Scilab case study project on Scilab

Early Detection of Chronic Kidney Disease Using an Interpretable AI Model in Scilab

Elakhya S G

Department of Artificial Intelligence and Data Science

Rajalakshmi Institute of Technology, Chennai, India

Artificial Intelligence and Machine Learning

13 April 2026

Abstract

Chronic Kidney Disease (CKD) is a serious and progressive medical condition that often remains undetected during its early stages, leading to severe complications and increased mortality rates. The primary purpose of this case study is to develop a simple, cost-effective, and interpretable Artificial Intelligence (AI) model for early detection of CKD using Scilab. The study focuses on designing a transparent and efficient prediction system that can be easily understood and applied in real-world healthcare scenarios.

The design of the study follows a structured data analysis pipeline consisting of data preprocessing, feature selection, model development, and performance evaluation. A publicly available clinical dataset is utilized, containing multiple attributes related to kidney health. Data preprocessing techniques such as handling missing values, normalization, and encoding are applied to improve the dataset quality and consistency.

The methodology emphasizes selecting a minimal set of significant features that strongly influence CKD prediction. Specifically, hemoglobin level, specific gravity, and hypertension are chosen to build the classification model. A suitable classification algorithm is

implemented in Scilab to categorize patients as CKD or non-CKD. The model's performance is evaluated using accuracy measures and confusion matrix analysis.

The approach prioritizes interpretability by reducing model complexity and incorporating graphical representations to explain feature contributions. This improves transparency and supports better clinical decision-making. Overall, the study demonstrates that a simplified and explainable AI model can effectively assist in early CKD detection, especially in resource-limited healthcare environments.

1. Introduction

Chronic Kidney Disease (CKD) is a long-term medical condition in which the kidneys gradually lose their ability to filter waste and excess fluids from the blood. If not detected early, CKD can progress to kidney failure, requiring dialysis or transplantation. Early diagnosis is therefore essential to slow disease progression and reduce healthcare costs. However, CKD is often referred to as a "silent disease" because it shows minimal or no symptoms in its early stages.

With the rapid advancement of Artificial Intelligence and Machine Learning, data-driven approaches are increasingly being used in medical diagnosis. These models can analyze large volumes of patient data and identify patterns that may not be visible through traditional methods. However, many existing machine learning models are complex and function as "black boxes," making them difficult for healthcare professionals to interpret and trust.

This project aims to address this issue by developing a simple and interpretable machine learning model using Scilab. The focus is on achieving early detection of CKD using a minimal set of clinical features while maintaining transparency in the decision-making process. The proposed approach ensures that the model is not only accurate but also understandable and practical for real-world implementation.

2. Problem Statement

Chronic Kidney Disease is difficult to detect at an early stage due to the absence of noticeable symptoms and the reliance on multiple medical tests. In many cases, diagnosis occurs only after significant kidney damage has already taken place. While machine learning models have shown promise in improving diagnostic accuracy, many of these models are complex and lack interpretability, which limits their adoption in clinical practice.

The problem addressed in this case study is to develop a low-cost, accurate, and interpretable model for early detection of CKD using Scilab. The solution focuses on simplifying the prediction process without compromising reliability. This is achieved through systematic data preprocessing, selection of the most relevant features, and implementation of a classification model that provides clear and understandable results.

The proposed method involves cleaning and preparing medical data, identifying key features that strongly influence CKD, building a classification model, and presenting results through graphical interpretations. The overall objective is to reduce computational complexity while ensuring that the model remains effective, transparent, and suitable for practical healthcare applications.

3.Basic concepts related to the topic

3.1 Machine Learning

Machine Learning is a subset of Artificial Intelligence that enables systems to learn patterns from data and make predictions without being explicitly programmed. In this project, machine learning techniques are used to analyze patient data and classify individuals as CKD or non-CKD based on selected features.

3.2 Feature Selection

Feature selection is the process of identifying the most relevant input variables that contribute significantly to the prediction outcome. Reducing the number of features helps in simplifying the model, improving interpretability, and reducing computational cost..

The correlation between variables is measured using the formula:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

This equation helps in determining the strength of the relationship between features and the target variable.

3.3 Classification

Classification is a supervised learning technique used to categorize data into predefined classes. In this case study, the model classifies patients into two categories:

- CKD (1)
- Non-CKD (0)

The classification model learns from labeled data and predicts the class of new input samples.

3.4 Explainable AI

Explainable Artificial Intelligence (XAI) refers to techniques that make machine learning models transparent and easy to interpret by humans. In critical domains such as healthcare, understanding how a model arrives at a decision is as important as the accuracy of the prediction itself. Black-box models often fail to provide this clarity, which reduces trust among medical professionals.

In this case study, interpretability is achieved by selecting a minimal number of highly relevant features and visualizing their impact using graphs. By analyzing how hemoglobin levels and specific gravity influence CKD prediction, the model provides clear insights into the reasoning behind each classification. This approach ensures that the system remains both reliable and clinically meaningful.

4. Flowchart

The flowchart represents a step-by-step workflow of the CKD prediction system. The process begins with data collection from a clinical dataset, followed by preprocessing steps such as handling missing values and normalizing the data. Once the data is cleaned, relevant features are selected based on their importance in CKD prediction.

The selected features are then passed into a classification model implemented in Scilab. The model analyzes the input values and categorizes each patient as either CKD or non-CKD.

Finally, the results are visualized using graphs and evaluated using performance metrics such as the confusion matrix.

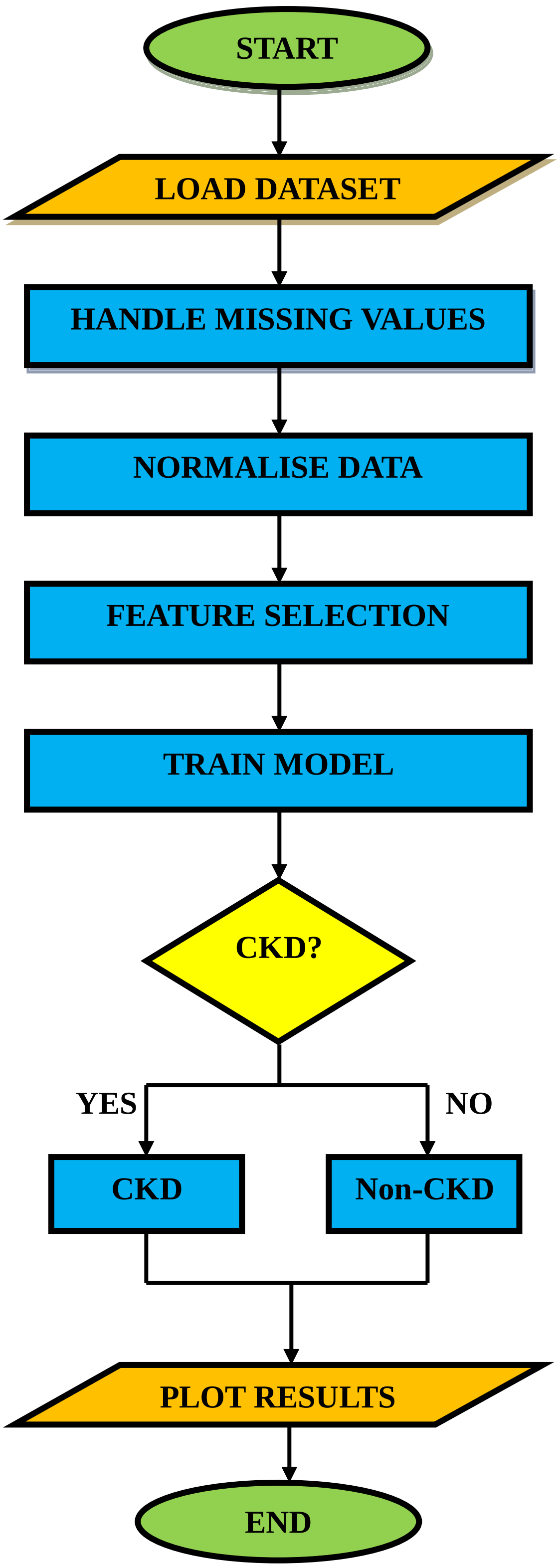

Figure 1: Flowchart

5. Software/Hardware used

The implementation of this case study is carried out using Scilab, an open-source numerical computation software that provides a powerful environment for data analysis and algorithm development. The system is developed and tested on a Windows-based platform to ensure compatibility and ease of use.

The details of the software and hardware used are as follows:

- **Operating System:** Windows 10 (64-bit)
- **Software:** Scilab version 6.1.0
- **Hardware Requirements:** Personal computer with a minimum of 8 GB RAM and standard processor

Scilab is chosen due to its flexibility, ease of implementation, and strong support for mathematical computations and graphical visualization.

6. Procedure of execution

1. Install Scilab software.
2. Place dataset file (ckd.csv) in working directory.
3. Open Scilab and set directory.
4. Create main.sce file.
5. Load dataset using:

```
data = csvRead('ckd.csv');
```

6. Handle missing values using mean/mode.
7. Normalize numerical values.
8. Select features: hemo, sg, htn.
9. Train classification model.
10. Predict outputs.
11. Plot results using graph functions.

7. Result

The developed Scilab-based model successfully predicts the presence of Chronic Kidney Disease (CKD) using a minimal set of selected features, namely hemoglobin level, specific gravity, and hypertension. The results are analyzed using graphical representations and classification performance metrics to ensure both accuracy and interpretability.

Graph 1: Hemoglobin vs CKD

The first graph illustrates the relationship between hemoglobin levels and CKD classification. It is observed that patients diagnosed with CKD (represented by ‘+’ markers) generally have lower hemoglobin levels, mostly concentrated in the range of approximately 6 to 12 g/dL. In contrast, non-CKD patients (represented by diamond markers) are clustered at higher hemoglobin values, typically between 13 and 17 g/dL. This clear separation indicates that hemoglobin is a strong indicator for CKD detection. Lower hemoglobin levels are associated with impaired kidney function, making this feature highly significant for prediction.

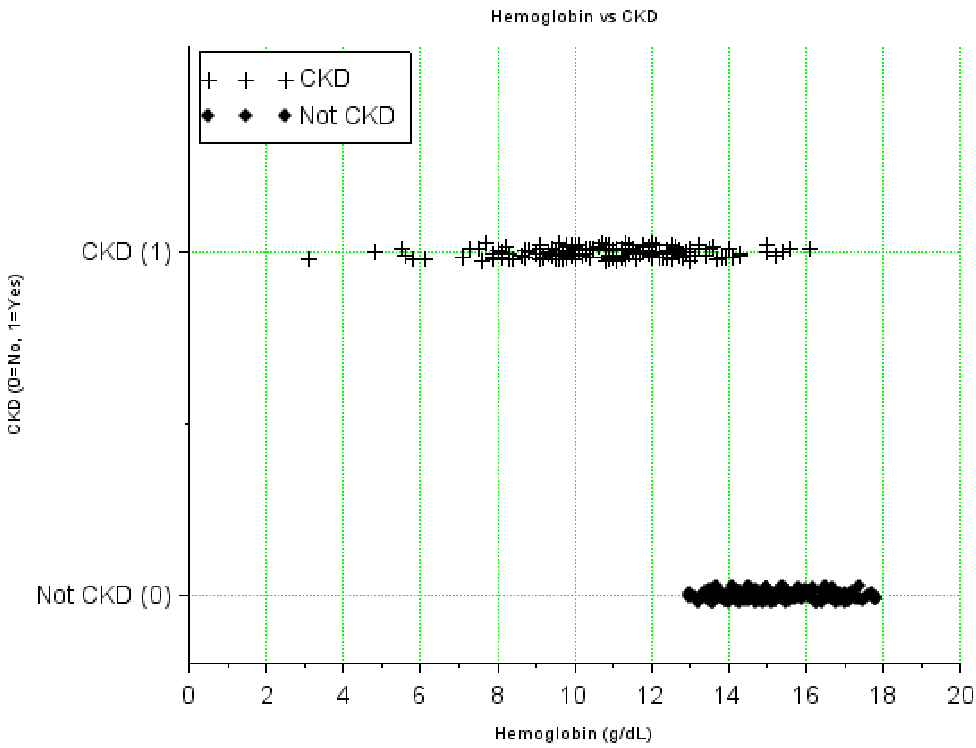

Figure 2: Hemoglobin Vs CKD

Graph 2: Specific Gravity vs CKD

The second graph shows the distribution of specific gravity values for CKD and non-CKD patients. CKD cases are spread across lower specific gravity values (around 1.005 to 1.015), whereas non-CKD cases are concentrated at higher values (around 1.020 to 1.025). This pattern suggests that reduced specific gravity is associated with kidney dysfunction, as the kidneys lose their ability to concentrate urine effectively. The visual separation between the two classes further confirms the importance of this feature in classification.

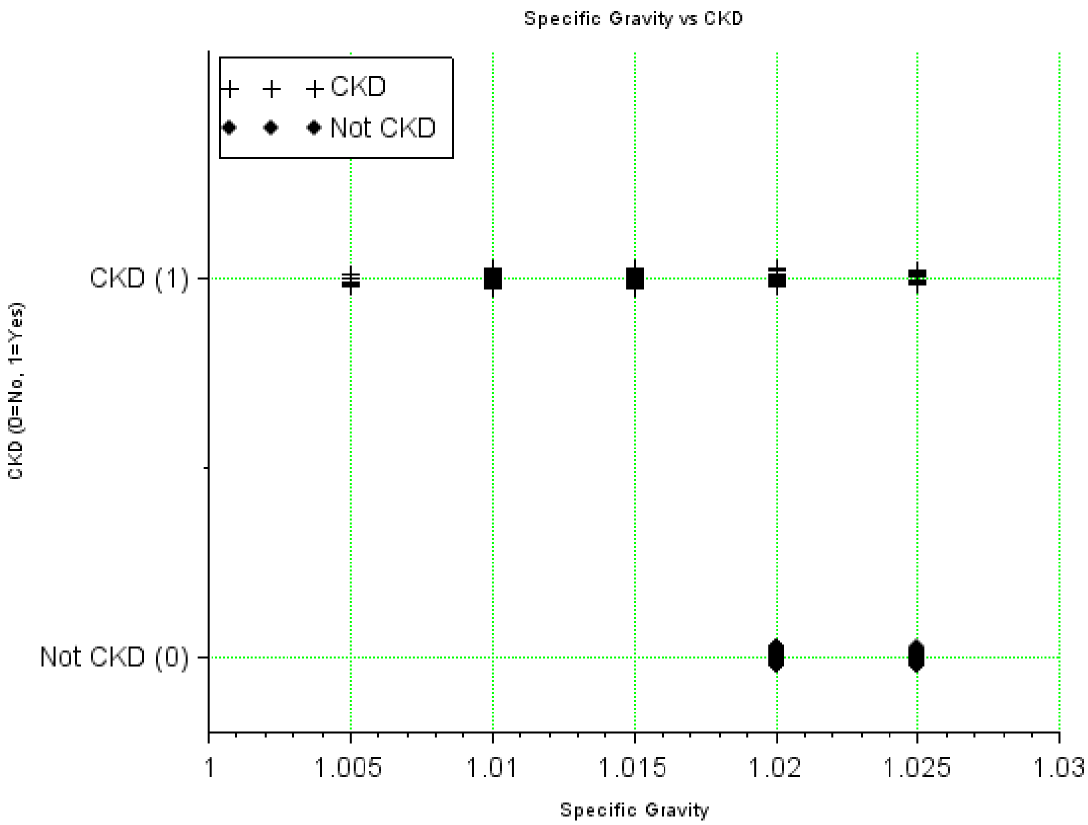

Figure 3: Specific Gravity Vs CKD

Confusion Matrix Analysis

The performance of the model is evaluated using a confusion matrix, which includes True Positives (correctly identified CKD cases), True Negatives (correctly identified non-CKD cases), False Positives, and False Negatives. The results indicate that the model achieves high classification accuracy with minimal misclassification, demonstrating its effectiveness despite using only a limited number of features.

Inference

From the results, it can be inferred that reducing the number of input features does not significantly affect the prediction performance when the selected features are highly relevant. Hemoglobin and specific gravity play a crucial role in distinguishing CKD from non-CKD cases. The graphical representation enhances interpretability by clearly showing how feature values influence the classification outcome. Overall, the model achieves a balance between simplicity, accuracy, and explainability, making it suitable for practical healthcare applications, particularly in resource-constrained environments.

8. References

- P. A. Moreno-Sánchez, "Data-Driven Early Diagnosis of Chronic Kidney Disease: Development and Evaluation of an Explainable AI Model," *IEEE Access*, vol. 11, pp. 38359–38367, 2023.
- D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2017. [Online]. Available: http://archive.ics.uci.edu/ml
- A. C. Webster, E. V. Nagler, R. L. Morton, and P. Masson, "Chronic Kidney Disease," *The Lancet*, vol. 389, no. 10075, pp. 1238–1252, 2017.
- G. Stiglic et al., "Interpretability of Machine Learning-Based Prediction Models in Healthcare," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 5, 2020.
- Scilab Enterprises, "Scilab: Free and Open Source Software for Numerical Computation," [Online]. Available: https://www.scilab.org