PHASE-1 SUBISSION TEMPLATE

Student Name: Rajalatha.B

Register Number: 510923243018

Institution: GLOBAL INSTITUTE OF ENGINEERING AND TECHNOLOGY

Department: B.tech Artificial intelligence and data science

Date of Submission:07 -05-2025

1.Problem Statement

AI can predict acute kidney injury (AKI) by analyzing various data points, such as electronic health records (EHRs), vital signs, and laboratory results, to identify patients at high risk. For example, models using machine learning algorithms like logistic regression, gradient boosting, and deep learning have demonstrated strong predictive capabilities in intensive care units (ICUs). These models can predict AKI up to 48 hours in advance, allowing for timely interventions to potentially improve patient outcomes.

2.Objectives of the Project

* AI models can be trained to predict AKI before it's clinically evident, potentially allowing for timely interventions.
* This can involve using machine learning algorithms to analyze patient data, such as vital signs, lab results, and medical history, to identify individuals at high risk of developing AKI.
* AI-powered tools can provide clinicians with real-time risk scores and alerts to help them identify patients at high risk of AKI.
* This can enable clinicians to adjust treatment strategies, such as avoiding nephrotoxic medications or implementing hydration protocols, to prevent or mitigate AKI.

3. Scope of the Project

* Data integration

Here’s a detailed explanation of **Data Integration** in the context of an AI-powered **Acute Kidney Injury (AKI) prediction** project:

### **Data Integration**

* + **Objective:**  
    To consolidate and harmonize patient data from multiple sources into a unified dataset suitable for AI model training **Electronic Health Records (EHRs)**: Lab test results (e.g., creatinine, BUN), clinical notes.**Monitoring Devices**: ICU monitors (blood pressure, heart rate, urine output).**Pharmacy Records**: Medications administered (e.g., nephrotoxic drugs).
  + Build a data warehouse or data lake to aggregate and organize cleaned data for access by the AI model.

Data Preprocessing

AI can enhance the prediction of Acute Kidney Injury (AKI) by processing large datasets from Electronic Health Records (EHRs) and other sources. This data processing allows for early detection, risk stratification, and potentially personalized treatment strategies. The scope includes leveraging various AI algorithms like machine learning (ML) to analyze patterns in patient data and identify risk factors for AKI development.

Data Sources for AKI Prediction:

* **Electronic Health Records (EHRs):**

EHRs provide comprehensive patient information, including demographics, medical history, laboratory results, medication details, and diagnosis codes, making them a valuable source for building predictive models.

* **Dynamic, Real-Time Data:**

Incorporating real-time data, such as vital signs, physiological parameters, and intraoperative/postoperative factors, can improve the accuracy of AKI risk prediction compared to static models.

* **Specific Clinical Settings:**

Research focuses on predicting AKI in various settings, including post-cardiac surgery patients, ICU patients, and those with acute myocardial infarction.

* **Specific Patient Populations:**

Studies often target specific patient populations, such as older adults, pediatric patients, and those with chronic kidney disease.

High levels methodology

AI and machine learning (ML) algorithms offer promising avenues for predicting Acute Kidney Injury (AKI) with high accuracy and early detection, enabling timely interventions and potentially improving patient outcomes.

Data collection

Prior to these AI investigations, AKI risk prediction was done through static prediction based on baseline (or preoperative) data. However, newer AI models incorporate baseline (pre-AKI) as well as evolving, dynamic, real-time data collected during a hospital admission which show improved predicative abilities.

* **Data Cleaning:**

This involves identifying and removing errors, inconsistencies, and missing values in the data. This can include:

* + **Identifying and removing non-physiologic values:** Ensuring that data points, such as vital signs or lab values, are within the expected range.
  + **Handling missing data:** Deciding how to treat missing values, such as using imputation (e.g., replacing missing values with the mean or median) or keeping them as missing.
  + **Data alignment:** Ensuring that variables are correctly labeled and aggregated across different institutions, as the same variable may be labeled differently in different systems.

Exploratory Data Analysis (EDA) in AKI Prediction:

1. **1. Data Collection and Preprocessing:**

EDA begins with gathering relevant patient data, including demographics, medical history, laboratory results, and vital signs. This data needs to be cleaned, transformed, and prepared for analysis. Missing values, inconsistencies, and outliers must be addressed.

1. **2. Data Exploration and Visualization:**

EDA involves exploring the data to understand its characteristics, identify patterns, and uncover potential relationships between variables and AKI development. Techniques like histograms, scatter plots, and box plots can be used to visualize the data and identify trends.

1. **3. Feature Selection:**

EDA helps identify the most relevant features or variables that are strongly associated with AKI. This includes variables like age, creatinine levels, urine output, presence of comorbidities, and specific medications.

1. **4. Understanding Data Distributions and Correlations:**

EDA helps assess the distribution of each feature and identify any potential correlations or relationships between them. For example, EDA might reveal a strong positive correlation between serum creatinine levels and the risk of AKI.

1. **5. Identifying Outliers and Anomalies:**

EDA helps detect unusual data points or anomalies that might skew the analysis or training of the ML model.

Features engineering

* Transforming raw data into features that can be used by ML models.
* Here’s an overview of **Feature Engineering** for an AI-powered **Acute Kidney Injury (AKI) prediction** project:

**Objective:**  
To create meaningful, predictive features from raw patient data that enhance the performance of machine learning models in forecasting AKI.

### **Key Feature Categories:**

#### 1. **Demographic Features**

* Age
* Sex
* Ethnicity
* Weight and BMI
* Comorbidities (e.g., diabetes, hypertension)

#### 2. **Laboratory Results (Time-Series & Trends)**

* Serum creatinine (most critical indicator)
* Blood Urea Nitrogen (BUN)
* Electrolytes: Potassium, Sodium
* Glomerular Filtration Rate (GFR)
* Hemoglobin, Platelet count

**Derived features:**

* Creatinine change over time (e.g., delta over 48 hours)
* Rate of creatinine increase
* Abnormal value flags

#### 3. **Vital Signs**

* Blood pressure (systolic/diastolic)
* Heart rate
* Respiratory rate
* Oxygen saturation
* Temperature

**Derived features:**

* Variability or deviation from baseline
* Abnormal trend detection (e.g., hypotension)

#### 4. **Medication Features**

* Use of nephrotoxic drugs (e.g., NSAIDs, ACE inhibitors)
* Dosage and duration of medication
* Recent contrast media exposure (from imaging)

#### 5. **Clinical Event Features**

* Recent surgeries or ICU admission
* Urine output (volume and trend)
* Diagnoses and procedure codes (ICD/CPT)

#### 6. **Temporal Features**

* Time since admission
* Time of day when lab/vitals were recorded
* Time gaps between measurements

#### 7. **Text-derived Features (optional/NLP-based)**

Keywords or phrases from clinical notes (e.g., "fluid overload", "oliguri. Sentiment or concern levels flagged by physician.

7.Team Members and Roles

|  |  |  |
| --- | --- | --- |
| **Team Member** | **Role** | Responsibilities |
| 1: Sharamini R | Team leader | Oversees project planning, task delegation, timeline management, and coordination among team members. |
| 2: Rajalatha B | Data Analyst | Collects patient datasets, performs data cleaning, handles missing values, and prepares data for modelling. |
| 3: Pavalya J | Machine learning engineer | Develops and trains AI/ML models for disease prediction, tunes hyperparameters, and ensures model efficiency |
| 4: Veena R | Documentation & Deployment Lead | Prepares reports, creates project presentation, and integrates the final model into a user-friendly interface or system. |