

# Chronic Kidney Disease Stage Identification In HIV Infected Patients Using Machine Learning

Adeeba Seema<sup>1</sup>, Dr. Laxmi Math<sup>2</sup>

<sup>1</sup>Student, Department of Artificial Intelligence and Data Science, Faculty of Engineering and Technology, Sharnbasva University, Kalaburagi, India.

<sup>2</sup>Assistant Professor, Department of Artificial Intelligence and Data Science, Faculty of Engineering and Technology, Sharnbasva University, Kalaburagi, India.

## ABSTRACT

Chronic Kidney Disease (CKD) is a chronic and progressive medical condition that is often not associated until it is in its advanced stages, which consequently leads to extreme morbidity and increased mortality. The condition aggravates this risk in individuals who have contracted Human Immunodeficiency Virus (HIV) because of the long-term use of antiretroviral drugs, persistent systemic inflammation, and co morbid metabolic disturbance. In turn, timely diagnosis of CKD and accurate staging of the disease is of utmost importance in order to support the process of clinical intervention. The current study suggests a machine-learning-based decision support system in identifying CKD stages in HIV-infected patients using routine clinical and laboratory parameters. The system uses the supervised learning algorithm such as Support Vector Machine, Random Forest, Decision Tree, and XGBoost to classify the renal status as either normal or abnormal. Meanwhile, the estimated Glomerular Filtration Rate (eGFR) is determined by CKD-EPI equation and it is further matched with standard CKD levels, that is, Stage 1 to Stage 5. Experimental evaluation shows that models built in the form of an ensemble achieve better predictive accuracy with the Random Forest algorithm having the highest accuracy, precision and recall rates compared to the other analyzed models. The proposed approach will increase the reliability and interpretability of diagnostics by integrating data-driven machine -learning predictions with clinically validated eGFR-based staging. The resulting system is likely to provide healthcare professionals with an opportunity to screen CKD in its early stages, to reduce the diagnostic time, and to facilitate evidence-based decisions, especially in relation to the high-risk group of HIV-positive patients.

**Keywords:** Chronic Kidney Disease, Machine Learning, eGFR, CKD Stage Classification, Clinical Decision Support System.

## I. INTRODUCTION

Chronic Kidney Disease (CKD) is a serious and long-lasting health condition where the renal function gradually decreases and is experienced by millions of people in the world. It has been identified as a serious community health issue due to its great prevalence, incurable course, and strong correlation with heart-related complications and deaths. Unattended or left latent, CKD can eventually develop into End stage Renal Disease (ESRD), requiring long-term dialysis or kidney transplant. The major issue with the management of CKD is that it remains silent and only has limited perceptible symptoms at its early stages; as such, most patients are only diagnosed at a point when significant levels of renal damage has already been done to limit therapeutic choices and undermine patient outcomes. The weight of CKD is significantly increased in people who are infected with HIV. The presence of long-term antiretroviral therapy, chronic immune activation, opportunistic infections, and metabolic imbalance, in turn, increases the risk of renal impairment among this demographic. Although there is a general clinical monitoring, kidney failure in HIV-infected individuals may go unnoticed until progression leading to the need to put in place efficient and timely screening products. The growing number of electronic health records and large volumes of clinical data has enabled the appearance of machine-learning (ML) methods as potential tools to identify non-trivial patterns in medical records and aid in the early detection of disease. The proposed research thus focuses on the application of ML models together with clinically validated eGFR-based staging to improve the CKD stage identification among HIV-infected patients.

### Chronic Kidney Disease (CKD)

Chronic Kidney Disease refers to the long term disease in the structure or functioning of the kidneys which continues to affect the overall wellbeing in a period of more than three months. The condition often co-exists with diabetes mellitus, hypertension, cardiovascular disease, and chronic infections including HIV. Clinically, CKD is evaluated through estimated eGFR, which demonstrates the ability of the kidney to remove waste materials in the blood, and biomarkers of kidney injury including albuminuria. According to eGFR results, CKD is divided into five levels, the first being Stage 1, which means the normal or slightly impaired kidney functioning, and the last level in Stage 5, denoting the renal failure. With the advancement of CKD, the patients are faced with increased risk of anaemia, electrolyte imbalance, bone diseases, cardiovascular diseases, and reduced quality of life. It is thus necessary to detect diseases early and stage them properly to slow down the process and enhance the outcome in the long run.

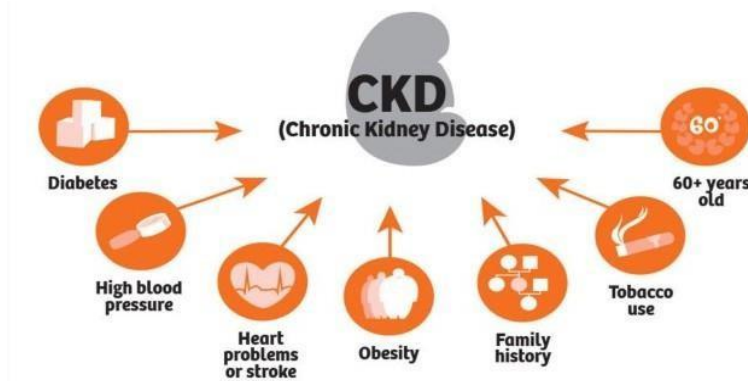


Fig. 1 Demonstrates the etiological factors of CKD development that are dominant.

### Difficulties in Diagnosis of CKD at an Early Stage

The clinical issue of early CKD detection is a challenging one, especially in the high-risk patient groups, including HIV-infected individuals. During the early stages, renal damage progresses in an insidious fashion and does not produce noticeable clinical features making a routine screening tedious. The traditional methodology of diagnosis is highly dependent on laboratory values of serum creatinine and blood urea. Even though these indicators are clinically relevant, they might be insensitive to detect early functional impairment when assessed individually. Additionally, proper diagnosis determines the concomitant determination of several clinical and biochemical variables to increase the complexity of decision-making. Manual interpretation of laboratory reports can be time consuming and prone to human error in the resource-limited healthcare environment or where there is high patient turnover. Differences in patient demographics, comorbidities like diabetes and hypertension, and incongruent clinical measurements also make the diagnosis more challenging. Such restrictions often lead to delayed intervention especially in HIV-infected people whereby the renal damage could have been fastened by long durations of medication exposure. As a result, there is an increased need of automated, data-based systems with the ability of analysing multidimensional clinical data and provide early warnings of CKD development.

### Contributions and Motivation of the Proposed Work

The motivation of the paper is the development of a smart and clinically interpretable system of identifying the stage of CKD among HIV-positive individuals through a combination of machine-learning algorithms and medical principles. Although many studies that are done to date indicate high predictive accuracy in terms of ML models, many of them lack the inclusion of standardised disease staging, which limits their practical application in clinical practice. This contribution fills this gap by integrating machine-learning-based CKD classification with eGFR-based stage classification, thereby guaranteeing both predictive and clinical means.

The main findings of this study are as follows:

1. Training and analysis of numerous trained machine-learning algorithms to predict CKD in HIV-infected individuals.
2. The most effective model has been identified automatically by the measures of experimental performance.
3. Introduction of eGFR values by CKD-EPI equation.
4. Mapping of eGFR values to standard CKD stages between Stage 1 and Stage 5.
5. Introduction of a web-based clinical decision-support tool that provides health care personnel with real-time CKD status, eGFR value and disease stage.

## II. LITERATURE REVIEW

Dritsas et al. [1] have carried out a comprehensive survey of the machine learning techniques that are relevant to the early detection of CKD and assessed algorithms like support vector machines (SVM), Random Forests (RF), XGBoost, and neural network architectures. This research highlighted typical barriers, such as the lack of clinical values, laboratory measurements heterogeneity, and atypical features descriptions. They also discovered that ensemble and tree-based models were always better than a single classifier and thus indicated that care should be taken in pre-processing data and in selecting features.

Ghosh et al. [2] proposed an ensemble learning framework that could predict CKD progression and stratify the disease stages based on multisource clinical data. Their empirical evidence proved to have better predictive capability and strength as compared to stand alone models. The writers also used explainability methods in identifying clinically meaningful predictors, which is why interpretable machine learning constructs are needed in healthcare decision-support systems.

Chen et al. [3] conducted a comparative evaluation of the classical and boosted or gradient approaches to the classification of CKD. Their experiments showed that XGBoost obtained better precisionrecall equilibrium especially in the early-stage CKD. The authors suggested that operationalisation of machine learning systems in decision making in medicine requires threshold optimisation and compatibility with clinical context as a prerequisite.

Proper estimation of renal functioning is still invaluable in CKD diagnosis and staging. Inker et al. [4] gave revised creatinine-based CKD-EPI equations and cystatin C-based equations that improve the accuracy of eGFR estimation but eliminate race-based adjustments. These equations have been widely adopted in clinical settings and they form the foundation of machine learning predictions combined with rule-based CKD staging.

Halder et al. [5] were interested in explainable machine learning to predict CKD, showing that SHAP-based interpretations increase clinician trust and transparency. Their work shed light on the fact that the predictive accuracy and interpretability are crucial to the actual clinical adoption.

Raihan et al. [6] introduced a hybrid pipeline of domain specific feature engineering and machine learning suite of CKD detection that combines clinical indicators with automated feature selection. Their results showed that domain-sensitive preprocessing significantly enhances early-stage CKD detection, which supports the introduction of medical information into machine learning systems.

Arif et al. [7] explored the CKD prediction through ensemble in low-resource healthcare settings. According to their findings, Random Forest and decision tree models offer a practical compromise between performance and computational efficiency, and hence are applicable to be deployed in real world clinical context.

Deb et al. [8] tested the cross-cohort generalisation of CKD prediction models and found considerable deterioration in the performance of the model when the preprocessing steps were not consistent. The paper highlighted the need to adopt standardised data handling and validation plans in using machine learning in various patient groups.

Praveen et al. [9] compared the methods of gradient-boosting to forecast clinical risk, specifically CKD, based on XGBoost and LightGBM. Their analysis showed that boosted models are well-suited to nonlinear relations between lab variables, unless the issue of class imbalance and hyperparameter fine-tuning are properly addressed.

Patel et al. [10] provides an extensive outline of the applications of machine learning in the nephrology field and analyzes opportunities, limitations, and challenges related to integration with clinical work. The paper has noted the need to establish systems to integrate predictive modeling with previously set medical rules notably in chronic illnesses which require longitudinal surveillance.

## III. SYSTEM ANALYSIS

### 1. Problem Statement

Chronic kidney disease is an irreversible, progressive disease which is often asymptomatic in its initial stages. Lack of overt clinical presentations tends to result in diagnoses made when severe renal impairment has already set in then therapeutic interventions would be limited. Late diagnosis of CKD is associated with grave after-effects such as cardiovascular disease, electrolyte imbalance, anaemia, and increased mortality. This issue is particularly acute in high-risk groups, including HIV-infected citizens, who demonstrate more susceptibility to renal dysfunction due to chronic infection, long-term exposure to pharmacologic agents, and comorbidities. Traditionally, CKD diagnosis is also based significantly on the series of laboratory examinations and the manual analysis of the clinical reports by medical experts. Although these methodologies are clinically approved, they are labour intensive, resource consuming and also dependent on the availability of specialists, which might be minimal in specific health care

facilities. Furthermore, the independent examination of the single laboratory parameters often does not offer a complex evaluation of renal health. Considering the growing amount of clinical data, there is an undisputable demand of automated and intelligent decision-support systems that can help clinicians early detect and stage CKD. A promising way of overcoming these limitations is through machine learning methods, which are synergised with proven clinical measurements, like eGFR.

## 2. Objectives of the Study

The overall intention of the research is to produce a smart and automated framework of detecting and diagnosing chronic kidney disease at an early stage in HIV-infected individuals, through machine learning techniques. The most important aspect of this high-risk population is to detect the disease at an early stage to prevent disease progression and improve the clinical outcomes in the long-term.

The specific objectives are: Purpose to develop a machine-learned predictive framework with a classification of CKD status (Normal or Abnormal) based on routinely collected clinical and laboratory data.

- To estimate the glomerular filtration rate (eGFR) of each patient using known medical equations and find out the corresponding CKD stage.
- To train and test various machine-learned models and compare the performance of the models on clinical data.
- To reduce the use of manual diagnosis and expert interpretation, providing an automated clinical decision-support system.
- To help healthcare professionals recognize high-risk HIV-infected patients in the primary phase, and thus be able to provide medical care within a short time and reduce risks of renal failure.

## Dataset Description

The dataset that is used in this study is a sample of the clinical records of HIV-positive patients obtained and summarises the demographic variables in addition to the laboratory investigation results. It is a realistic clinical situation and thus a good model in development and validation of machine-learning models that can predict and stage kidney diseases. The corpus has around 400 patient records, each of which is described by between 18 and 25 attributes, depending on the preprocessing steps and feature selection strategies.

The important numerical characteristics include age, blood pressure, specific gravity, albumin concentration, glucose level, serum creatinine, blood urea, sodium, potassium, hemoglobin and white-blood-cell count. Categorical variables refer to clinical conditions and comorbidities like diabetes mellitus, hypertension, coronary artery disease, appetite status and pedal edema. These determinants are clinically important as they have a direct effect on renal activity and chronic kidney disease (CKD) development. The outcome variable is binary, and the cases will be categorized as CKD (Abnormal) or Non-CKD (Normal), which will be easily supervised. The data attributes on which the experiments were done are as follows based on Table 1.

**Table I. HIV-CKD Patient Dataset Attributes.**

| Sr no. | Data                    | Type        |
|--------|-------------------------|-------------|
| 1      | Gender                  | Categorical |
| 2      | Diabetes Mellitus       | Categorical |
| 3      | Coronary Artery Disease | Categorical |
| 4      | Pedal Edema             | Categorical |
| 5      | Appetite                | Categorical |
| 6      | Age                     | Numerical   |
| 7      | Ethnicity               | Numerical   |
| 8      | Hypertension            | Numerical   |
| 9      | Blood Pressure          | Numerical   |
| 10     | Specific Gravity        | Numerical   |
| 11     | Albumin                 | Numerical   |
| 12     | Sugar                   | Numerical   |
| 13     | Red Blood Cells         | Numerical   |
| 14     | Pus Cell                | Numerical   |
| 15     | Bacteria                | Numerical   |
| 16     | Blood Glucose Random    | Numerical   |

|    |                        |           |
|----|------------------------|-----------|
| 17 | Pus Cell clumps        | Numerical |
| 18 | Blood Urea             | Numerical |
| 19 | serum creatinine       | Numerical |
| 20 | Sodium                 | Numerical |
| 21 | Potassium              | Numerical |
| 22 | Hemoglobin             | Numerical |
| 23 | Packed Cell Volume     | Numerical |
| 24 | White Blood Cell Count | Numerical |
| 25 | Red Blood Cell Count   | Numerical |

### 3. Existing System

The diagnosis of CKD in clinical practice today is largely based on the manual evaluation of the laboratory reports and clinical findings by medical workers. To examine the renal functioning physicians examine the individual test results such as serum creatinine, blood urea and urinalysis. Although the estimated eGFR is critical measure of renal well-being, it is usually calculated separately and is not included in automated predictive systems. This traditional method does not incorporate a unified analytical system, which is capable of integrating a series of clinical variables to provide a comprehensive evaluation of the risk of CKD. As a result, the detection of CKD is not proactive and most of the cases are diagnosed when significant damage has already occurred in the kidney. There is also a high risk of late diagnosis due to the limited availability of automated screening tools especially in patients who are infected with HIV where renal impairment can creep in unnoticed.

### 4. Proposed System

The proposed system offers a smart and computerized system of early identification and stage detection of CKD in HIV-infected individuals through the combination of machine-learning frameworks with clinical laboratory information. Unlike the traditional ones, the system analyzes many health indicators of a patient simultaneously to provide timely and accurate predictions. A carefully selected data set of demographic data, laboratory data, and comorbidity data is pre-processed and used to learn a number of machine learning packages such as Support Vector Machine, Random Forest, Decision Tree, and XGBoost. The trained models are accessed against the standard performance measures, and the most effective model is automatically chosen to be used in final prediction. The system takes a calculated estimation of the GFR through the CKD-EPI equation and maps the obtained figure to predetermined CKD stages in accordance with the current clinical guidelines. Combining machine-learning-based prediction and clinically validated eGFR estimation, the suggested system will provide an extensive and interpretable indication of renal health. The methodology is useful in supporting early identification, facilitating the decision-making process within the clinical setting, and supporting preventive care among HIV-positive patients who are at risk of CKD.

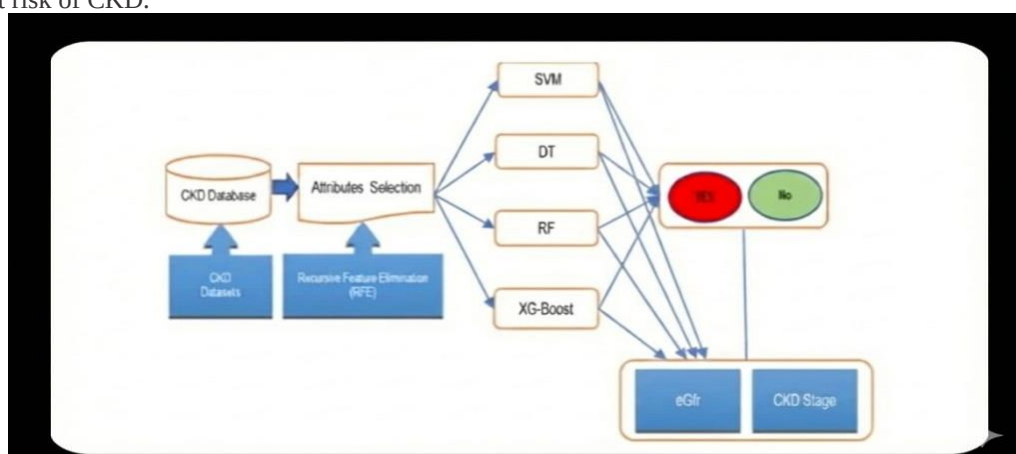


Fig. 2 Block diagram of the proposed system

## IV. METHODOLOGY

**A. Data Preprocessing:** Preprocessing of data is an important process of building a trustworthy machine-learning model especially when the data are clinical and often limited by missing data, inconsistencies and presence of different data types. The raw data obtained in patients who have HIV infection include demographic information, laboratory results, and clinical measurements that are related to the performance of renal functions. There are a number of preprocessing steps that are used to improve the quality of the data and the model performance.

**Handling Missing Values:** The absence of values in vital clinical parameters including blood pressure, serum creatinine, albumin, and blood glucose is dealt with through statistical imputation measures. Mean or median values of the data distribution are used to mitigate the numerical features, and the most common category is used to impute the categorical attributes. This plan maintains the data integrity and removes the chances of warping the underlying patterns.

**Encoding of Categorical Attributes:** There are several categorical variables in the dataset, such as sex, red-blood-cell status, pus-cell presence, bacteria presence, diabetes mellitus and hypertension. These characteristics are coded into numbers using label encoding; e.g., sex (male (1) and female (0)) and binary clinical variables normal and abnormal (0 and 1). This conversion helps machine learning algorithms to effectively process categorical information.

**Feature Scaling and Normalization:** To obtain uniformity in the features and avoid bias from having disparate scales in the measure, numerical attributes undergo normalization using standard scaling techniques. The distance-based algorithms are especially affected by feature scaling, as in the case of Support Vector Machine, unscaled features can cause undue influence on the behaviour of the model.

**Removal of Noise and Inconsistency:** Preprocessing identifies outliers, duplicate records, and records that are logically inconsistent, and is used to eliminate them. Elimination of noisy data enhances the reliability of data sets and advances the learning ability of machine-learning models hence providing more reliable and correct predictions.

## B. Proposed Methodology

The offered methodology is a combination of machine-learning-based classification with clinically validated rule-based assessment of CKD staging among HIV-infected individuals. The general workflow is made up of the following consecutive steps:

**Data Input:** As input to the system, patient demographic, and clinical measures are provided. These are laboratory findings, physiological signs, and comorbidity signs that are applicable to the renal health of HIV-infected patients.

**Data Preprocessing:** The input data are cleaned, encoded, normalized and validated in order to be compatible with the trained machine-learning models and to be clinically consistent.

**Model Training:** Multiple supervised machine learning algorithms, namely Support Vector Machine, Random Forest, Decision Tree and XGBoost are trained with the pre-processed dataset. Each model learns discriminative patterns that separate normal kidney function from abnormal conditions that occur in patients infected with HIV.

**Prediction and Model Choosing:** For new patient data, the trained models are used to generate predictions. Using predefined evaluation metrics such as accuracy, precision and recall, based on these metrics, the system internally selects the best performing model and outputs the final CKD classification as either Normal or Abnormal. No manual model selection is required from the end user.

**Computation of eGFR and CKD Staging:** To improve clinical interpretability, the estimated Glomerular Filtration Rate (eGFR) is calculated using the CKD epi equation. The resulting eGFR value is then mapped to the internationally recognised CKD stages which varies from Stage 1 to Stage 5 as per the established medical guidelines.

## C. Machine Learning Models Used

In order to attain credible detection of chronic kidney disease (CKD) stages in patients with HIV, a number of machine learning models were implemented to elicit complex clinical trends. The Support Vector Machine (SVM) has been chosen due to its efficiency in high-dimensional data classification, and the feature normalization has been used to improve the results. Ensemble based approaches such as the Random Forest have been utilized due to their resilience in face of noise and overfitting reduction due to the presence of several decision trees. The use of the Decision Tree models was due to their readability and ability to represent the consistent rules thus making them easier to understand by the clinicians of the predictions. XGBoost was as well taken advantage of its high capacity to depict nonlinear association and feature interactions using gradient boosting. In order to minimize the chances of user bias, the system can automatically pick the highest performing model when training on evaluation metrics, thus able to make accurate and reliable predictions when deployed.

#### D. Training and Evaluation of the Model

##### Train–Test Partitioning:

The data set is split into training and testing data sets using a traditional train-testing method. The training set is used to learn the model parameters while the testing set is used to assess the generalisation performance of the model on unseen data.

**Evaluation Metrics for Performance:** Model performance is evaluated using several evaluation metrics to obtain a complete analysis:

- Precision: Refers to the reliability of positive CKD predictions.
- Recall: Reflects the capability of the model to identify the correct cases of CKD, which is important for early diagnosis in HIV-infected individuals.
- Confusion Matrix: Confusion matrix gives the detailed break down of true positive, true negative, false positive and false negative gives more insight about classification behavior.

#### E. eGFR Estimation and Staging CKD

Estimated Glomerular Filtration Rate is also a clinical accepted indicator for determining kidney function and severity of CKD. In the proposed system, eGFR is calculated by the CKD-EPI equation, which is commonly used because of its enhanced accuracy in different age groups and levels of renal function.

##### CKD-EPI Equation:

The CKD-EPI equation estimates kidney filtration capacity from the serum creatinine concentration, age and sex of the individual. Serum creatinine indicates the efficiency of kidneys whereas age and sex explain the physiological differences in creatinine formation.

##### Input Parameters:

- Patient age
- Sex (male or female)
- Serum creatinine level (mg/dL)

The resulting eGFR value is an expression of mL/min/1.73 m<sup>2</sup>.

##### CKD Staging Based on eGFR:

The eGFR calculated is mapped to standard CKD as follows:

- Stage 1: eGFR  $\geq 90$
- Stage 2: eGFR 60–89
- Stage 3: eGFR 30–59
- Stage 4: eGFR 15–29
- Stage 5: eGFR  $< 15$

## V. SYSTEM ARCHITECTURE

### System Architecture Overview

The system proposed for the identification of CKD stages in patients infected with HIV is implemented by means of a web based client server architecture, therefore the system can guarantee its accessibility, scalability, and efficiency in the processing of clinical data. The architecture has been made to incorporate the predictions based on machine learning together with medically validated eGFR estimation and produce clinically interpretive outputs which can be used in routine healthcare use cases. The client layer includes user interaction and data entry; healthcare professionals view the system using a standard web browser and enter clinical and laboratory parameters of HIV infected patients using an intuitive web interface. This approach eliminates the need for special software installations and allows one to deploy it easily in hospitals and diagnostic centers. The server layer serves as the processing core and is implemented using the Flask web framework. It handles incoming requests and checks the input by the patients and orchestrates the entire flow of execution. The server combines the use of preprocessing modules, a trained machine learning model, and a rule-based medical computation engine. During deployment, the backend uses the pre-selected optimal model determined during the training phase and hence guarantees accurate predictions without having to manually select the model. A variable residence layer dedicated to medical computation calculations is able to calculate the estimated Glomerular Filtration Rate (eGFR) using the CKD-EPI equation, and the linked CKD stage. This ensures that outputs from this system are in accordance with internationally accepted nephrology guidelines. The last output layer shows CKD status, eGFR value, and disease stage in a clear and

interpretable manner which is good for clinical decision making for HIV-positive patients with an increased risk of renal complications.

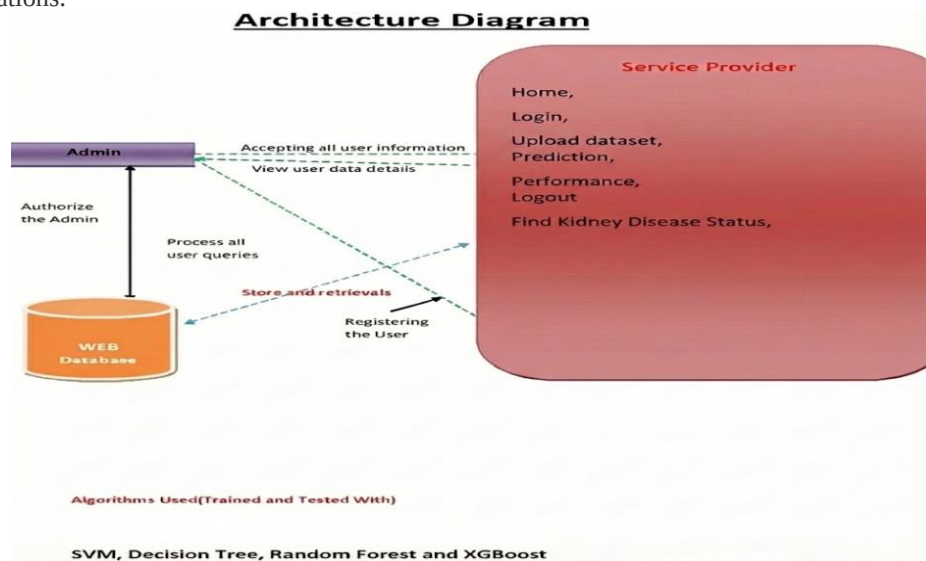


Fig. 3 An overview of the system architecture

### System Components

The proposed CKD prediction framework includes multiple interrelated components which all have a distinct role to maintain accurate prediction and meaningful clinical interpretation. The User Interface component is a web-based platform for users to input demographic information of and laboratory values for HIV-infected patients. It is designed in using the coding languages of Hyper Text Markup Language (HTML), CSS, Bootstrap to ensure simplicity, responsiveness, ease of use and thus enables clinicians to enter patient data without the need for technical expertise. Flask Backend component is the heart of the controller in the system. It deals with the handling of the request over the network protocol, the input validation and the coordination of one module with another and performs the feature formatting and scaling to be ready for machine learning prediction. Machine Learning Model component is responsible for Chronic Kidney Disease present or not. During the training phase multiple algorithms are tested, i.e. Support Vector Machine, Random Forest, Decision Tree and XGBoost, and the best performing model is automatically selected and saved. In the deployment stage, this trained model takes data fed to it from the HIV-infected patients and generates a CKD classification. The eGFR Calculation Module calculates the estimated Glomerular Filtration Rate using the CKD-EPI equation by using clinically relevant input parameters such as age, sex and serum creatinine which are key persons of kidney function for HIV-infected individuals. The CKD Staging Module correlates the computed eGFR value with standard stages of CKD from Stage 1 to stage 5 improving the clinical interpretability. Finally, the Result Visualization component displays CKD status, estimated eGFR and CKD stage in a structured and presentable format that facilitates healthcare professionals to quickly evaluate the health status and provide treatment respectively.

### Architecture Flow

The architecture flow of the proposed system follows a structured and sequential process in order to ensure accurate prediction and clinical relevance. Initially, patient demographics information and laboratory measurements related to renal function is provided by the user at the web interface; these usually correspond to HIV-affected patients for routine clinical evaluation. Upon submission, the data is sent to the backend (Flask) allowing the input validation and data preprocessing. These processes involve data normalization, categorical encoding, function feature alignment in order to provide compatibility with the trained machine learning model. The preprocessed data is then fed to machine learning prediction module after which the classification of the patient's renal condition will either be Normal or Abnormal which is required to provide a preliminary assessment of CKD presence or absence. Concurrently, the estimated eGFR is calculated by the system using the CKD-EPI based upon serum creatinine, age and sex. The calculated value of eGFR is then matched with the relevant stage of the CKD and has been shown to provide information on disease severity. Finally, the results of the predictions in terms of CKD status, eGFR value,

and CKD stage are fed back to the user interface and presented in an interpretable format. This end-to-end workflow ensures efficient prediction, medical interpretability and effective clinical decision supported may occur.

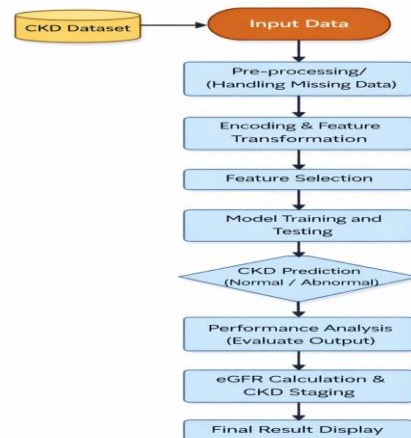


Fig. 4 presents the flow of architecture of the proposed system.

## VI. RESULT

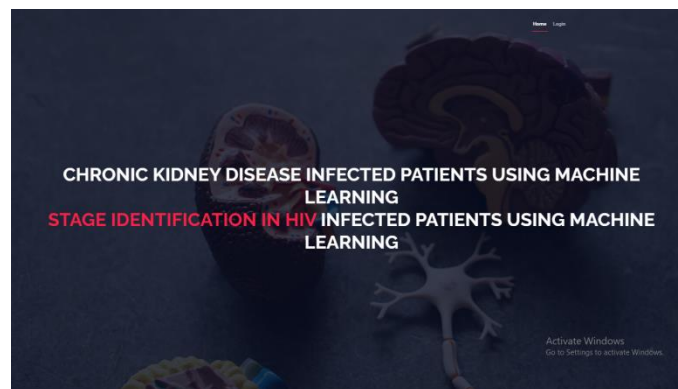


Fig. 5: Home Page of CKD Stage Identification System

This figure 5 shows the home interface of the proposed Chronic Kidney Disease stage identification system. It introduces the objective to detect CKD and identify stages of the disease in HIV infected patients with machine learning techniques. The interface offers simple navigation including the login, upload, prediction and performance analysis modules.

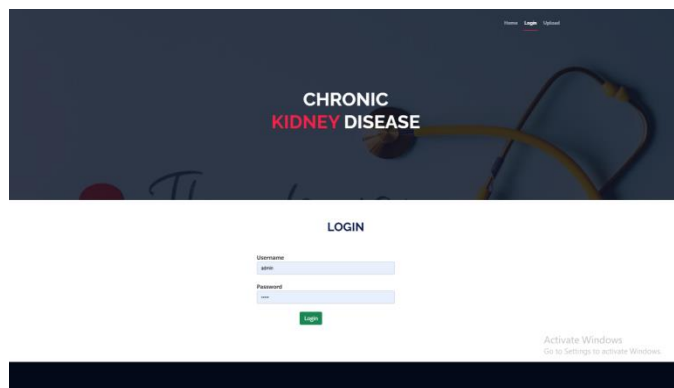


Fig. 6: User Login Interface

This figure 6 shows the authentication portal of the system, where authorized people are given authentic access into the application. Authentication mechanisms are used to ensure that there is controlled access to sensitive patient data and functionalities of the system, thereby protecting data integrity and confidentiality. Users, if authenticated successfully, are taken to modules for uploading of data and predictive analytics.

| Patient ID | Age | Sex | Blood Pressure | Albuminuria | Glycemic Readings | Serum Creatinine | Other Parameters |
|------------|-----|-----|----------------|-------------|-------------------|------------------|------------------|
| 19         | 19  | 010 | 1010           | 1.0         | 0.0               | normal           | abnormal present |
| 20         | 20  | 010 | 00.0           | 1010        | 1.0               | 0.0              | abnormal         |
| 21         | 21  | 000 | 00.0           | 1010        | 1.0               | 0.0              | abnormal         |
| 22         | 22  | 400 | 00.0           | 1020        | 4.0               | 0.0              | normal           |
| 23         | 23  | 210 | 70.0           | 1010        | 0.0               | 0.0              | normal           |
| 24         | 24  | 420 | 000.0          | 1010        | 4.0               | 0.0              | abnormal         |
| 25         | 25  | 010 | 00.0           | 1020        | 0.0               | 0.0              | normal           |
| 26         | 26  | 710 | 00.0           | 1010        | 0.0               | 0.0              | normal           |
| 27         | 27  | 000 | 70.0           | 1010        | 1.0               | 0.0              | normal           |
| 28         | 28  | 710 | 70.0           | 1.0         | 0.0               | normal           | abnormal         |
| 29         | 29  | 000 | 70.0           | 1005        | 1.0               | 0.0              | abnormal         |
| 30         | 30  | 710 | 70.0           | 1010        | 1.0               | 0.0              | normal           |
| 31         | 31  | 710 | 00.0           | 1010        | 0.0               | 0.0              | abnormal         |
| 32         | 32  | 010 | 00.0           | 1010        | 0.0               | 0.0              | normal           |

Fig. 7: Preview of Uploaded Dataset

This figure 7 shows the preview interface after upload of CKD dataset used for model training and evaluation purposes. It shows an abundance of clinical characteristics including blood pressure, albuminuria, glycemic readings, serum creatinine and other relevant laboratory parameters. The preview helps the user to check the quality of the data before the model is run, which also lowers the chances of erroneous predictions due to any inconsistent data entries.

Prediction Form

Age: 00  
Sex: Male  
Blood Pressure: 120  
Albuminuria: 1  
Glycemic Readings: 0  
Serum Creatinine: 0  
Other Parameters: 0  
Predict

Fig. 8: Predicting CKD Input Form

This figure 8 shows the structured form of the input used for the CKD prediction module. Demographic data and laboratory data of patients are entered into a standardized interface based on the feature set required by the underlying machine learning algorithms.

### Prediction Result

Abnormal

Estimated eGFR: 9.33 mL/min/1.73m<sup>2</sup>

CKD Stage: Stage 5 (eGFR < 15) — kidney failure

**Fig. 9: The Prediction Result Severe CKD (Stage 5)**

This figure 9 shows the result of the CKD prediction module in case of severe kidney impairment. Patient demographics and laboratory values are input into trained machine learning models to come to the result of an estimated eGFR below 15 mL/min/1.73 m<sup>2</sup>. The system considers the condition to be at CKD Stage 5, which means that the patient suffers from end-stage renal disease and should be closely monitored by a medical professional urgently.

**Prediction Form**

Age: 55  
 Sex: Male  
 Blood Pressure: 150/90  
 Albumin: 30  
 Sugar: 100  
 Red Blood Cells: 4.5  
 Hemoglobin: 12.5  
 ESR: 20  
 Blood Urea Nitrogen: 10  
 Blood Creatinine: 1.2  
 Potassium: 4.0  
 White Blood Cell Count: 12000  
 Hypertension: Yes  
 Diabetes Mellitus: No  
 Coronary Artery Disease: No  
 Pulmonary Disease: No  
 Asthma: No

**Prediction Result**

Normal

Estimated eGFR: 128.43 mL/min/1.73m<sup>2</sup>

CKD Stage: Stage 1 (eGFR ≥ 90) — confirm with albuminuria

**Fig. 10: CKD Prediction Input and Result Normal Kidney Function (Stage 1)**

This figure 10 depicts the prediction procedure for a good profile for the kidneys. The input form is filled with laboratory parameters which indicate normal kidney function. As a result, the estimated eGFR on the model must be equal to or above 90 mL/min/1.73 m<sup>2</sup>, which corresponds to that patient being in CKD Stage 1. The use of albuminuria assessment is recommended by the system to confirm the diagnosis of the early stages of renal involvement.

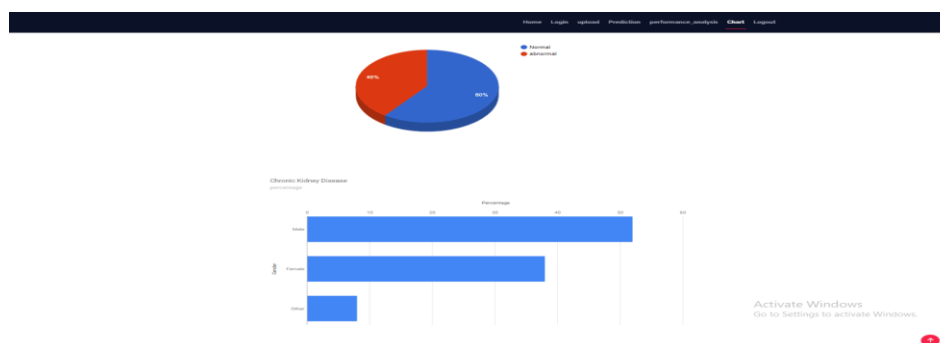


**Fig. 11: Graphical Analysis on Distribution of CKD**

This figure 11 has visual representations of predictive outcomes in the form of pie and bar charts. It shows the number of people with normal versus abnormal cases of CKD, as well as the distribution among the different individual stages of CKD. Such graphical analyses allow insight into dataset characteristics and general system performance, that helps to identify trends as well as possible biases.

## VII. CONCLUSION

This study has presented a global, clinically significant framework for the classification of Chronic Kidney Disease stages among HIV Patients by incorporating methods of supervised machine learning in combination with empirically tested rule-based evaluation. CKD is most likely to be an insidious process, especially in the HIV population, because of cumulative exposure to antiretroviral medication, immune dysfunction and coexisting comorbidities. Consequence of this delayed diagnosis is often an end result of advanced renal impairment and increased morbidity. The proposed system overcomes this issue by allowing early detection and precise stage of CKD based on analysis of routinely collected clinical and laboratory data.

Multiple supervised learning models have been trained and carefully evaluated including Support Vector Machine, Random Forest, Decision Tree and XGBoost algorithms for capturing complex inter-relationships within the clinical dataset. During the training phase, the system has autonomously picked out the best performing model based on pre-defined evaluation metrics, thus removing the need for manual model picking during the deployment phase. This design improves the robustness and predicts the performance in the clinical world.

In order to enhance the clinical reliability of the system, the system included estimation of the Glomerular Filtration Rate, based on the CKD -- EPI equation, and then maps the derived eGFR values to standardized CKD stages. The last predictive mechanism combines the outputs of machine learning classification with domain-specific nephrology guidelines, and ensures that the final decision of CKD status is based on both patterns that are identified from the data and those protocols that are accepted in the medical community. This hybrid approach enhances interpretability, lowers the risk of misclassification, and enhances clinician's confidence in the resultant predictions. The web-based architecture developed in this context provides scalability, accessibility and user-friendly interaction that makes the system easily deployable in the context of hospital and diagnostic laboratory settings. In sum, this research has shown that the combination of machine-learning algorithms and specific medical knowledge produces a powerful decision support system for the early detection and stage of CKD in people living with HIV. The findings highlight the potential of intelligent health information systems to help clinicians, reduce the diagnosis and delay time and improve the patient's outcomes with timely, evidence-based intervention.

## REFERENCES

- [1] A. Dritsas et al., "Machine learning approaches for early detection of chronic kidney disease," *Journal of Medical Systems*, 2022.
- [2] R. Ghosh et al., "Predicting chronic kidney disease progression using ensemble learning," *Scientific Reports*, 2023.
- [3] F. Chen et al., "A comparative study of classical and gradient-boosted models for CKD classification," *PLOS ONE*, 2023.
- [4] L. A. Inker et al., "New creatinine- and cystatin C-based equations to estimate GFR without race," *New England Journal of Medicine*, 2021.
- [5] S. Halder et al., "Explainable machine learning for CKD prognosis: SHAP and clinical validation," *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [6] M. Raihan et al., "Hybrid feature engineering and machine learning pipeline for CKD detection," *IEEE Access*, 2022.
- [7] P. Arif et al., "Kidney disease prediction using ensemble methods in low-resource settings," *Journal of Healthcare Informatics Research*, 2023.
- [8] H. Deb et al., "Comparative performance of machine learning classifiers for CKD: Cross-cohort evaluation," *Computers in Biology and Medicine*, 2021.
- [9] A. Praveen et al., "XGBoost and LightGBM in clinical risk prediction: A CKD case study," *International Journal of Medical Informatics*, 2020.
- [10] J. K. Patel et al., "Machine learning in nephrology: Opportunities and challenges," *Clinical Kidney Journal*, 2019–2020.

- 
- [11] Dr. Rekha Patil, Vidya Kumar Katrabad, Mahantappa, & Sunil Kumar. (2023). Image Classification Using CNN Model Based on Deep Learning. Journal of Scientific Research and Technology, 1(2), 60–71.
- [12] Bhagyashree, & Dr. Swaroopa Shastri. (2025). A Machine Learning Approach To Classify Medicinal Plant Leaf By Using Random Forest And KNN. Journal of Scientific Research and Technology, 3(7), 100–115.  
<https://doi.org/10.61808/jsrt261>