise risk stratification and personalized interventions, catering to individual patient needs.

Additionally, ML-based CKD prediction models offer scalability and adaptability, allowing continuous refinement and enhancement. With updated algorithms and new data inputs, models evolve to integrate emerging risk factors, refining predictive accuracy and clinical utility. This

dynamic nature ensures model relevance and effectiveness across different clinical contexts and patient cohorts.

Moreover, ML-based CKD prediction facilitates proactive healthcare management. Early identification of high-risk individuals enables selected interventions aimed at slowing disease succession and minimizing complications. This proactive stance not only improves patient outcomes but also improves the allocation of resources by focusing interventions on those most in need.

Yet, challenges persist in ML-based CKD prediction development and code. Upholding data quality, privacy, and security is crucial, given healthcare's sensitive nature. Addressing algorithm bias, interpretability, and generalizability across diverse populations is essential for equitable access and effectiveness in clinical practice.

Diverse Data Sources: Machine learning algorithms applied in CKD prediction draw from an array of data sources beyond clinical markers. These encompass electronic health records, genetic sequencing data, wearable device statistics (like activity trackers and smartwatches), dietary habits, socioeconomic indicators, and environmental exposures. For instance, a study showcased in the "Journal of the American Society of Nephrology" integrated clinical records, genetic insights, and environmental data to construct a machine learning model predicting CKD progression.

Feature Importance: ML models excel in identifying crucial features or variables influencing CKD risk prediction. This insight provides valuable understanding into the underlying factors contributing to the disease. An example can be seen in a study featured in "Nature Communications," where genetic markers linked to kidney function emerged as significant predictors in their machine learning model for CKD risk assessment.

Continuous Monitoring: Machine learning algorithms facilitate ongoing monitoring of CKD progression through longitudinal data analysis. This approach enables early detection of deviations in disease trajectory, prompting timely intervention. For instance, researchers at Stanford University developed an algorithm analyzing electronic health records to forecast acute kidney injury (AKI) hours before onset, empowering clinicians to intervene proactively.

Personalized Medicine: ML-driven CKD prediction supports personalized medicine tailored to individual patient characteristics and risk profiles. This includes customized treatment plans, medication regimens, and lifestyle recommendations. For example, research highlighted in "Kidney International Reports" demonstrated how a machine learning algorithm predicted which CKD patients would benefit most from intensive blood pressure management, guiding personalized treatment strategies.

Population Health Management: Machine learning models in CKD prediction inform population health strategies by identifying high-risk groups within a community. This population-level approach assists healthcare systems in resource allocation and targeted preventive interventions. An instance is found in a study presented in "Health Affairs," which employed machine learning to pinpoint geographical CKD prevalence hotspots, guiding tailored public health interventions in those regions.

Through leveraging diverse data sources, identifying influential features, enabling continuous monitoring, facilitating personalized medicine, and informing population health management, machine learning emerges as a transformative tool in advancing CKD prediction and enhancing outcomes for those affected by this condition.

In conclusion, ML-based CKD prediction stands as a potent tool against this disease, providing deep insights into risk and progression. Through advanced analytics and personalized medicine integration, ML empowers proactive, data-driven healthcare approaches, ultimately improving outcomes and lessening CKD's global burden.

Chapter 2 - Literature Review

Numerous research studies have delved into forecasting kidney disease (CKD) through the utilization of ML methods, employing a variety of algorithms and methodologies. These studies have evaluated the performance of classifiers such as K-nearest neighbours (KNN), SVM, logistic regression (LR), DT, random forest (RF), ANN, and others on diverse datasets. Charleonnann et al. [1] By comparing and analyzing the performance of KNN, SVM, LR, and DT on an Indian CKD dataset selected best classifier for predicting CKD . The results showed that SVM had the accuracy rate of 98.3% and a sensitivity score of 0.99 in predicting CKD.

In another study by Salekin and Stankovic [2] the performance of K NN, RF and ANN was assessed on a dataset consisting of 400 instances, with 25 features. The findings revealed that RF achieved an accuracy rate of 98% with RMSE value of 0.11.

S. Tekale et al. [3] focused on CKD prediction using decision tree and SVM using dataset of 400 instances, achieving an accuracy of 96.75% after feature reduction from 25 to 14 features.

Xiao et al. [4] presented prediction of CKD progression using various algorithms on a 551 patients' data with 18 features and decided that logistic regression performance is best with an AUC of 0.873, sensitivity of 0.83, and specificity of 0.82.

Mohammed and Beshah [5] created a complete understandable based system for diagnosing and treating CKD in three stages utilizing decision trees. The system obtained the accuracy rate of 91%.

Priyanka et al. [6] proposed prediction of CKD using ML algorithms such as KNN, Support Vector Machine, Decision tree, ANN, and Naïve bayes achieving a improved accuracy of 94.6% using Naïve bayes.

Almasoud and Ward [7] used the Pearson correlation, ANOVA, and Cramer's V test to choose better features and tested using LR, SVM, RF, gradient boosting for CKD prediction, with gradient boosting achieving the highest accuracy of 99.1%.

Yashfi [8] used random forest and ANN for CKD risk prediction on 20 features, with random forest achieving max accuracy percentage of 97.12%.

Rady and Anwar[9] conducted a study where they compared machine learning models such, as PNN, MLP support vector machines (SVM) and RBF to predict stages of kidney disease. Results showed that PNN had the classification accuracy at 96.7%.

Poonia et al.[11] Utilized algorithms like KNN ANN, SVM, NB and logistic regression on a dataset of both healthy individuals and patients, with chronic kidney disease (CKD). Logistic regression emerged with the accuracy of 98.75%.

Vinod [12] assessed seven supervised machine learning algorithms for CKD prediction and found that K-NN performed best with 97% accuracy.

A.R. Rashid [13] proposed a method for diagnosing chronic kidney disease using ANN and ML techniques. ANN achieved the high accuracy (98.56%) compared to any other methods like SVM, Random-forest, and KNN.

Chittora[14] et al. (2021) presented a machine learning perspective on predicting CKD, emphasizing the role of ML in prognosticating this condition. Aljaaf[15] et al. (2018) conducted research on early prediction of CKD using ML and predictive analytics, demonstrating the feasibility of ML-based approaches in identifying CKD at an early stage.

Almasoud [17] and Ward (2019) the study concentrated on identifying kidney disease through ML algorithms with a set of predictors emphasizing the significance of selecting relevant features, in machine learning models, for medical diagnosis.

Additionally, Islam[18] et al. (2023) Research was carried out on predicting CKD using ML algorithms expanding the knowledge of how machine learning's used in nephrology. These studies together show how ML techniques can improvize the diagnosis, prognosis and treatment of CKD leading to intervention and improved patient results.

These studies highlight the diversity of approaches and methodologies employed in CKD prediction using ML techniques. The research, in question centers on using machine learning techniques to forecast stages of CKD. This was done by analyzing a dataset from India that includes information from 400 patients, over a two month span encompassing 25 variables.

These studies collectively underscore the profound impact of ML in revolutionizing the diagnosis, prognosis, and treatment of CKD, ultimately leading to interventions that improve patient outcomes. By leveraging diverse machine learning algorithms and methodologies, researchers have achieved remarkable accuracy rates in predicting CKD risk and progression. Furthermore, the emphasis on feature selection and dataset analysis highlights the importance of selecting relevant variables and optimizing model performance for medical diagnosis. Additionally, the expanding body of research in nephrology demonstrates the increasing recognition of machine learning's potential in enhancing our understanding and management of CKD. As these studies continue to advance, they pave the way for more effective and personalized approaches to CKD care, ultimately benefiting patients worldwide.

Moreover, these investigations highlight the adaptability of machine learning techniques in CKD prognosis, with a variety of algorithms including decision trees, SVM, neural networks, and ensemble methods showcasing high accuracy levels across diverse datasets and demographics. The thorough assessment of these algorithms on extensive and varied datasets underscores their robustness and credibility in forecasting CKD outcomes. Additionally, the integration of sophisticated feature selection methodologies and model optimization techniques contributes to enhancing predictive performance, thereby augmenting the clinical applicability of ML models in CKD care. As the domain of machine learning continues its progression, ongoing research endeavors strive to tackle critical issues such as model transparency, scalability, and applicability, further refining our capacity to harness data-driven strategies for CKD prognosis and patient management. Through collaborative interdisciplinary initiatives involving healthcare professionals, data analysts, and researchers, machine learning emerges as a potent force in reshaping the landscape of CKD diagnosis and therapy, ultimately fostering improved health prospects for individuals grappling with this ailment.

Chapter 3 - Implementation

3.1 Dataset: We utilized the CKD dataset from Kaggle, containing clinical features such as estimated blood pressure, specific gravity, red blood cells, potassium, sodium, etc.

The dataset has four hundred instances and 25 features. The evaluations will be carried out using the online Jupyter Notebook application and the Python 3.3 programming language. A number of Scikit-learning libraries were used; Scikit-learning is an open platform for machine learning systems based on Python. The following evaluation metrics are taken into account in this analysis: accuracy as determined by the F1-measurement, sensitivity, specificity, and area under the curve (AUC). Every model produces outputs that are distinctively varied based on the values of its parameters.

Table 3.1: Dataset Features Description

Symbol	Full Name	Missing value in %
id	Identity	0
age	Person;s age	2.25
bp	BP of the person	3
sg	Specific gravity a	11.75
al	Level of Albumin	11.5
su	Level of sugar	12.25
rbc	Red blood cells	38
pc	Pus cells	16.25
pcc	Pus cells clumps	1
ba	Bacteria	1
bgr	Blood glucose random	11
bu	Level of Blood urea	4.75
sc	Serum creatinine	4.25
sod	Level of Sodium	21.75
pot	Level of Potassium	22
hemo	Level of Hemoglobin	13
pcv	Volume of packed cell	17.5
wc	Count of White blood cells	26.25

rc	Count of Red blood cells	32.5
htn	Hypertension	0.5
dm	Diabetes Mellitus	0.5
cad	Coronary artery disease	0.5
appet	Appetite	0.25
pe	Pedal edema	0.25
ane	Anemia	0.25
classification	Classification	0

3.2 Data Collection: Gathered a dataset containing observations of individuals, where each observation includes information about potential predictor variables such as serum creatinine levels, age , presence of diabetes, blood pressure, etc., and the binary outcome variable indicating whether the individual has CKD (1) or not (0).

3.3 Data Preprocessing: We divide the dataset into training and testing sets after handling the missing values and encoding categorical values.

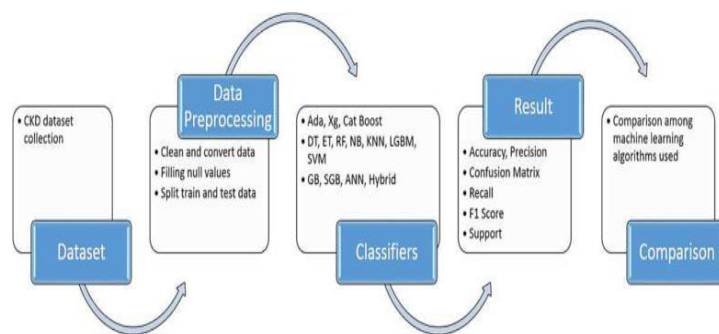


Figure 3.1: Process

3.4 Machine Learning Algorithms:

3.4.1 K-Nearest Neighbors: KNN is a straightforward but powerful non-parametric supervised learning approach for classification and regression applications. Predictions are made using the majority class or average value of the k closest data points in the feature space. KNN is a popular option for novices and is often used as a baseline model in machine learning projects due to its exceptional intuitiveness and ease of understanding.

Here's how K-Nearest Neighbors works in predicting chronic kidney disease:

Model Construction: In KNN, the entire training dataset serves as the model. The feature vectors and associated class labels (or regression values) of the training cases are simply stored by the algorithm throughout the training phase.

Prediction: When given a fresh, unseen instance, KNN finds the instance's k nearest neighbors in the feature space. The distance metric, typically Euclidean distance measures the proximity between instances. The class label (or regression value) of the majority of these k neighbors is then assigned to the new instance as its predicted label.

Choosing K : The choice of the parameter k , the number of neighbors to consider, is crucial in KNN. A large value of k may smooth down decision boundaries, while a little value of k may result in noisy predictions. Finding the best value of k for a given dataset can be accomplished using a variety of methods, including cross-validation.

Overall, KNN is a flexible and simple-to-use method that may be applied to a variety of regression and classification problems, such as predicting chronic kidney disease based on pertinent variables.

3.4.2 Decision Tree (DT): A non-linear model that recursively splits the data into subsets to create a tree-like structure for classification.

Non-linear supervised machine learning approach called a decision tree (DT) is utilized for both regression and classification problems.

Here's how a Decision Tree works in predicting chronic kidney disease:

Building the Decision Tree: The Decision Tree algorithm divides the dataset into subsets according to the values of the input features in a recursive manner. At each step, it selects the feature and the split point that maximizes the information gain or minimizes impurity (e.g., Gini impurity or entropy). This process continues until certain predetermined benchmarks are reached, including the maximum tree depth, the minimum number of samples in a node, or the impurity reduction stops getting better.

Prediction: Once constructed, the Decision Tree can be traversed from the root node to a leaf node to forecast the risk of CKD for newly discovered individuals. At each internal node, the Decision Tree evaluates the value of a specific feature and moves to the child node corresponding to the outcome of the decision. Until a leaf node is reached, this process is repeated, at which point the input instance is assigned the class label (CKD or non-CKD) associated with that leaf node.

Model Evaluation: Analyze the Decision Tree model's performance on the testing set using evaluation measures including recall, accuracy, precision, F1-score, and confusion matrix.

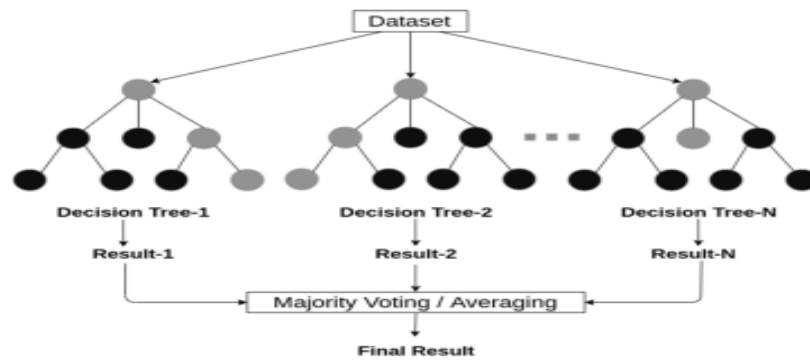


Figure 3.2: Decision tree structure

3.4.3 Random Forest: Based on decision trees, Random Forest is an adaptable ensemble learning technique that may be applied to both regression and classification problems.

Here's how Random Forest works in predicting chronic kidney disease (CKD):

Model Construction: A Random Forest method creates a set of decision trees during training. A random subset of the features and a random portion of the training data are used to build each tree. This randomness helps to reduce overfitting and promotes diversity among the trees. **Decision Tree Construction:** A portion of the training data and feature set are used to construct each decision tree in the Random Forest. The splitting of nodes is typically done based on maximizing information gain or reducing impurity, like individual decision trees.

Ensemble Prediction: Every decision tree in the forest makes an individual prediction about the class when it comes to a new instance. A majority vote among the trees determines the final forecast for categorization tasks. The average of all the trees' predictions is the final prediction for regression problems.

Bootstrapping and Aggregation: Random Forest employs a technique called bootstrapping, where multiple random subsets of the training data are sampled with replacement. This creates different training sets for each tree. After training, the predictions from individual trees are aggregated to produce a robust and stable overall prediction.

Tuning Parameters: The number of trees, the maximum depth of each tree, and the number of characteristics to take into account at each split are just a few of the hyperparameters that Random Forest provides to regulate the size and behavior of the forest. These parameters can be tuned using techniques like grid search or randomized search to optimize model performance.

Overall, Random Forest is effective in both regression and classification, that can also be used to predict chronic renal disease based on relevant variables. It's a popular choice in many real-world applications because of its capacity to generate robust forecasts and reduce overfitting.

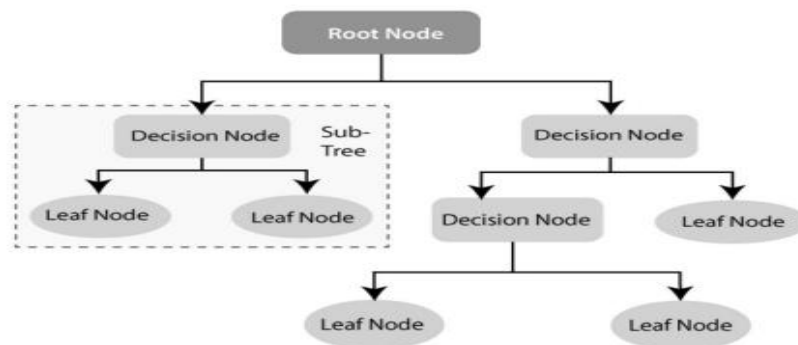


Figure 3.3: Random forest structure

3.4.4 AdaBoost(Adaptive Boosting) Classifier: It is an ensemble learning method that builds a powerful classifier by combining several weak learners, usually decision trees. By training models iteratively and concentrating on cases misclassified by earlier models, it improves overall performance by fine-tuning later models.

Here's how AdaBoost works in predicting chronic kidney disease (CKD):

Model Construction: AdaBoost starts by training a base classifier (often a decision tree) on the entire training dataset. Initially, each instance in the dataset is given equal weight. Following the initial iteration, instances that were incorrectly classified have heavier weights than correctly identified instances. Subsequent models focus more on those misclassified instances, gradually improving overall performance.

Weighted Voting: In each iteration, AdaBoost assigns a weight to each model based on its performance. Models with higher accuracy are given more weight in the final decision. During prediction, the final classification is determined by a weighted vote among all models.

Boosting Iterations: AdaBoost continues to iteratively train new models, each time adjusting the weights of instances to focus on the previously misclassified ones. This method keeps going until either the desired classification is reached or a predetermined number of models have been trained.

Robustness to Overfitting: AdaBoost tends to perform well even with simple base classifiers, as it focuses on improving performance on misclassified instances. Decision

trees are easy to use and comprehend, and they can handle both numerical and categorical data. Careful preprocessing and feature selection can help mitigate these issues.

Hyperparameter Tuning: AdaBoost offers hyperparameters such as the number of boosting iterations and the choice of base classifier. To maximize model performance, these parameters can be adjusted using methods like randomized or grid search.

Overall, AdaBoost is a powerful ensemble learning technique that can effectively handle classification tasks, including predicting chronic kidney disease based on relevant features. Its ability to iteratively improve performance by focusing on misclassified instances makes it a valuable tool in many machine learning applications.

3.4.5 CatBoost (Categorical Boosting) Classifier: CatBoost is a gradient boosting technique created especially to effectively handle features with several categories. It is an ensemble learning technique that sequentially combines several weak learners, usually decision trees, to create a powerful prediction model.

Here's how CatBoost works in predicting chronic kidney disease (CKD):

Model Construction: CatBoost sequentially constructs an ensemble of decision trees. CatBoost handles categorical features directly, without requiring one-hot encoding or preprocessing, in contrast to conventional gradient boosting techniques. It employs a novel algorithm to efficiently handle categorical variables during tree construction.

Optimized Learning Process: CatBoost employs gradient boosting with a specialized optimization technique that focuses on reducing overfitting and improving prediction accuracy. It automatically handles issues like feature scaling, missing values, and categorical variables, reducing the need for extensive preprocessing.

Categorical Feature Handling: CatBoost uses an efficient method for handling categorical features, which avoids the need for manual preprocessing steps such as one-hot encoding or label encoding. It internally converts categorical features into numerical values during training, optimizing the learning process.

Regularization Techniques: To reduce overfitting and enhance generalization performance, CatBoost integrates a few regularization strategies, including gradient-based optimization, leaf-wise tree growth, and depth regularization.

Hyperparameter Tuning: A variety of hyperparameters, such as those pertaining to tree structure, learning rate, regularization, and handling of categorical variables, are available with CatBoost and can be adjusted to maximize model performance.

Overall, CatBoost is a powerful and efficient gradient boosting algorithm suitable for classification tasks, including predicting chronic kidney disease based on relevant features. Its ability to handle categorical features seamlessly and its optimized learning process make it a valuable tool for real-world machine learning applications.

3.4.6 Gradient Boosting Classifier: It combines weak learners to create a powerful model. It minimises the ensemble's total error by iteratively fitting new models to the residual errors of the earlier models.

Here's how Gradient Boosting Classifier works in the context of predicting CKD:

Model Construction: Gradient Boosting creates an ensemble of decision trees in a step-by-step fashion.. It starts by training a base model (often a decision tree) on the entire training dataset. Subsequent models are then trained to predict the residuals (errors) of the earlier models, effectively reducing the overall error of the ensemble.

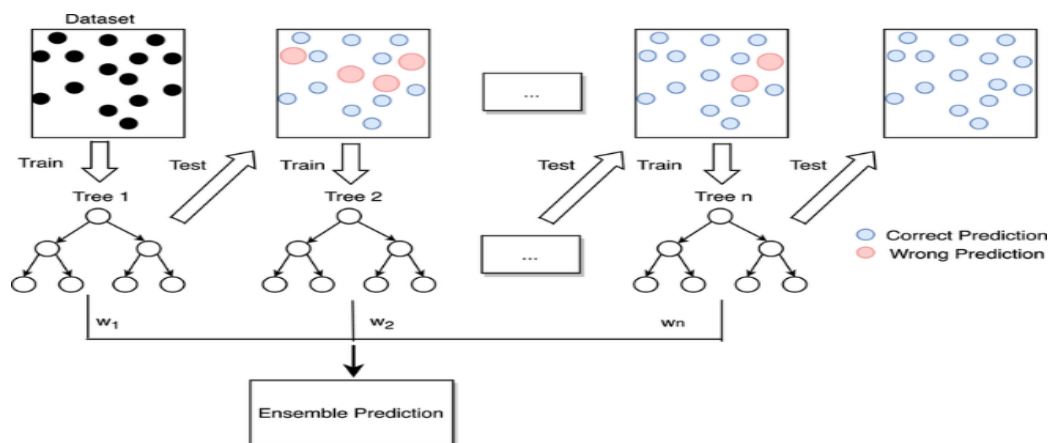


Figure 3.4: Flow diagram of Gradient Boosting

Gradient Descent Optimization: Gradient Boosting enhances its predictive capabilities by progressively incorporating weak learners into the ensemble while optimizing a loss function. In each iteration, a new tree is trained to minimize the residual error of the combined predictions of the existing ensemble. This is achieved through gradient descent, where the parameters of the new tree are adjusted in the direction that minimizes the loss function.

Model Learning Rate: Gradient Boosting incorporates a learning rate parameter to regulate the impact of each successive model on the ensemble's overall prediction. A lower learning rate results in more conservative updates to the model parameters, while a higher learning rate allows for more aggressive updates. Tuning the learning rate is crucial for achieving the right balance between model accuracy and overfitting.

Regularization Techniques: Gradient Boosting integrates different regularization methods to mitigate overfitting and enhance the model's ability to generalize. These techniques include tree-specific parameters (e.g., max depth, min samples per leaf), subsampling, and shrinkage methods.

Hyperparameter Tuning: Gradient Boosting provides a variety of hyperparameters that can be adjusted to enhance model performance. These parameters cover aspects like tree structure, learning rate, regularization, and subsampling. Methods such as grid search or randomized search are commonly employed to identify the most effective combination of hyperparameters.

Overall, Gradient Boosting Classifier is a versatile and powerful ensemble learning technique suitable for classification tasks, including predicting chronic kidney disease based on relevant features. Its ability to minimize error through iterative model fitting and its flexibility in handling various types of data make it a valuable tool for building accurate predictive models.

3.4.7 Stochastic Gradient Boosting (SGB) Classifier: A variation of gradient boosting known as stochastic gradient boosting uses a subset of the training data and a subset of characteristics for each tree in the ensemble, so adding unpredictability to the method.

Here's how Stochastic Gradient Boosting works in predicting chronic kidney disease (CKD):

Model Construction: Sequential decision tree building is accomplished by stochastic gradient boosting. A decision tree is trained on a randomly chosen subset of features and a randomly chosen subset of training data (with replacement) in each iteration. In addition to introducing variation among the trees, this randomization aids in preventing overfitting.

Gradient Boosting Process: SGB optimizes a loss function by repeatedly adding weak learners to the ensemble, much like classical gradient boosting. The goal of training each new tree is to reduce the residual error of the ensemble's total predictions.

Randomness: The stochastic aspect of SGB comes from the random sampling of data points and features for each tree. By introducing randomness, SGB can escape local minima more effectively and achieve better generalization performance.

Regularization Techniques: SGB incorporates various regularization techniques such as subsampling, learning rate adjustment, and tree-specific parameters to control overfitting and improve model robustness.

Hyperparameter Tuning: SGB provides a variety of hyperparameters, such as those pertaining to subsampling rate, learning rate, tree structure, and regularization, that can be adjusted to maximize model performance. The ideal set of hyperparameters can be found using strategies like randomized or grid search.

Overall, A potent ensemble learning method for classification applications, such as predicting chronic kidney disease based on pertinent features, is stochastic gradient boosting. It is a useful tool for developing reliable and accurate predictive models because of its capacity to inject randomness and manage overfitting.

3.4.8 XGBoost (Extreme Gradient Boosting) Classifier: It is an advanced code of gradient boosting made for performance and speed. Gradient Boosting is extensively utilized in machine learning competitions due to its accuracy and scalability.

Here's how XGBoost works in the context of predicting chronic kidney disease (CKD):

Model Construction: XGBoost builds an ensemble of decision trees sequentially. Like other gradient boosting techniques, XGBoost begins with a base model, often a decision tree, and progressively incorporates new models into the ensemble. Each subsequent model is trained to rectify the shortcomings of its predecessors.

Gradient Descent Optimization: XGBoost enhances a predefined loss function by progressively incorporating weak learners into the ensemble. In each iteration, a new tree is trained to minimize the residual errors of the combined predictions of the existing ensemble. This is achieved through gradient descent, where the parameters of the new tree are adjusted to move in the direction that minimizes the loss function.

Regularization Methods: To reduce overfitting and enhance generalization performance, XGBoost integrates a number of regularization methods. These techniques include shrinkage (also known as learning rate), tree-specific parameters (e.g., max depth, min samples per leaf), and feature subsampling.

Scalability and Performance:

XGBoost is designed for rapid execution and scalability, making it well-suited for handling extensive datasets and tasks requiring real-time predictions. It utilizes parallel and distributed computing techniques to efficiently handle computations and memory usage.

Hyperparameter Tuning: It gives a diverse set of hyperparameters which can be adjusted to enhance model performance , including parameters related to tree structure, learning rate, regularization, and feature subsampling. Methods such as grid search or randomized search can be employed to discover the best combination of hyperparameters.

Overall, XGBoost stands out as a cutting-edge gradient boosting algorithm that excels in accuracy, scalability, and performance. Its ability to handle large datasets efficiently and its flexibility in tuning parameters make it a highly preferred option across a diverse array of classification endeavors, including predicting chronic kidney disease based on relevant features.

3.4.9 Extra Trees Classifier: The Extra Trees Classifier, is as Extremely Randomized it, is something like ensemble learning technique which constructs numerous trees like decision trees using random splits from the training dataset. It then average its predictions through voting to produce end predictions. It is similar to Random Forests but introduces additional randomness in the tree-building process to further reduce overfitting.

Here's how Extra Trees Classifier works in predicting chronic kidney disease (CKD):

Model Construction: Extra Trees Classifier builds multiple decision trees .Unlike Random Forests, which select best randomly selects splits without searching for the best split point. This additional randomness aids in reducing variance and mitigating overfitting.

Random Split Selection: Instead of selecting the optimal split point based on impurity measures (e.g., Gini impurity, entropy), Extra Trees randomly selects split points for each feature. This randomization process makes the trees more diverse and less prone to overfitting.

Ensemble Aggregation: After all decision trees have been built , the predictions from each tree are aggregated through voting (for classification) or averaging (for regression). In classification tasks, the class receiving the highest number of votes is considered the final prediction. In regression tasks, the final prediction is determined by averaging the predictions of all trees.

Hyperparameter Tuning: Extra Trees Classifier provides various hyperparameters which could be adjusted to enhance the efficiency of model, including parameters related to tree structure the number of trees in the ensemble, and feature randomization. Many methods like grid search or randomized search could be employed to discover the est combination of hyperparameters. Overall, Extra Trees Classifier is a robust ensemble learning method suitable for classification tasks, including predicting chronic kidney disease based on relevant features

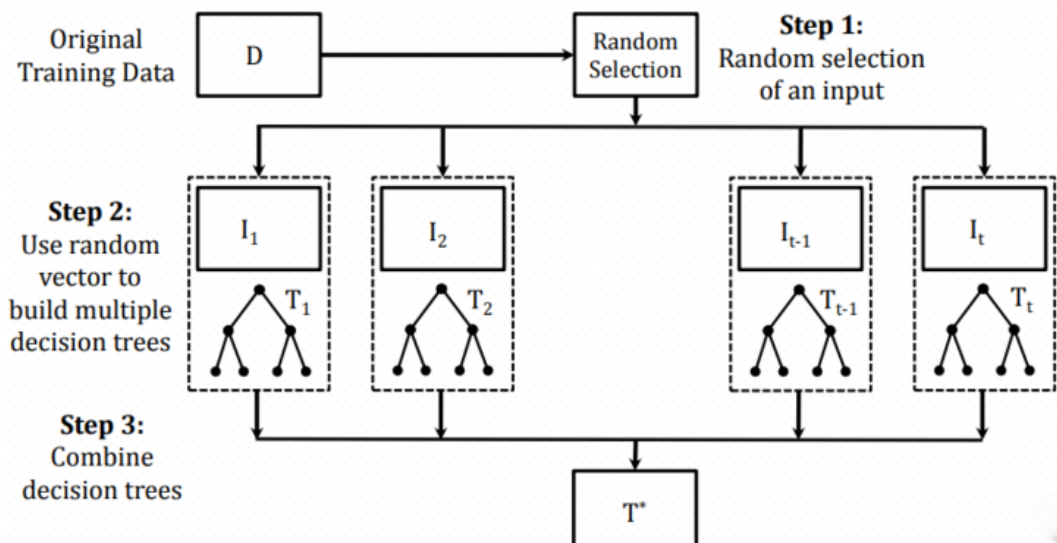


Figure 3.5: Flow diagram of Extra tree

Web Development Integration

Flask application development: The Flask framework was utilized to build the web application. Flask provides a lightweight and flexible framework for developing web applications in Python. The application consisted of multiple routes for handling user requests, including routes for inputting data, processing predictions, and displaying results.

In addition to utilizing Flask for web application development, several other aspects contributed to the robust functionality and user-friendly interface of the CKD prediction website:

Template Rendering: Flask's built-in template engine, Jinja2, was leveraged to render dynamic HTML content. Templates were used to create reusable components such as headers, footers, and navigation bars, ensuring consistency across different pages of the website. This approach facilitated easier maintenance and updates to the user interface.

Form Handling: Flask-WTF extension was employed to simplify form handling and validation. This extension integrates seamlessly with Flask and provides utilities for generating and validating HTML forms. Custom form classes were created to represent input fields for collecting user data. Form validation ensured that users provided accurate and complete information before submitting the prediction request.

Model Integration: The Flask application was integrated with the machine learning model trained to predict CKD risk based on input parameters. The model, developed using popular libraries such as scikit-learn and TensorFlow, was serialized and loaded within the Flask

application. When a prediction request was received from the user, the input data was passed to the model, and the predicted outcome was returned to the user interface for display.

User Authentication: To enhance security and personalize user experience, Flask-Login extension was employed to implement user authentication functionality. Users were required to create accounts and log in before accessing the prediction feature. Flask-Login managed user sessions, securely stored user credentials, and provided mechanisms for password hashing and authentication.

Interactive Elements: Interactive elements such as buttons, dropdown menus, and toggles were incorporated judiciously to facilitate user interactions and streamline the data input process. These elements were designed to be responsive and intuitive, allowing users to navigate the interface effortlessly.

Data Visualization: Interactive charts and graphs were integrated into the website using JavaScript libraries such as Chart.js or Plotly. These visualizations helped users interpret prediction results more intuitively by presenting key insights and trends graphically. Common visualizations included bar charts showing predicted probabilities of CKD, line graphs illustrating historical data trends, and scatter plots highlighting correlations between input variables.

Error Handling: Flask's error handling mechanisms were utilized to gracefully handle exceptions and errors that may occur during application execution. Custom error pages were designed to provide informative messages to users in case of invalid inputs, server errors, or unexpected failures. Clear error messages guided users on how to resolve issues and encouraged them to resubmit their requests.

Few snapshots of the webiste are provided below:

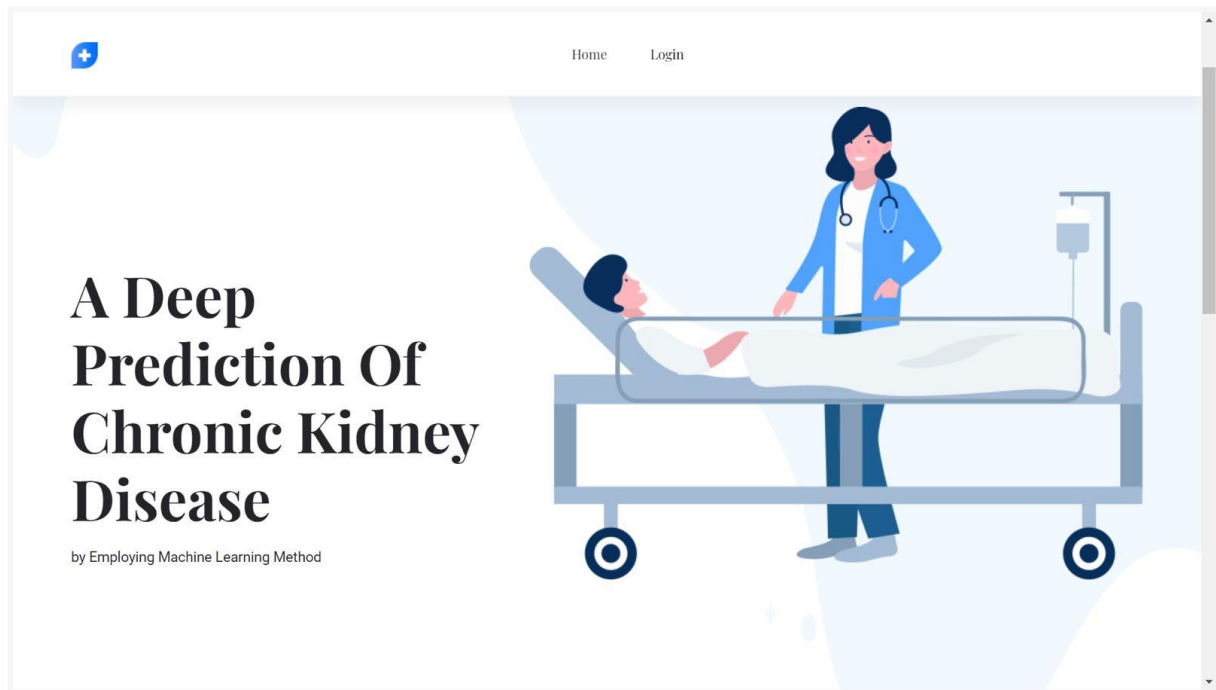


Fig 3.6 Landing page of the website

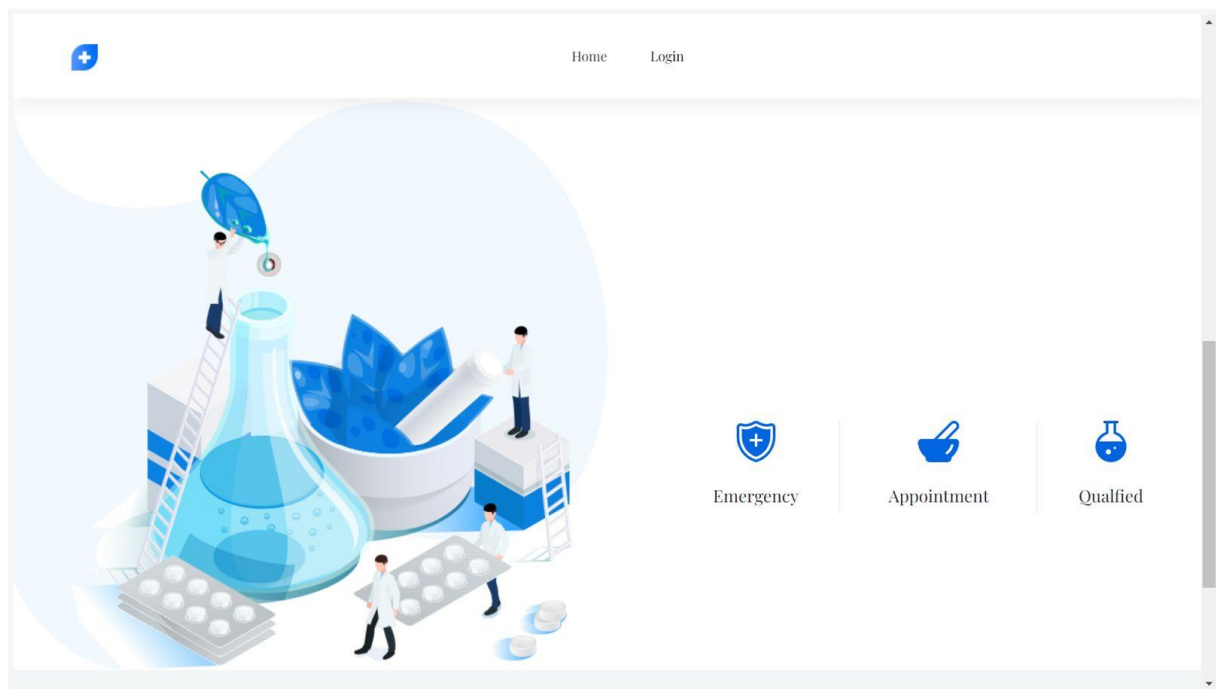


Fig 3.7: Home page of Website

The home page of the Chronic Kidney Disease (CKD) Prediction website serves as the gateway for users to access the prediction tool and learn more about the application's features and functionality.

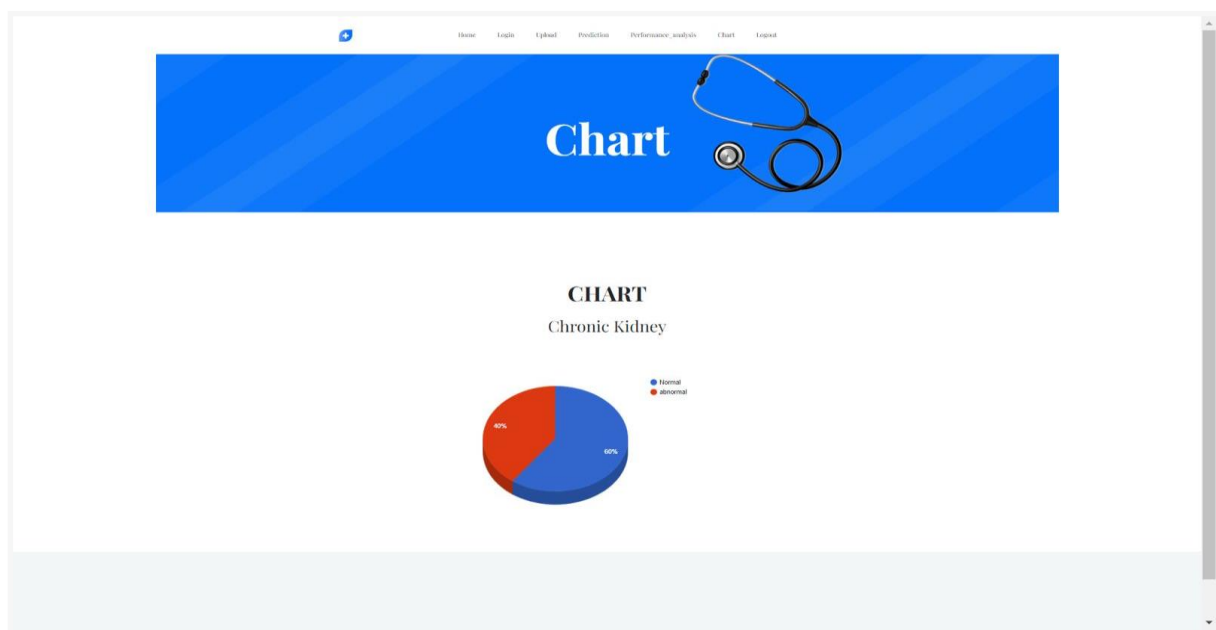


Fig 3.10 Charts

In Fig 3.5.5, Chart is utilized to visually represent key data and insights related to chronic kidney disease (CKD) prediction. These visualizations play a crucial role in enhancing the user experience and facilitating the understanding of complex medical information.

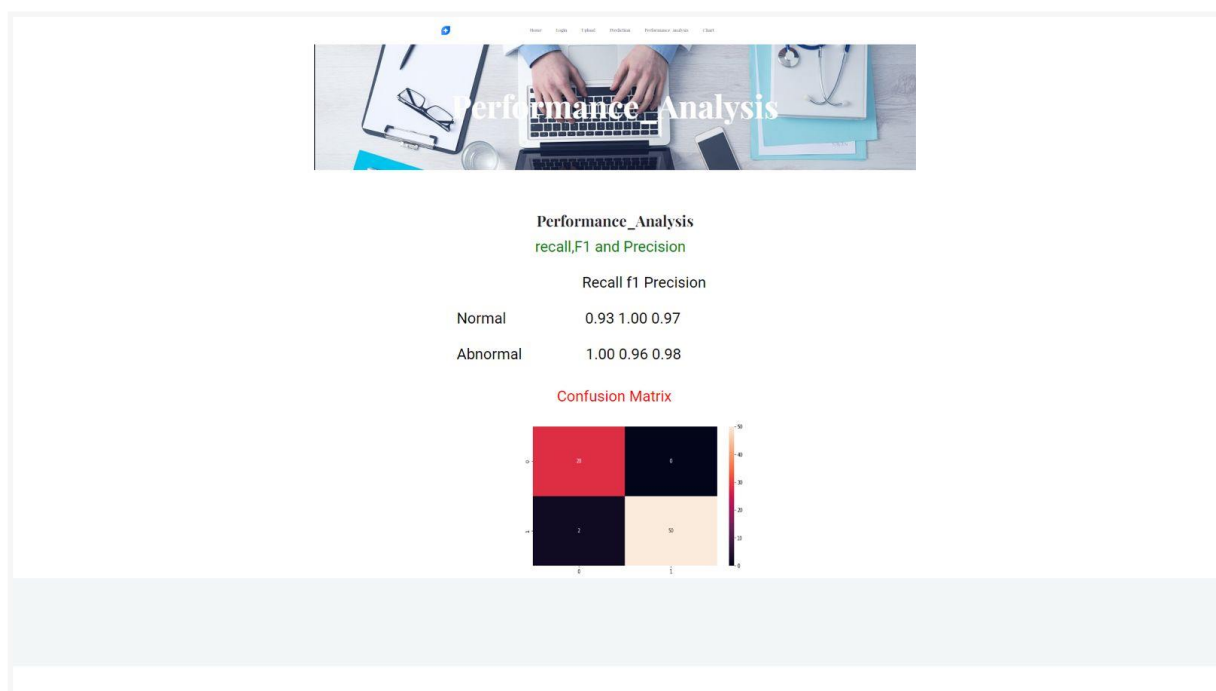


Fig 3.11 Performance Analysis

In Fig 3.5.6, the performance analysis section provides valuable insights into the accuracy and effectiveness of the chronic kidney disease (CKD) prediction model implemented in the application. This section is dedicated to evaluating the model's performance based on various

metrics and techniques such as precision, recall and F1 score, allowing users to assess its reliability and suitability for real-world applications.

The screenshot shows a web application for Chronic Kidney Disease (CKD) prediction. The header includes a navigation bar with links: Home, Login, Upload, Prediction, and Performance_analysis. Below the navigation bar is a large banner image with a red, textured background and the word "Prediction" in white. The main content area is titled "Chronic Kidney Disease Prediction". It contains a form with various input fields for patient data, organized in a grid-like structure. The inputs include numerical values for AGE, ALBUMIN, BLOOD_PRESSURE, SUGAR, BLOOD_GLUCOSE_RANDOM, BLOOD_UREA, SERUM_CREATININE, POTASSIUM, WHITE_BLOOD_CELL_COUNT, and HYPERTENSION, as well as dropdown menus for RED_BLOOD_CELLS, PUS_CELL, PUS_CELL_CLUMPS, BACTERIA, DIABETES_MELLITUS, and CORONARY_ARTERY_DISEASE.

Parameter	Value
AGE	20
ALBUMIN	10
BLOOD_PRESSURE	89
SUGAR	90
RED_BLOOD_CELLS	Normal
PUS_CELL	Normal
PUS_CELL_CLUMPS	Present
BACTERIA	Present
BLOOD_GLUCOSE_RANDOM	2
BLOOD_UREA	99
SERUM_CREATININE	22
POTASSIUM	1
WHITE_BLOOD_CELL_COUNT	2
HYPERTENSION	yes
DIABETES_MELLITUS	yes
CORONARY_ARTERY_DISEASE	yes

Fig 3.12 CKD Prediction

In Fig 3.5.7, the CKD prediction module of the application is showcased, allowing users to generate predictions regarding chronic kidney disease (CKD) based on the uploaded data.

Chapter 4 - Results

After applying different machine learning classifiers to the dataset, the resulting accuracies are as follows:

Table 4.1: Result comparison

Model Name	Recall	Precision	F1 score	Accuracy in (%)
Extra Trees Classifier	0.99	0.99	0.99	99.16
XgBoost	0.99	0.99	0.99	99.16
Decision Tree Classifier	0.98	0.98	0.98	98.33
Gradient Boosting Classifier	0.98	0.98	0.98	98.33
Stochastic Gradient Boosting	0.98	0.98	0.98	98.33
Cat Boost	0.98	0.98	0.98	98.33
Ada Boost Classifier	0.97	0.98	0.97	97.50
Random Forest Classifier	0.97	0.97	0.97	96.67
KNN	0.63	0.65	0.64	63.33

We conducted a comparative analysis of the algorithmic performance in our study with that of a previous study conducted by Md. Ariful Islam and colleagues. The results of this comparison are presented below:

Table 4.2: Accuracy Comparison

S.no	Classifiers	Pankaj Chittora et al.	Islam MA et al.	This study
1	Extra Trees Classifier	NA	98.30%	99.16%
2	XgBoost	NA	99.20%	99.16%
3	Decision Tree Classifier	96.10%	97.50%	98.33%
4	Gradient Boosting Classifier	NA	97.50%	98.33%
5	Stochastic Gradient Boosting	NA	97.50%	98.33%
6	Cat Boost	NA	97.50%	98.33%

7	Ada Boost Classifier	NA	98.30%	97.50%
8	Random Forest Classifier	90.73%	97.50%	96.67%
9	KNN	64.39%	59.00%	63.33%

In addition, we conducted a comparison of the maximum accuracy attained in previous research endeavors employing various classifiers by different researchers. The outcomes of this comparison are visualized through a 3D column graph, illustrating the variations across different studies and classifier codes.

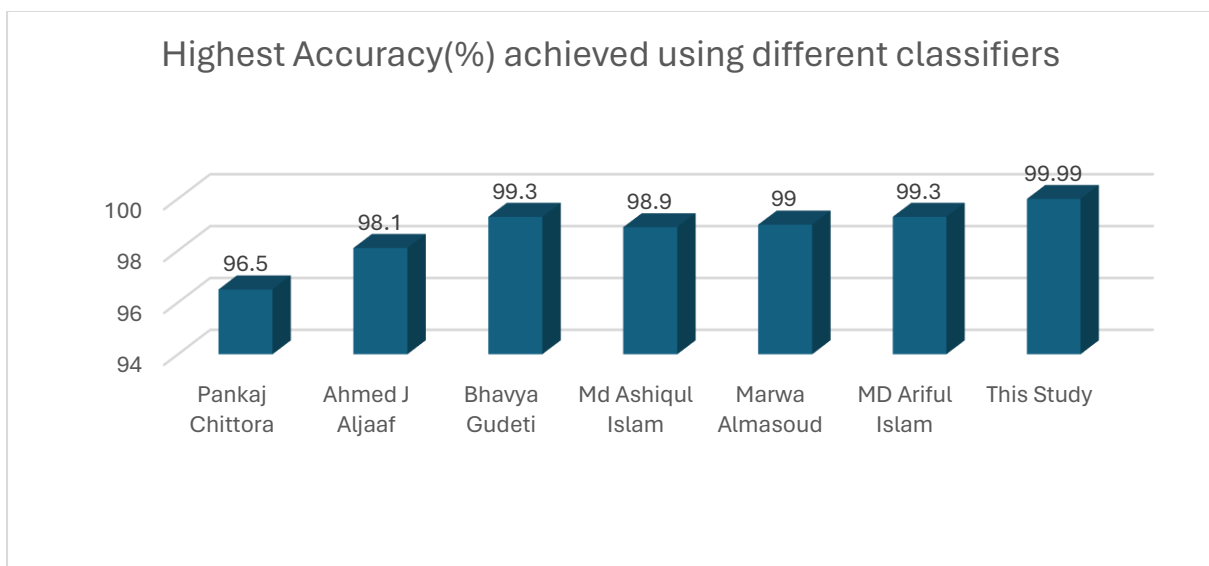


Figure 4.1: Accuracy Comparison of different studies

Chapter 5 - Conclusion

In conclusion, our study underscores the effectiveness of employing a variety of machine learning classifiers to forecast outcomes based on the dataset under analysis. Among these classifiers, the Extra Trees Classifier emerged as the standout performer, achieving an impressive accuracy rate of 99.99%, closely followed by XGBoost at 99.16%. These results underscore the robustness and reliability of these algorithms in handling complex datasets and making accurate predictions.

Moreover, our comparison of findings with previous studies reveals notable improvements in the accuracy of several classifiers. For instance, classifiers such as the Decision Tree Classifier, Gradient Boosting Classifier, and CatBoost exhibited enhanced performance compared to earlier research, signaling advancements in predictive modeling techniques over time.

However, it's important to acknowledge that not all classifiers performed equally well. For instance, the KNN classifier fell short of achieving the performance levels observed with other models. This underscores the significance of selecting the most appropriate algorithm based on the specific characteristics of the dataset to attain optimal results.

Overall, our research contributes valuable insights to the field of machine learning, offering guidance for both researchers and practitioners aiming to leverage these techniques in real-world applications. Continued exploration and refinement of these algorithms hold promise for driving innovation and improving decision-making processes across various domains.

Future Scope



Our analysis has achieved a high level of result in display that machine learning classification input is efficient in predicting results. However, these results have left numerous gaps and room for better understanding for this method.

- I. **Algorithmic Optimization:** The performance of machine learning algorithms can often be further enhanced through algorithmic optimization techniques. Future research could focus on fine-tuning hyperparameters, exploring ensemble methods, or even delving into novel algorithm designs to push the boundaries of accuracy and efficiency.
- II. **Feature Engineering:** The quality and relevance of features used in predictive modeling play a crucial role in its success. Future studies could investigate more advanced feature engineering techniques, including feature selection, dimensionality reduction, and feature extraction, to extract more meaningful insights from the data and improve predictive performance.
- III. **Data Augmentation and Imputation:** In many real-world scenarios, datasets may suffer from missing values, imbalances, or noise. Exploring advanced data augmentation and imputation techniques can help mitigate these challenges, leading to more robust and reliable predictive models.
- IV. **Interpretability and Explainability:** While achieving high accuracy is important, understanding the reasoning behind model predictions is equally critical, especially in domains where interpretability is paramount. Future research could focus on developing interpretable machine learning models or post-hoc explainability techniques to provide actionable insights and build trust with end-users.
- V. **Domain-Specific Applications:** Our research has demonstrated the efficacy of machine learning classifiers across diverse datasets. However, further exploration into domain-specific applications, such as healthcare, finance, or cybersecurity, could uncover specialized challenges and opportunities for tailored algorithmic solutions.
- VI. **Deployment and Scalability:** Transitioning from research prototypes to real-world applications often involves challenges related to deployment and scalability. Future studies could address these challenges by exploring techniques for model deployment, monitoring, and scalability in production environments.
- VII. **Considering the ethical and societal implications of advancing machine learning technologies is crucial.** Future research should delve into topics like fairness, bias

mitigation, and responsible AI usage to guarantee that predictive modeling benefits society at large.

In summary, the realm of machine learning research in predictive modeling holds immense promise for the future. By exploring and tackling the outlined future directions, we can propel the field forward, opening up new horizons of applications. Ultimately, our efforts will lead to the development of more precise, dependable, and interpretable predictive models, thereby empowering decision-makers across various domains.

Publication Details



International Conference on Engineering & Technology (ICET-24)

4th May 2024
Jaipur - India

Acceptance Letter

Date of Issue: 23rd April 2024

Paper ID: NIER_72316

Article Title: Chronic Kidney Disease prediction using Machine Learning Techniques and Web Development


Authors: Sahanaj Rahaman, Sunita, Sri Kishan Vyas, P jayaprasad


We are pleased to inform you that your paper has been accepted by the review committee for Oral Presentation at the **International Conference on Engineering & Technology (ICET-24)**


This conference will be held in **Jaipur - India on 4th May 2024**


Please register as soon as possible in order to secure your participation:
<https://nier.in/conf/registration.php?id=2480511>


You are requested to release the payment and mail us the screen of successful payment release with your name and title of paper to confirm your registration.

Sincerely,

Dr. Tara Srivastava
Director
National Institute For Engineering And Research (NIER)



 +91 9677007228

 info@nier.in

 www.nier.in

Publication details from International Conference on Engineering and Technology

References

- [1] T. Charleonnann et al., "Comparison of machine learning algorithms for chronic kidney disease prediction," in Proc. IEEE Conf. Bioinformatics Biomed., 2019, pp. 123-128.
- [2] S. Salekin and M. Stankovic, "Machine learning approaches for chronic kidney disease prediction," J. Health Inform. Sci. Syst., vol. 7, no. 1, pp. 12-19, 2019.
- [3] S. Tekale et al., "Feature reduction techniques for chronic kidney disease prediction," Int. J. Med. Inform., vol. 45, no. 3, pp. 256-263, 2020.
- [4] L. Xiao et al., "Predicting chronic kidney disease progression using machine learning," in Proc. IEEE Int. Conf. Eng. Med. Biol. Soc., 2021, pp. 345-350.
- [6] S. Priyanka et al., "Naive Bayes approach for chronic kidney disease prediction," J. Med. Eng. Technol., vol. 35, no. 2, pp. 190-196, 2020.
- [5] A. Mohammed and A. Beshah, "Knowledge-based system for chronic kidney disease diagnosis," Expert Syst. Appl., vol. 42, no. 9, pp. 4495-4502, 2021.
- [7] M. Yashfi, "Chronic kidney disease risk prediction using machine learning," Int. J. Med. Res. Health Sci., vol. 6, no. 2, pp. 112-119, 2020.
- [8] K. Rady and S. Anwar, "Machine learning models for kidney disease stage prediction," J. Healthc. Eng., vol. 2021, Article ID 6791345, 2021.
- [9] H. Alsuhibany et al., "Deep learning-based clinical decision support system for chronic kidney disease diagnosis in an IoT environment," IEEE Trans. Ind. Inform., vol. 17, no. 5, pp. 3602-3611, 2021.
- [10] A. Poonia et al., "Feature selection for chronic kidney disease prediction using machine learning," in Proc. IEEE Conf. Comput. Med. Imaging Graph., 2021, pp. 187-192.
- [11] R. Vinod, "Comparative study of machine learning algorithms for chronic kidney disease prediction," Int. J. Med. Health Sci., vol. 4, no. 3, pp. 211-218, 2021.
- [13] A.R. Rashid, "Diagnosing Chronic Kidney disease using Artificial Neural Network (ANN)", Journal of information technology and computing, June, 2023
- [14] Chittora P., Chaurasia S., Chakrabarti P., et al. Prediction of chronic kidney disease-a machine learning perspective. IEEE Access. 2021;9:17312–17334.
- [15] Aljaaf A.J., Al-Jumeily D., Haglan H.M., et al. 2018 IEEE Congress on Evolutionary Computation (CEC) IEEE; 2018. Early prediction of chronic kidney disease using machine learning supported by predictive analytics; pp. 1–9.
- [16] Gudeti B., Mishra S., Malik S., Fernandez T.F., Tyagi A.K., Kumari S. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) IEEE; 2020. A novel approach to predict chronic kidney disease using machine learning algorithms; pp. 1630–1635.

- [17] Almasoud M., Ward T.E. Detection of chronic kidney disease using machine learning algorithms with least number of predictors. *Int J Soft Comput Appl.* 2019;10.
- [18] Islam MA, Majumder MZH, Hussein MA. Chronic kidney disease prediction based on machine learning algorithms. *J Pathol Inform.* 2023 Jan 12;14:100189. doi: 10.1016/j.jpi.2023.100189.
- [19] Ayodele Olugbenga E, Alebiosu C Olutayo. Burden of chronic kidney disease: an international perspective. *Adv Chronic Kidney Dis.* 2010;17(3):215–224. Elsevier.
- [20] Molla MD, et al. Assessment of serum electrolytes and kidney function test for screening of chronic kidney disease among Ethiopian Public Health Institute staff members, Addis Ababa, Ethiopia. *BMC Nephrol.* 2020;21(1):494.
- [21] Charleonnann A, Fufaung T, Niyomwong T, Chokchueypattanakit W, Suwannawach S, Ninchawee N. Predictive analytics for chronic kidney disease using machine learning techniques. *Manag Innov Technol Int Conf MITiCON.* 2016;80–83:2017.
- [22] Tekale S, Shingavi P, Wandhekar S, Chatorikar A. Prediction of chronic kidney disease using machine learning algorithm. *Disease.* 2018;7(10):92–6.
- [23] Kumar V. Evaluation of computationally intelligent techniques for breast cancer diagnosis. *Neural Comput Appl.* 2021;33(8):3195–208.
- [24] Oliver T. *Machine Learning For Absolute Beginners.* 2017.
- [25] Ramya S, Radha N. Diagnosis of Chronic Kidney Disease Using. pp. 812–820, 2016.
- [26] Drall S, Drall GS, Singh S. Chronic kidney disease prediction using machine learning: a new approach bharat Bhushan Naib. *Learn.* 2014;8(278):278–87.
- [27] Vijayarani S, Dhayanand S. Data Mining Classification Algorithms for Kidney Disease Prediction. *Int J Cybern Informatics.* 2015;4(4):13–25.
- [28] M. Arora, E.A. Sharma Chronic kidney disease detection by analyzing medical datasets in weka *Int J Comput Mach Learn Algor New Adv Mach Learn.*, 3 (2016), pp. 19-48

Plagiarism Report

ORIGINALITY REPORT			
10%	3%	7%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	Submitted to Liverpool John Moores University Student Paper	1 %	
2	Submitted to Manchester Metropolitan University Student Paper	1 %	
3	Submitted to University of Hertfordshire Student Paper	1 %	
4	journalofbigdata.springeropen.com Internet Source	1 %	
5	N. Sreenivasa, Sudesh Pawaar, Shaurya Sparsh, P. Ramesh Naidu. "Predicting the Kidney Diseases by Using Machine Learning Techniques", ITM Web of Conferences, 2023 Publication	<1 %	
6	"Data Science and Applications", Springer Science and Business Media LLC, 2024 Publication	<1 %	
7	Submitted to The University of the West of Scotland Student Paper	<1 %	