**CHRONIC KIDNEY DISEASE**

A Project Report in partial fulfillment of the degree

# Bachelor of Technology

in

# Computer Science&Engineering

## By

**2203A51655** V.SRESHTA

**2203A51620** T.AISHWARYA

**2203A51647** NAVEENA

**Under the Guidance of**

# D.Ramesh

# Assistant Professor, Department of CSE.

**Submitted to**



# DEPARTMENT OF COMPUTER SCIENCE&ENGINEERING

## SRUNIVERSITY,ANANTHASAGAR,WARANGAL





**DEPARTMENTOFCOMPUTERSCIENCE&ENGINEERING**

**CERTIFICATE**

This is to certify that the Project Report entitled “CHRONIC KIDNEY DISEASE” is a record of Bonafide work carried out by V.Sreshta ,T.Aishwarya ,.Naveena, bearing RollNo (s) 2203A51655, 2203A51620,2103A51647 during the academic year 2022-2023 in partial fulfillment of the award of the degree of Bachelor of Technology in Computer Science Engineering by the SR UNIVERSITY, WARANGAL.

# Supervisor HeadoftheDepartment

Mr. D. Ramesh Dr. M. Sheshikala

Asst. Professor, HOD (CSE)

SR University SR University

**ACKNOWLEDGEMENT**

We express our thanks to course coordinator Mr. D. Ramesh, Asst. prof. for guiding us from the beginning through the end of the course project. We express our gratitude to head of the department CS&AI, Dr. M. Sheshikala, for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean of School of Computer Science and Artificial Intelligence.

Finally, we express our thank to all teaching and non-teaching staff of the department for their suggestions and timely support.

# ABSTRACT

Chronic kidney disease (CKD) is a complex disease which affects approximately 13% of the world’s population. Over time, CKD can cause renal dysfunction and progression to end-stage kidney disease and cardiovascular disease. Complications associated with CKD may contribute to the acceleration of disease progression and the risk of cardiovascular-related morbidities. Early CKD is asymptomatic, and symptoms only present at later stages when complications of the disease arise, such as a decline in kidney function and the presence of other comorbidities associated with the disease. In advanced stages of the disease, when kidney function is significantly impaired, patients can only be treated with dialysis or a transplant. With limited treatment options available, an increasing prevalence of both the elderly population and comorbidities associated with the disease, the prevalence of CKD is set to rise. This review discusses the current challenges and the unmet patient need in CKD.

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Content** | **PageNo** |
| 1 | Introduction | 6 |
| 2 | LiteratureReview | 7 |
| 3 | Design | 8 |
| 4 | Methodology | 9 |
| 5 | DataPre-processing | 11 |
| 6 | Results | 16 |
| 7 | Conclusion | 18 |
| 8 | Future scope | 18 |
| 9 | References | `19 |

**1.INTRODUCTION:**

Chronic kidney disease (CKD) is a complex and multifaceted disease, causing renal dysfunction and progression to end-stage kidney disease (ESKD) and cardiovascular disease. Complications associated with the disease contribute to the acceleration of CKD progression and risk of cardiovascular-related morbidities.Despite its high prevalence and the clinical and economic burden of its associated complications, disease awareness remains profoundly low. Worldwide, only 6% of the general population and 10% of the high‐risk population are aware of their CKD statuses [1]. In addition, CKD recognition in primary care settings is also suboptimal, ranging from 6% to 50%, dependent upon primary care specialty, severity of disease, and experience. Awareness of CKD remains low in part because CKD is usually silent until its late stages. However, diagnosis of CKD during the later stages results in fewer opportunities to prevent adverse outcomes. Physician awareness of CKD is critical for the early implementation of evidence-based therapies that can slow progression of renal dysfunction, prevent metabolic complications, and reduce cardiovascular-related outcomes.Currently CKD is not curable, and management of the disease relies on treatments which prevent CKD progression and cardiovascular disease. Despite available treatments, a residual risk of adverse events and CKD progression remains. This article reviews the challenges associated with CKD and the treatments available for patients, highlighting the unmet need for cardio-renal protection in patients with CKD.

**2.LITERATURE REVIEW:**

This area presents the review of different literatures which was carried out to create an adequate framework for the current study. Some of the latest and significant researches, relevant to the current study have been briefly reviewed in this chapter. The chapter begins with introducing CKD, Further, the machine and the data processing models adopted by this research.**A. Chronic Kidney Disease**s:

The definition and classification of Kidney Disease Outcomes Quality Initiative (K/DOQI) were approved with refinements. Chronic kidney disease is defined as kidney damage or a glomerular filtration rate (GFR) of 60 ml/min/1.73 m2 for at least three months, regardless of the cause, in other words, chronic kidney disease (CKD) means that your kidneys are damaged and cannot filter the blood as it should. This disease is called chronic because kidney damage occurs slowly over a long period of time. This damage can cause waste products in the body. Kidney disease does not develop overnight. It will happen slowly and in stages. Most people in the early stages have no symptoms. They may not know anything is wrong. But if detected and treated, kidney disease can often be slowed or stopped. **B. Machine Learning Application:**Machine learning (ML) is an umbrella term that refers to many algorithms that make intelligent predictions based on a set of data. These datasets are often large, consisting of potentially millions of unique data points. Recent advances in machine learning have achieved human-like semantic understanding and information, and sometimes the ability to perceive abstract patterns more accurately than a human expert. Machine learning is defined as a field of research that gives computers the ability to learn without being specially programmed. Machine learning algorithms are divided into a taxonomy based on the desired result of the algorithm.

# 3.DESIGN:

**Requirement Specifications**

## Hardware Requirements

## System

## RAM

## Hard Disk

## Input

## Output

## Software Requirements

* **OS**
* **Platform**
* **Program Language**

# 4. METHODOLOGY:

After Data pre-processing and data visualization the next step is to apply the models on the dataset. Our dataset comes under supervised learning as it contains the labeled data (target variables, feature variables). First the dataset is splitted into training set and testing set. Then the model is trained on training set and then tested on testing set.

**4.1 Logistic regression algorithm:**

Logistic regression is a machine learning algorithm which comes under supervised learning. It is a parametric method, where an equation is formed to solve. The equation returns continues values. These continues values should to converted to categorical values.so, we use a activation function called “sigmoid”.by using log error function we calculate the error.

* from sklearn.linear\_model import LogisticRegression
* lr=LogisticRegression()
* mm=lr.fit(x\_resem\_train,y\_resem\_train)

**4.2 K-Nearest Neighbor algorithm:**

K-Nearest Neighbor algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression. This algorithm is non parametric. This is also called as lazy learning algorithm. This algorithm works by first selecting the k value which is an integer value and less than the number of rows. When a new data point is given, KNN finds the nearest neighbors to that data point based on the distance using various methods like Euclidean distance or Manhattan distance. And assigns the data point to that class.

* from sklearn.neighbors import KNeighborsClassifier
* classifier=KNeighborsClassifier(n\_neighbors=5,metric='minkowski',p=2)
* classifier.fit(x\_resem\_train,y\_resem\_train

**4.3 Naive Bayes algorithm:**

# Naive Bayes algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression.This algorithm is non parametric. This algorithm works based on the bayes theorem. Naive Bayes algorithm is a probabilistic classifier. It predicts the probability of an object. And also it does not require much training data.

* from sklearn.naive\_bayes import GaussianNB
* gnb=GaussianNB()
* gnb.fit(x\_resem\_train,y\_resem\_train)

# 4.4 Desicion Tree algorithm:

# Decision tree algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression problems. This algorithm is also known as ID3 algorithm. This algorithm is non parametric method. It forms a tree from the given dataset. It has two nodes decision nodes and leaf nodes. Decision nodes are used for taking decisions and leaf nodes are the output of that decisions. The attribute selection happens by entropy and information Gini.

* from sklearn.tree import DecisionTreeClassifier
* classifier=DecisionTreeClassifier(criterion='entropy',random\_state=0)
* mm=classifier.fit(x\_resem\_train,y\_resem\_train)

# 4.5 Support vector machine algorithm:

# Support vector machine algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression problems. SVM works by constructing a hyperplane or a line that separates the different classes of data points. SVM has support vectors. The distance between positive hyperplane and negative hyperplane is called margin.

* from sklearn.svm import SVC
* svm\_model=SVC(kernel='linear')
* svm\_model.fit(x\_resem\_train,y\_resem\_train)

# 5.DATASET PRE-PROCESSING:

# DATASET DESCRIPTION

# Attributes:

# ID

# BP

# AGE

# RBC

# PC

# PCC

# HEMO

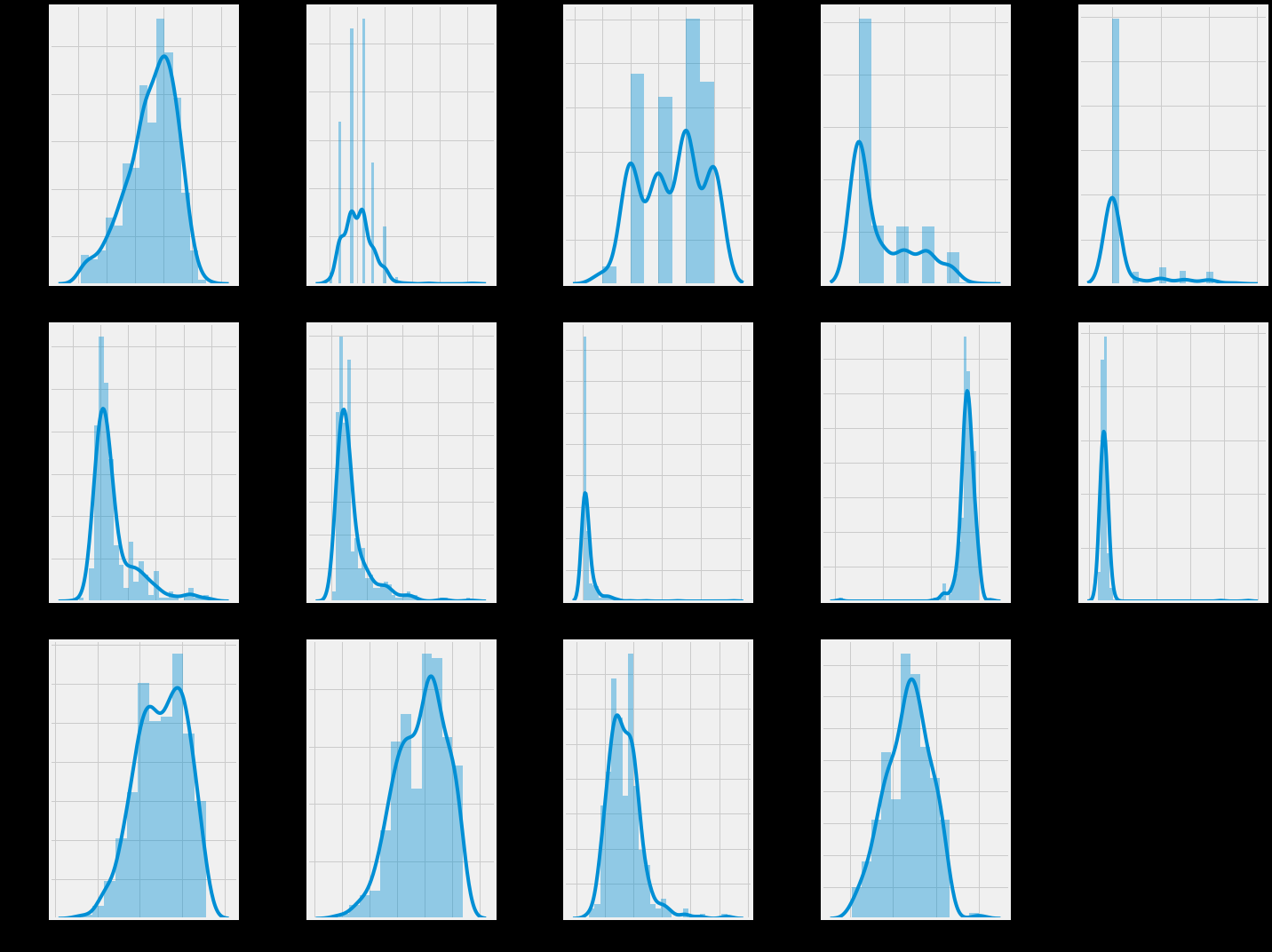
# RC etc.,

**10.DATASET :**



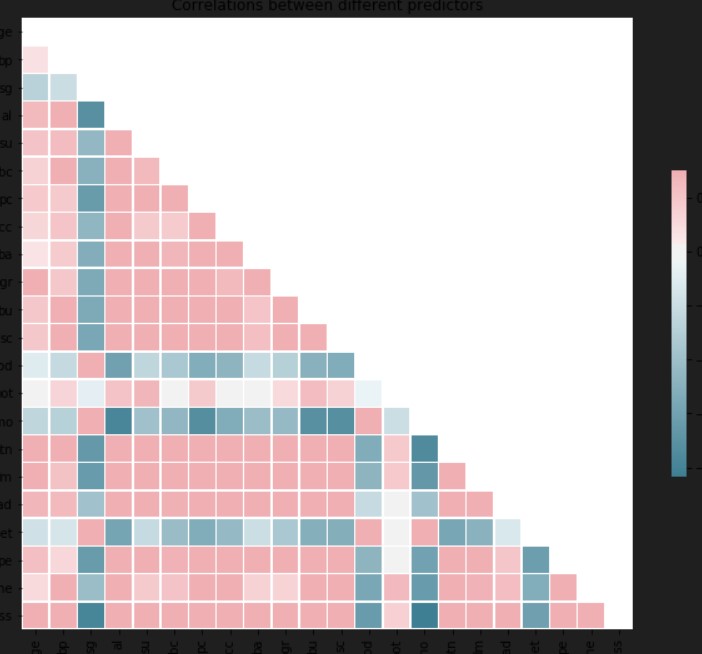


**GRAPHS PLOTTED:**



# The image shows multiple graphs that appear to represent various probability density functions or distributions. Each graph depicts a curve plotted on a grid, with the area under the curve representing the total probability or frequency. Some distributions are symmetric and bell-shaped, while others are skewed or have multiple peaks. Without reproducing any specific material, I can analyze and describe the general characteristics and shapes of these distributions based on the provided visual information.

# 6. RESULTS:



This image appears to be a correlation matrix or heatmap, which is a visual representation of correlations or relationships between different variables or predictors. The matrix is composed of small colored squares, with the colors ranging from pink to dark green or teal.

The diagonal line running from the top-left to the bottom-right shows a solid dark line, likely indicating a perfect correlation of 1 between each variable and itself.

The colored squares surrounding the diagonal represent the correlation coefficients or strengths of the relationships between different pairs of variables. The pink squares indicate positive correlations, with darker shades of pink implying stronger positive correlations. The green or teal squares suggest negative correlations, with darker shades indicating stronger negative correlations.

Correlation matrices like this are commonly used in data analysis and machine learning to identify patterns, relationships, and potential multicollinearity issues among multiple predictors or features in a dataset.

# 

# The image shows a correlation matrix, which displays the pairwise correlation coefficients between different variables or features. The values range from -1 to 1, with 1 representing a perfect positive correlation, -1 representing a perfect negative correlation, and 0 indicating no correlation.

# The diagonal elements all have a value of 1, as each variable is perfectly correlated with Itself.The remaining values in the matrix represent the correlation coefficients between pairs of variables.

# Some observations from the matrix:

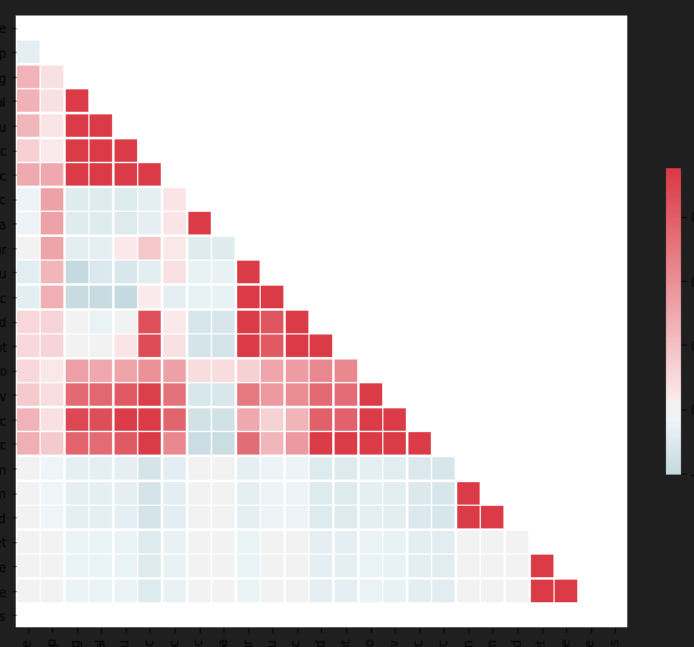
# 1. There are several strong positive correlations, indicated by values close to 1 (e.g., 0.9, 0.79, 0.77).

# 2. There are also strong negative correlations, shown by values close to -1 (e.g., -0.73, -0.69, -0.63).

# 3. Many values are close to 0, suggesting little or no correlation between those pairs of variables.

# 4. The matrix is symmetric about the diagonal, as the correlation between variable A and variable B is the same as the correlation between variable B and variable A.

# Correlation matrices like this are useful for identifying potential multicollinearity issues and understanding the relationships between different features in a dataset, which can be valuable for feature selection and model building.

  
The image appears to be a visual representation of a correlation matrix or heatmap, using

different color shades to depict the strength and direction of correlations between variables or features.

The diagonal line running from the top-left to the bottom-right represents the perfect correlation of 1 between each variable and itself, indicated by the solid red line.

The colored squares surrounding the diagonal represent the correlation coefficients between different pairs of variables. The red squares suggest positive correlations, with darker shades of red implying stronger positive correlations. The blue squares indicate negative correlations, with darker shades of blue representing stronger negative correlations.

The overall pattern in the matrix shows a clustering or block-like structure, with regions of strong positive correlations (dark red squares) and regions of strong negative correlations (dark blue squares). This structure may suggest underlying relationships or groupings among subsets of the variables.

Correlation matrices and heatmaps are useful tools for visualizing and identifying patterns, relationships, and potential multicollinearity issues among multiple variables or features in a dataset, which can inform feature selection, dimensionality reduction, and model building processes.

# 7. CONCLUSION:

Among worldwide, Chronic Kidney Disease (CKD) is a prevalent global health issue affecting approximately 13% of the population. Early detection remains challenging due to its asymptomatic nature in early stages. Risk factors include hypertension, diabetes, and obesity. Diagnostic markers such as estimated glomerular filtration rate (eGFR) guide staging, ranging from mild (Stage 1) to severe (Stage 5). Management strategies focus on slowing progression, controlling blood pressure, and addressing underlying causes. Ongoing research aims to improve risk prediction models and personalized treatment approaches. Collaborative efforts are essential to combat CKD’s impact on public health.

**8. FUTURE SCOPE :**

The future scope of Chronic Kidney Disease (CKD) holds several promising developments. First, precision medicine will play a pivotal role, tailoring treatments based on individual characteristics. Bioartificial kidneys, currently under research, could revolutionize CKD management by providing an alternative to dialysis and transplantation. Telemedicine and remote monitoring will enhance patient care, allowing real-time consultations and adjustments. Improved risk prediction models, driven by machine learning, will identify high-risk patients early. Additionally, regenerative therapies and patient education efforts will contribute to better outcomes. Collaborative efforts across disciplines will be essential in addressing CKD’s global impact.

**9. REFERENCES :**

* National Kidney Foundation. K-DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification and stratification. Am J Kidney Dis. 2002;39(2 Suppl 1):S1-266.
* KDIGO. KDIGO 2012 Clinical practice guideline for the evaluation andmanagement of chronic kidney disease. Kidney Int Suppl. 2013;3(1):1-150.
* Pereira BJ. Optimization of pre-ESRD care: the key to improved dialysisoutcomes. Kidney Int. 2000;57(1):351-65.
* Moyer VA; U.S. Preventive Services Task Force. Screening for chronic kidney disease: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012;157(8):567-70.
* Ginsberg JM, Chang BS, Matarese RA, Garella S. Use of single voidedurine samples to estimate quantitative proteinuria. N Engl J Med 1983;309(25):1543-6.
* Vassalotti JA, Centor R, Turner BJ, Greer RC, Choi M, Sequist TD; National Kidney Foundation Kidney Disease Outcomes Quality Initiative. Practical approach to detection and management of chronic kidney disease for the primary care clinician. Am J Med. 2016;129(2):153-62.
* Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA. 2007;298(17):2038-47.
* Lugon JR. Doença renal crônica no Brasil: um problema de saúde pública.J Bras Nefrol. 2009;31(supl 1):2-5.
* Marinho AWGB, Penha AP, Silva MT, Galvão TF. Prevalência de doençarenal crônica em adultos no Brasil: revisão sistemática da literatura. CadSaúde Colet. 2017;25(3):379-88.
* Thomé FS, Sesso RC, Lopes AA, Lugon JR, Martins CT. Brazilian chronicdialysis survey 2017. J Bras Nefrol. 2019;41(2):208-14.
* Thomas B, Matsushita K, Abate KH, Al-Aly Z, Ärnlöv J, Asayama K, et al;Global Burden of Disease 2013 GFR Collaborators; CKD Prognosis Consortium; Global Burden of Disease Genitourinary Expert Group. Globalcardiovascular and renal outcomes of reduced GFR. J Am Soc Nephrol. 2017;28(7):2167-79.
* Moyer VA, U.S. Preventive Services Task Force. Screening for chronic kidney disease: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012;157(8):567-70.