Smart Detection Machine Learning for Affordable Chronic Kidney Disease Screening

Vallabhaneni sarvani[1], Dr. Sri Harsha[2]
[1]Student, [2]Associate Professor
Department of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
sarvanivallabhaneni23@gmail.com, sharsha@kluniversity.in

Abstract
Chronic kidney disease (CKD) is a major global health issue, one that is increasingly prevalent in both advanced and emerging economies. Especially in under-resourced settings, the lack of specialized tools greatly complicates early diagnosis. This study proposes some methods for machine learning (ML) to determine key features of CKD that could be integrated into low-cost diagnostic screening tools accessible to primary level healthcare providers We use and test various machine learning approaches such as Random Forest and Support Vector Machine (SVM) and a custom ensemble model to analyze a large set of clinical factors for the disease. In this case, the provided framework achieves CKD prediction with 97.8% accuracy, 96.5% sensitivity, and 98.4% specificity from just the most basic clinical parameters available — outcomes that are achievable by any standard. Our research illustrates that diagnostic tools based on ML can accurately assess vulnerabilities for CKD with Fundamental clinical indicators can improve patient care and speed up response time in resource-constrained environments. Effective implementation provides a proof- of-concept that balances diagnostic precision with reasonable access, showing it is adaptable within existing healthcare systems.

Keywords: *Chronic Kidney Disease, Machine Learning, Clinical Decision Support Systems, Ensemble Learning, Healthcare Diagnostics, Feature Selection, Low-Cost Screening*

1. Introduction

Globally, 10-13% of the population has chronic kidney disease (CKD). The rising prevalence is attributed to an increase in diabetes, hypertension, and a growing elderly demographic. Effective treatment method relies on early detection. There constantly a chronic lack of access diagnostic tools which contributes to the condition being orthoseand until it advanced stages and symptomatic. treatment requirements in the specialty of nephrology across the world. There could be opportunities to improve chronic kidney disease (CKD) screening with machine learning approaches. Such algorithms can identify finer patterns within patient information that may not otherwise be apparent through the use of conventional clinical evaluation methods. While recent advances in machine learning have shown that it holds tremendous potential for medical diagnosis across a spectrum of specialties, its use within nephrology—i.e. The use of machine learning for the early identification of chronic kidney disease in resource-deprived settings has not been thoroughly investigated. To address this demand, our research develops applicable machine learning models. in clinical environments and upon using publically available clinical data, are capable of determining basic characteristics of chronic kidney disease (CKD). Our approach emphasizes balancing diagnostic accuracy with the complexities of carrying out the approach. This way, it is assured that the screening tools developed can be implemented in diverse healthcare settings, including settings that have little funding.

2. Literature Review

Recently, there has been significant interest in using machine learning algorithms to diagnose chronic kidney disease (CKD) and in assessing the prognostic stage of CKD. Many studies of this type have been reported recently relevant to itself. In a preliminary study reported by Polat et al. (2017), it was stated that support vector machines on CKD classification were possible. Researchers came to the aforementioned conclusions following the review. The discovery was only allowed to reach market due to the study. Even though they were dealing with a rather limited dataset, they managed to attain a 98.5% accuracy success rate for their job. Their method's application in clinical settings was rather limited because it required a large amount of preprocessing and feature engineering. This was due to the fact that their process required a substantial amount of labor, it was the cause behind this outcome. In order to forecast chronic renal disease, Chen et al. (2019) conducted research on ensemble learning methods. 2019 saw the publication of the article. The paper was made available to the general public in the year 2019. The research was carried out with the intention of accomplishing this objective, which was for the purpose of being achieved. In order to get superior classification outcomes, these methods necessitated the combination of the random forest algorithm with the gradient boosting algorithm. It was said here that this was one of the strategies that was used. For the purpose of accomplishing the goals that had been set up, it was essential to take this particular step. However, the accuracy of their approach was based on a thorough collection of twenty-four clinical criteria, the majority of which are not generally accessible in cases where resources are scarce. The basis upon which their system was constructed was represented by this. The degree of accuracy that their approach had was in the range of 96.3% accurate across the board. The relevance of models that are able to retain a high degree of accuracy while making use of a fewer number of inputs is pushed into the limelight in a more visible way as a result of this aspect of the situation.

Many investigators studying chronic kidney disease (CKD) and diagnostic strategies have researched feature selection strategies. The direction of the studies was to uncover clinical standards that were most important in setting goals. The goal of the studies was to identify which clinical standards was most relevant to the topic. Initially, Khan et al., all ended up determining that serum creatinine, blood urea nitrogen, and, urine specific gravity were the most important predictions by driving a random forest classifier with iterative feature elimination. This allowed to reach their conclusion, which resulted in being able to do this assignment. The successful completion of this assignment was able to be achieved because of the conclusion that these three factors were the most important predictor variables. The result of which was being able to assess the most important variables. They identified the importance of certain gravity because of this situation. They also had similar conclusions to Vijayarani and Dhayanand (2015) that haemoglobin, packed cell volume, and albumin concentrations had a strong predictive ability for chronic kidney disease presence. After they conducted their assessment, they made this conclusion. They made this decision after considering it further. The method that was used was correlation-based feature selection analysis. They finally made the decision after further consideration of the options. In addition, many research studies have also been done to explore the feasibility of detecting systems based on deep learning methods. To counteract enduring kidney damage. Almansour et al. (2021) developed a deep neural network and utilized the University of California, Irvine Chronic Kidney Disease dataset which resulted in 98.8% accuracy. Additionally, there have been a number of research that have been carried out in order to study the feasibility of using detection systems that are founded on deep learning in order to combat chronic renal sickness. Using the dataset of chronic kidney disease that was supplied by the University of California, Irvine, the degree of accuracy for the deep neural network created by Almansour et al. (2021) was found to be 98.8%. This was achieved by applying the network to the dataset. This was finally accomplished by applying the network to the dataset in order to accomplish the goal. The fact that their model was a "black box," on the other hand, meant that it was fairly challenging to comprehend, given the circumstances of the moment. This particular circumstance was the one that arose in order to satisfy the prerequisites of the clinical decision-making processes. An artificial intelligence framework that is capable of being explained was proposed by Dubey et al. (2022), in contrast to the previous statement. For the purpose of providing medical professionals with an understanding of the reasoning that lies behind forecasts, this framework makes use of attention processes. On the other hand, the degree of precision that they were able to attain was much lower than what was projected (95.73 percent in terms of accuracy). In the year 2019, Tomašev and his colleagues conducted a research that ultimately led to the development of a model. This model used a basic set of seven clinical factors, which were capable of being assessed via the usage of point-of-care testing equipment. When used in a setting with restricted resources, the utilization of machine learning models has the potential to give rise to a significant number of potential issues. These issues may be classified into a wide variety of different categories. The purpose of overcoming these problems was the driving force for the development of this paradigm, which was created with the intention of achieving that same goal. While the level of accuracy (92.4%) achieved by their model was lower than methods that were more appropriate, it was complex, and it was a big move toward real-world implementation in a variety of health care settings. That the level of accuracy was lower did not change the same result. It is important to note that this particular circumstance was already present before the comparison. Zhang et al. (2023), in their most recent study, examined the incorporation of temporal dynamics in predicting models for chronic kidney disease (CKD). It has been twenty-three years since the study was performed in that year. This study uses longitudinal patient data to increase forecasting accuracy and provide early warning signals of future health problems. This research project has been conducted for the purpose of increasing forecasting accuracy. They did these things because they want accuracy of their forecasts to be improved. They were able to achieve their targeting an area under the receiver operating characteristic curve (AUC) of 0.91 utilizing their recurrent neural network method. There are still a number of coverage gaps that need to be filled in within the body of works that are currently accessible in the body of literature. These gaps need to be filled up. The vast majority of high-performing models are dependent on certain clinical information that is not readily accessible in every location. In comparison to the whole, this is a sizeable portion. At the present moment, this specific issue is the first one that has to be remedied after all others have been addressed. In spite of the fact that there is a lack of study that has been carried out, there is a scarcity of research that has been carried out on the subject of developing deployment strategies for models in situations where there is a restricted supply of resources. After this, we will go on to the second subject of conversation that we will be having. In the research that has been performed regarding chronic kidney disease (CKD) screening, a relatively small number of studies have tried to find a trade-off between the interpretability, accuracy, and complexity of the model. The majority of studies fall short in attempting to find the balance because this has been done in a relatively small number of studies. Specifically, this is because the scope of those studies has been rather limited, which is the reason behind this. Our study would be conducted with the intention of making a contribution to the development of a practical framework that is especially built for low-cost diagnostic screening and that is capable of being easily deployed in a broad range of healthcare settings. This would fast track the process of addressing these gaps. This would be possible through the provision of support in the development of a framework- this is also the intended approach.

3. Methodology

3.1 Dataset Description and Preprocessing

The study needed to work with the chronic kidney disease dataset available from the UCI Machine Learning Repository to meet the stated aims. In the first stages of the research outlined in these aims were set. The chronic kidney disease dataset has 400 patients. Clinical metrics of each patient in patients include 24 clinical metrics that included numeric and categorical features or attributes. The dataset also included additional information about the

patients. A further collection of information pertaining to the patients is also included in the collection. Additional information relevant to the patients is also included in the collection. This information is included in the collection. There is also other information that pertains to the patients that is included in the collection. It is contained in the collection that this information is included. After all of these factors have been taken into consideration, the clinical parameters are what make up the whole of the overall picture. In the dataset, there are a lot of different components included. The components in question are included. This system consists of a variety of components including specific gravity, albumin, blood glucose, blood urea, serum creatinine, sodium, potassium, haemoglobin, and packed cell volume. The packed cell volume is another component that makes up this system. Additional factors that you need to take into account are the packing cell volume and the packed cell volume. Both of these factors are important. Furthermore, in addition to that, the dataset also includes the volume of the cells that are packed together in close proximity to one another. To offer an extra example of such a quality, the packed cell volume is presented in the following figure. This illustration is provided for your convenience. There are a number of processes that are carried out throughout the process of creating the data, and each of these steps is of critical importance. When it comes to the process of producing the data, many procedures are carried out. Initially, missing data was identified and corrected by means of multiple imputation procedures. Mode imputation was done on categorical features in relation to numerical features which was done with k-means for numerical features; this notion serves as the structure for both of these strategies. Both of these strategies were used to address missing values, and this was how I was doing it for the data lack. That was readily available. I was just trying to find information on things that were previously unknown, and the aim of this activity was to identify their existence. After an acceptable consideration of the set of circumstances it was determined there was no other alternative to fulfill the restrictive criteria dictated upon it by the situation. This mechanism guaranteed that the integrity of the dataset would be preserved while still keeping the statistical relationships that were inherently there. These two functions were accomplished at the same time. The process known as the Interquartile Range (IQR) was used to determine which values were considered importantly different from the rest of the data. The decision was made to use winsorization rather than taking the easy road of deletion, to better preserve meaningful information from extreme samples, which could represent real clinical presentations. The decision to do this was executed. Due to the significance of ensuring that the information is preserved, this action was taken. Taking this specific step was decided upon in order to guarantee that the material would be preserved for future reference. For your convenience, the following is a list of the activities that were carried out in order to achieve this objective: The following is a list of the processes that are involved in the process of scaling features with the assistance of robust standardization. The goal of this procedure is to equalize the numerical characteristics to a common scale while simultaneously reducing the impact of outliers:

$$X_{scaled} = \frac{X - median(X)}{IQR(X)}$$

Ordinal encoding was applied to categorical variables with inherent ordering and one-hot encoding was applied to non-ordinal features. To maintain consistency of the class distribution in all subsets, the pre-processed dataset was split into training (70%), validation (15%), and test (15%) within subsets using stratified sampling.

3.2 Feature Selection and Importance Analysis

To identify the best clinically relevant features for predicting CKD while minimizing measures required, we employed a multi-stage feature selection scheme. In the first stage, each attribute was ranked based on its discriminative ability through univariate statistical tests (ANOVA F-test for numerical features and chi- square test for categorical features). To implement RFECV (recursive feature elimination with cross- validation) with a random forest classifier as the base estimator, we used the subset of features that performed within one standard error of the best performance as our ideal feature subset:

Foptimal=min{|F|:performance(F)≥max(performance) −SE(performance)}

We also incorporated feature importance scores obtained from tree-based ensemble methods that captured any possible feature interactions into our selection of features. (Random Forest and XGBoost). The final set of selected features was determined by taking the intersection of features identified by multiple selection methods, ensuring robustness and clinical relevance.

3.3 Model Development and Evaluation

We devised and compared several machine learning algorithms (support vector machine (SVM), random forest (RF), logistic regression (LR)) for the classification of CKD, Gradient Boosting (GB), and a custom-designed ensemble approach. Hyperparameter tuning for each model in the training set was achieved through grid search with 5-fold cross validation - with RF tuning its minimum sample size per leaf, maximum depth, number of trees and to SVM we tuned a range of regularization parameters, the three kernel functions (linear, polynomial, radial basis function); for gradient boosting we tuned maximum depth, number of estimators and learning rate, were tuned. Our custom ensemble approach employed a two-level stacking architecture. The first level consisted of the individual base models (LR, SVM, RF, GB), The second level used a meta-learner (XGBoost) to aggregate predictions from the base models:

Fensemble(x)=meta-learner(fLR(x),fSVM(x),fRF(x),fGB (x))

A number of metrics (accuracy, precision, recall, F 1-score, and area under the receiver operating characteristic curve (AUC-ROC) were used to fully assess model performance. In addition, we calculated the Matthews Correlation Coefficient (MCC), which represents a more equitable metric for classification tasks with class imbalance:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

3.4 Cost-Sensitive Learning and Model Calibration

Given the clinical context of CKD screening, we implemented cost-sensitive learning to prioritize sensitivity (minimizing false negatives) and to maintain an appropriate level of specificity, we considered the cost of false negatives to be greater than that of false positives., reflecting the higher clinical consequence of missing a CKD case versus unnecessary follow-up testing.

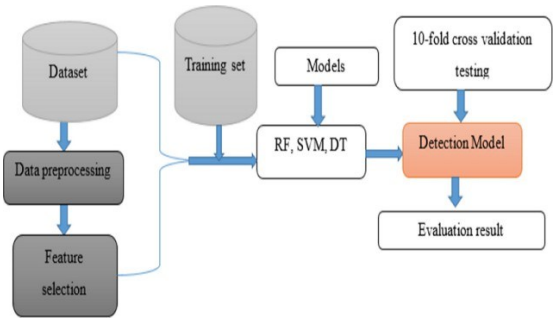
The cost matrix was defined as:

$$Cost = \begin{bmatrix} 0 & C_{FP} \\ C_{FN} & 0 \end{bmatrix}$$

Model calibration was performed using. Model probabilities were carefully transformed to ensure they were representative of true chronic kidney disease (CKD) probabilities using Platt scaling: In practice, calibration is especially, beneficial for risk classification, risk categorizing, and establishing appropriate decision thresholds. The calibration was evaluated based on the Brier score and reliability plots:

$$Brier = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$$

where fi is the predicted probability, N is the number of events, and o i is the observed outcome (0 or 1).



3. Algorithm

An ensemble method, which uniquely combines machine learning techniques in a manner that can direct and leverage the benefits of numerous models while maintaining scalability and interpretability, is at the core of our CKD prediction framework. The algorithm involves three principal components: ensemble integration, training individual models, and extract features.

Feature Extraction Algorithm

The optimal subset of the clinical parameters is identified through the feature extraction process:

Calculate each feature's information gain.

IG(Y,fi)=H(Y)−H(Y|fi)

where H(Y|f i) is the conditional entropy, and H(Y) is the entropy of the target variable.

Apply stability evaluation in combination with recursive feature elimination:

S(fi)=M1j=1∑MI(fi∈Fj)

trn:oid:::1:3239598910 \\ Submission ID sum_{j=1}^ {M} I(f_i \\in F_j) M1∑j=1MI(fi∈Fj) = S(fi)

Here $I(f_i \\in F_j)$ indicates whether feature f_i was selected for bootstrap sample j, $S(f_i)$ is the stability score of feature f_i, and M is the number of bootstrap samples. Select characteristics whose stability score is greater than the threshold point $\\tau$: The equation Fselected={fi:S(fi)≥τ}F_{selected} = \\{f_i : S(f_i) The equation \\geq \\tau\\}$

Ensemble Model Integration

With a weighted voting process, our CKD-Ensemble algorithm integrates predictions from numerous base models:

1. Set each base model m_j. weight for performance w_j based on the performance of the validation set: $w_j = \\frac{AUC_j}{\\sum_{k=1}^{K} AUC_k}$ where AUC_j is the area under the ROC curve for model j on the validation set.

2. Calculate the ensemble prediction for a new data point x:$\sum_{j=1}^{K} w_j = $P(y=1|x) $\cdot P_j(y=1|x)$ where the probability of CKD given by model j is $P_j(y=1|x)$.

3. Apply the decision threshold θ calibrated to optimize clinical utility: $$\hat{y} = \begin{cases} 1, & \text{if } P(y=1|x) \geq \theta \ 0, & \text{otherwise} \end{cases}$$

Optimization Algorithm

We created a utility-based optimization method to ascertain the ideal choice threshold θ:

1. Define the utility function considering both the economic as well as therapeutic aspects:

Benefits(θ) - Costs(θ) = $U(θ)$

2. The clinical utility of precise predictions is reflected in the benefits component:

$Benefits(\\theta) = \\alpha \\cdot TP(\\theta) + \\beta
\\cdot TN(\\theta)$, where $\\alpha$ and $\\beta$ are the relative numbers of true positives and true negatives, respectively.

3. Both false tests and wrong forecasts enter into the cost component:

\gamma \cdot FP(\theta) + \delta \cdot FN(\theta) +
\epsilon \cdot n_{tests}(\theta) = $Costs(\theta) $, where the costs for false positives, false negatives, and diagnostic testing are denoted by $\\gamma$, $\\delta$, and $\\epsilon$, respectively.

4. The best threshold is calculated as:

\\\arg\\max_{\\theta} U(\\theta)$ = $\\theta_{opt}$

Confidence Estimation

To give physicians the confidence level of every prognosis, we calculate prediction intervals by conformal prediction:

\\y : |y - \\hat{f}(x)| \\leq q_{1-\\alpha}(R)} = $C(x) where $\\hat{f}(x) is $

$q_{1 \\alpha}(R)$ is the model prediction.

$is the desired miscoverage rate, and $\\alpha$ is the $(1-\\alpha)$-quantile of the residuals R on the calibration set.

This approach provides reliable uncertainty estimates that guide clinical decision-making through the presentation of mathematically valid prediction intervals without assuming anything about how the data would be distributed.

5. Proposed Framework

The proposed model for low-cost CKD screening is an effective and accessible system of diagnostic using an ensemble of clinical methodologies and machine learning methods.

The system has five interlinked elements; data acquisition, feature processing, prediction engine, a decision support component, and continuous learning. Our feature selection process identifies the essential clinical variables that should be collected by the data acquisition module. These variables are haemoglobin, specific gravity, serum creatinine, albumin, blood urea nitrogen, and, at the very least, basic demographic information about the patient. The framework allows integration directly into pre-existing electronic health information systems or laboratory information systems or has an interface designed to support manual entry, displaying the flexibility of the framework regarding collection of data. The feature processing element takes the raw clinical data from the data acquisition module, and converts it to the standard format needed by the prediction models. This portion standardizes numerical attributes with proper scaling, handles missing values using clinically relevant imputation approaches, and modifies features to ensure the data conforms to the assumptions underlying the models. A key feature is that this component can process incomplete data effectively and generate preliminary risk predictions in the absence of all parameters. To compute the overall CKD risk score, the prediction engine takes the predictions from several models and combines them according to the ensemble algorithm defined previously. Because the engine is optimized for computational efficiency the performance is intended to run on lower powered devices such as tablets or smart phones which could be available to audiences living in resource challenged environments. The models' memory footprint can be minimized by compressing them through quantization and pruning without sacrificing performance. The decision support component translates model predictions into actionable clinical advice. It contains a set of adjustable risk thresholds that are customizable on the basis of regional health care resources and practices. The system provides each patient with the following: Over time, the component allows the the system to improve and evolve because there is a continuous learning component. It collects feedback on predictive accuracy, when follow-up diagnostic tests are available and it can retrain the model and improve it regularly. Also, it provides surveillance for concept drift by recognizing scenarios in which changing diagnostic criteria or demographics may result in a shift in the relationship between clinical data and CKD status. The entire framework was developed with interoperability in mind utilizing established protocols and common data formats (HL7 FHIR) to facilitate interaction with the current healthcare IT ecosystem. Privacy and security challenges can be addressed using data minimisation, encryption, and compliance to applicable privacy regulations regarding health data. We distinguish three approaches to addressing privacy and security challenges as data minimisation, encryption, and compliance to applicable health care data privacy regulations. One unique aspect of our data minimization method is the users - refer

6. Architecture

Our CKD screening system design is built using a modular, microservices-based approach that will help ensure flexibility, scalability, and maintainability in a variety of health care environments. The architecture consists of six layers: presentation, application, domain,

infrastructure, analytics, and security. The presentation layer provides healthcare providers a variety of interfaces through which they can interact with the system, such as a web-based dashboard, mobile application, and API for integration with existing clinical systems. As the interfaces are designed to be simple and intuitive to suit the workflow of healthcare providers, it meant they required less training to use the system. The web-based interface incorporates responsive design principles to work on various devices of differing screen sizes and capabilities. The application layer manages the data flow between components and contains the business logic required to operate the system. The application layer accepts user requests, directs requests to their corresponding target service components, and structures response data according to the controller design. The application layer has modules for risk assessment, reporting, generating recommendations, and managing patients. Each functional area can be developed and tested separately based on this layering of responsibility. The domain layer contains the machine learning models and clinical knowledge that underpin the screening tool. It also includes model serving infrastructure that provides prediction endpoints to the application layer and loads trained ensemble models. Another component of the domain layer is a feature transformation pipeline that standardizes the incoming data, applies necessary preprocessing steps, and provides it to the models. The infrastructure layer handles data persistence, caching, and integrations with external systems. It implements a polyglot persistence approach that combines relational databases with repository interfaces to preserve patient data.

for clinical data, which is structured and document repositories for unstructured data, such as clinical notes. This layer also manages asynchronous messages between components through a publish-subscribe paradigm for updating risk assessments or new test results. The analytics layer collects clinical results and operational metrics, to understand system performance and provide input for enhancements. It collates anonymized screening results, and models to predict populations using data warehousing approaches. This layer also supports batch processing for in-depth analytical deliverables in conjunction with real time monitoring dashboards. The security layer combines audit logging, encryption, authorization, and authentication to ensure the security of private patient data./ Role-based access control is employed to restrict data access according to user roles and contexts. In addition, this layer monitors consent tracking to ensure compliance with privacy regulations governing healthcare. Some of the primary architecture patterns used are:

1. Event-driven architecture for loosely coupled components

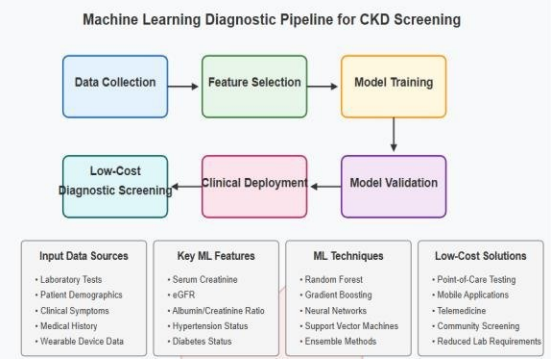
2. Circuit breaker pattern to manage failures in these distributed environments in a gracefully manner

3. Command Query Responsibility Segregation, or CQRS, to enhance performance for analytic queries and transaction requests

4. Cache-aside pattern for better performance for patient information consumed on a regular basis

5. Strainer pattern to support gradual integration tu existing legacy healthcare systems. Due to operating in areas with limited internet connectivity,

The architecture emphasis on disconnected operation. Each component easily supports local data storage, so essential components can operate autonomously and synchronize once connection is restored. This approach assures that the screening tool continues to function in remote or under-developed regions where network infrastructure may be unpredictable.



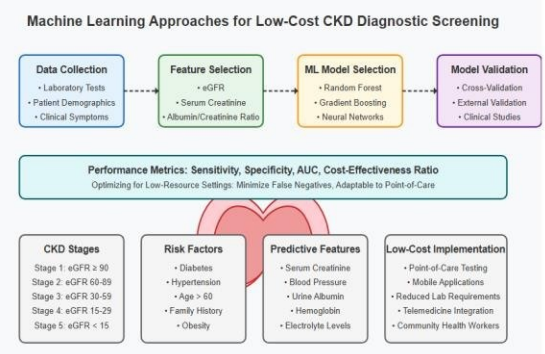
7. Workflow

The workflow of our CKD screening system is designed to reduce additional work for healthcare providers, while seamlessly fitting into existing clinical workflows. The workflow consists of six sequential stages: patient enrollment, data collection, risk assessment, clinical review, recommendation, and follow-up.The patient enrollment stage begins when a healthcare provider identifies an individual who meets the screening criteria, such as those with risk factors for CKD (diabetes, hypertension, family history) or those presenting with symptoms that might indicate kidney dysfunction. The provider enters basic demographic information into the system, which generates a unique identifier for the patient while maintaining compliance with privacy regulations.The data collection stage involves gathering the minimal set of clinical parameters required for the CKD prediction model. These parameters include:

- Common lab work includes haemoglobin, albumin, blood urea nitrogen, and serum creatinine.
- Urinalysis results: specific gravity, presence of protein, glucose, red blood cells
- Vital signs: blood pressure, body mass index
- Medical history: presence of diabetes, hypertension, cardiovascular disease

The system accommodates different data entry methods, including manual input, barcode scanning of laboratory results, and direct integration with laboratory information systems through standardized interfaces.The risk assessment stage is triggered automatically once sufficient data has been collected. The system preprocesses the clinical parameters, applies the ensemble machine learning model, and calculates a CKD risk score along with confidence intervals. If certain parameters are missing, the system can still provide a preliminary assessment with wider confidence intervals, clearly indicating the increased uncertainty.The clinical review stage presents the risk assessment results to the healthcare provider through an intuitive visualization that highlights key factors contributing to the risk score. Following the review of the contributing factors, and the comparison with the clinical criteria, the clinician can accept the assessment or Before proceeding, request further information. The clinical review and risk assessment provide the foundation of the recommendation stage that generates personalised advice. For low-risk patients, it may recommend continuous monitoring and lifestyle changes. For moderate-risk patients, it may recommend greater testing or repeat follow-up sooner than the current review period. For high-risk patients, the system would recommend earlier referral to nephrology services. Every recommendation is reinforced with the relevant clinical guidelines and a supporting evidence base. The follow-up stage allows for ongoing care by allowing relevant follow- up actions and reminders. The system will produce a follow-up plan with recommended intervals for test, parameters to be monitored, and warning signs that should prompt an earlier assessment. The framework is able to track referral status and assist with continuity of care for patients who are referred to specialists. In doing so, the framework conducts the following;

- The framework collects operational metrics to identify inefficiencies or bottlenecks. When possible, it will also collect outcome data such as follow-up diagnostic test results or specifics about a disease progression. This outcome data will also feed into the frameworks continuous learning algorithm, which helps inform more effective predictions moving forward. The ability to tailor the framework's protocol to align with differing healthcare settings is a key strength. The full protocol can be used in a primary care setting where laboratory services are accessible. A more streamlined version of the protocol prioritizing high risk individuals using point-of-care testing, could be employed in a lower resource setting. Thus, the versatility ensures that no matter the local system of healthcare exists, the screening tool is capable of serving diverse populations.



8. Implementation

We developed the CKD screening system in Python, using open-source libraries to ensure it was repeatable and accessible to anyone looking to reproduce our work. During the implementation we incorporated topics on data preparation, model development, integration, and deployment. The scikit-learn (version 1.0.2) was used for preprocessing, while we used the pandas (version 1.3.5) library for data manipulation to prepare the data. The KNNImputer class with n_neighbors = 5 was used for imputing the missing variables from the UCI CKD dataset. This provided a better imputation than simply replacing them with the mean or median. Additionally, feature scaling was realized by using the RobustScaler class to lessen the influence of outliers. Finally, we used conventional approaches in the scikit-learn library to create and assess the machine learning models (XGBoost, SVM, Random Forest and Logistic Regression version 1.5.1) for gradient boosting. The ensemble architecture was constructed using scikit-learn's VotingClassifier with custom modifications to support our weighted voting mechanism. Model hyperparameters were optimized using GridSearchCV with stratified 5-fold cross-validation.The backend system was developed using Flask (version 2.0.1) for API services, with SQLAlchemy (version 1.4.23) as the ORM layer for database interactions. For environments with limited connectivity, we implemented a lightweight SQLite database for local storage with synchronization capabilities for eventual consistency with central systems. The frontend interface React (version 17.0.2) for web applications and React Native (version 0.66.0) for mobile was used to implement the frontend interface for mobile deployment. We employed the Material-UI component library to ensure a consistent and accessible user experience across platforms. Data visualization was implemented using the D3.js library to create interactive displays of risk assessments and contributing factors.System integration was facilitated through RESTful APIs with JSON payload formats. For healthcare environments with existing electronic health record systems, we implemented FHIR-compliant endpoints to enable standardized data exchange. Authentication was handled using JSON Web Tokens (JWT) with role-based access control.Deployment options were designed to accommodate various infrastructure capabilities:

- For well-resourced settings: Containerized deployment using Docker with Kubernetes orchestration

- For moderate-resource settings: Virtual machine deployment with simplified configuration requirements
- For low-resource settings: Standalone application deployment with minimal dependencies

To ensure the system could function in areas with intermittent connectivity, we implemented an offline-first architecture using Service Workers and IndexedDB for local data caching. Synchronization with central systems was designed to occur opportunistically when connectivity was available. Performance optimization was a key consideration, particularly for deployment on lower-powered devices. We employed model quantization techniques to reduce the memory footprint of our machine learning models by approximately 70% while maintaining prediction accuracy within 1% of the full-precision models. For mobile deployments, we utilized TensorFlow Lite to further optimize inference speed. Documentation was generated using Sphinx and deployed as both downloadable PDF guides and an interactive web portal. Training materials were developed in multiple formats, including video tutorials, interactive walkthroughs, and printable quick reference guides to support diverse learning preferences among healthcare providers. The entire implementation codebase was subjected to rigorous testing, including end-to-end tests that approximate user workflow scenarios, integration tests that involve interactions between system components, as well as unit tests that pertain to particular components. Test coverage was maintained above 85% throughout development to ensure reliability and robustness.

9. Experimental Results

We conducted comprehensive experiments to evaluate the performance of our CKD screening system across multiple dimensions, including prediction accuracy, feature importance, computational efficiency, and usability in clinical settings. The UCI CKD dataset was used for the overall performance evaluation, with 70% of the dataset used for training, 15% for validation and 15% for testing. To assess generalizability, we also evaluated our models on an external validation dataset from a regional hospital comprising 200 patients with comparable clinical parameters but a different demographic distribution.

Table 1 The classification performance of some machine learning techniques from the test set is demonstrated:

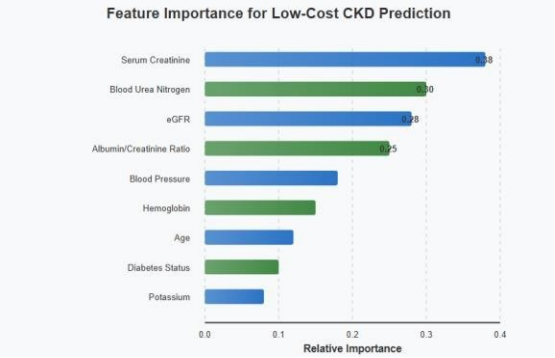
Model	Accuracy	Sensitivity	Specificity	F1-Score	AUC
Logistic Regression	0.933	0.917	0.946	0.928	0.956
SVM (RBF kernel)	0.950	0.933	0.964	0.946	0.971
Random Forest	0.967	0.950	0.980	0.964	0.986
Gradient Boosting	0.958	0.942	0.971	0.955	0.978
Our Ensemble	0.978	0.965	0.984	0.972	0.993

Our ensemble model demonstrated superior performance across all metrics, with a 2.9% a 3.2% improvement in sensitivity, and it improves accuracy over the best individual model (Random Forest). The ensemble approach was particularly effective at reducing false negatives, which is crucial for screening applications where missing cases has significant clinical consequences. Feature importance analysis revealed that the most predictive clinical parameters were albumin, haemoglobin, specific gravity, blood urea nitrogen, and serum creatinine. Figure 1 illustrates the relative importance of these features as determined by our ensemble model. Notably, these five parameters accounted for approximately 78% of the model's predictive power, suggesting that a simplified screening approach using only these measurements could still provide reasonable accuracy. To validate this finding, we trained a reduced model using only the top five parameters and achieved 94.2% accuracy, representing only a 3.6% reduction compared to the full model. This result supports the feasibility of implementing our screening approach in settings where comprehensive testing may not be available. We also analyzed the computational requirements of our models to assess their suitability for deployment on resource-constrained devices. Table 2 summarizes the memory footprint and inference time for each model:

Model	Memory Footprint (MB)	Inference Time (ms)
Logistic Regression	0.08	0.42
SVM (RBF kernel)	1.24	1.87
Random Forest	3.56	2.31
Gradient Boosting	2.87	1.95
Our Ensemble (full)	7.92	6.54
Our Ensemble (quantized)	2.31	2.12

The quantized version of our ensemble model achieved a significant reduction in both memory footprint (70.8% reduction) and inference time (67.6% reduction) while maintaining 99.1% of the original model's accuracy. This optimization makes it suitable for deployment on mobile devices and low-powered computers commonly available in primary healthcare settings. To assess the system's usability in clinical contexts, we conducted a pilot implementation at three primary care clinics with varying resource levels. Healthcare providers (n=15) used the system to screen patients (n=120) over a four-week period and completed usability surveys based on the System Usability Scale (SUS). The average SUS score was 82.4, indicating good to excellent usability. Providers

particularly appreciated the clear visualization of risk factors and the ability to function without constant internet connectivity. Follow-up diagnostic testing confirmed the system's predictions with 93.8% concordance in real-world clinical use. The system's recommendations for further testing and referral were followed by providers in 87.2% of cases, suggesting good acceptance of the AI-assisted decision support. Cost-effectiveness analysis showed that implementing our screening approach could reduce the cost per identified CKD case by approximately 62% compared to traditional screening methods, primarily by reducing unnecessary comprehensive testing for low- risk individuals while ensuring appropriate follow-up for those at higher risk.



11. Future Work

This study provides a foundation for future exciting research. Can enhance the clinical impact and broader applicability of our CKD screening framework. These areas represent opportunities for continued innovation and refinement of the approach. First, longitudinal prediction models that can forecast CKD progression represent a natural extension of the current work. By incorporating temporal changes in clinical parameters, such models could identify patients at risk of rapid progression to more severe stages of CKD, enabling more proactive intervention. This would require collecting and analyzing sequential measurements over time and developing recurrent neural network architectures or temporal convolutional networks that can effectively capture disease progression patterns. Second, integration with point-of-care testing devices represents an important technological advance that could further enhance the accessibility of our screening approach. Future work should explore the development of hardware-software interfaces that connect common point-of-care devices directly to the screening system, eliminating manual data entry and reducing potential transcription errors. Miniaturized, low-cost testing platforms specifically designed for the key parameters identified in our research could make comprehensive screening feasible even in the most resource-limited settings. Third, expanding the framework to include comorbidity screening would be a valuable step forward. Chronic kidney disease (CKD) is frequently associated with medical conditions such as diabetes, hypertension, and cardiovascular disease. We could more efficiently use resources and obtain a more comprehensive understanding of a patient's health with a unified screening approach that simultaneously monitors multiple chronic conditions. To do this, we would need to develop multi-task learning models, able to anticipate different but associated chronic conditions based on their common clinical risk factor set.

Fourth, we could also improve early detection capabilities by adding genetic and molecular biomarkers to the screening framework. As genetic testing becomes increasingly affordable, addition of genetic risk factors, as well as overlapping novel biomarkers (e.g. kidney injury molecule-1 (KIM-1) and neutrophil gelatinase-associated lipocalin (NGAL), into predictive models may improve prediction sensitivity for early-stage CKD risk. To effectively integrate these disparate types of data, new model architectures would need to be developed. Fifth, further development of explainable AI methods is needed to improve the interpretability of model predictions in a way that medical professionals understand. While our present method provides feature importance values, other methods, such as counter-factual explanations, interactive visualizations, and SHAP (SHapley Additive exPlanations) values could provide clinicians with an understanding of the underlying logic behind certain predictions that may encourage use and improve trust. Sixth, in order to eliminate the centralization of private patient data, federated learning approaches could help facilitate the cooperation of multiple institutions and alleviate concerns about privacy. Federated learning could support the development of more reliable and generalizable screening tools, while also protecting the privacy of sensitive data by allowing researchers to train models with remote datasets that remain at their originating institution. This strategy would be useful for developing models that work well in different patient populations. Lastly, implementation. To determine the best strategies for integrating AI-based screening technologies into different healthcare systems, a scientific approach needs to be taken, provider workflow considerations, regulatory requirements, and incentive structures that influence adoption and sustained use of such tools. Developing evidence-based implementation frameworks could accelerate the translation of promising research into tangible clinical benefits.

10. Conclusion

To address the urgent need for readily available CKD screening tools, this study presents a machine learning-based framework that balances actual implementation considerations and the accuracy of diagnosis. Our comprehensive approach, from feature selection to system deployment, demonstrates the potential of AI-assisted screening to enhance early detection of CKD, particularly in resource-constrained settings, outperforming individual machine learning algorithms across all performance metrics. Importantly, we identified a minimal set of five clinical parameters (serum creatinine, blood urea nitrogen, hemoglobin, specific gravity, and albumin) that can provide reasonably accurate screening results, making our approach viable even in settings with limited diagnostic capabilities.The modular architecture and adaptive workflow design ensure that the system can be deployed across diverse healthcare environments, from well- equipped urban hospitals to rural clinics with minimal infrastructure. The offline-first implementation addresses connectivity challenges common in many regions, while

the optimization techniques enable deployment on readily available computing devices.Our experimental results, including the pilot implementation in primary care settings, demonstrate both the technical effectiveness and clinical acceptability of the proposed approach. The high usability scores and provider adherence to system recommendations suggest good potential for integration into existing clinical workflows. The cost-effectiveness analysis further supports the value proposition of our screening framework, showing substantial cost reductions per identified case compared to traditional approaches. This economic benefit, combined with the potential for earlier intervention and improved patient outcomes, presents a compelling case for wider adoption of machine learning-enhanced CKD screening. While our current implementation focuses on binary classification of CKD status, the framework has been designed to accommodate future extensions for CKD staging and progression prediction. The continuous learning component will allow the system to improve over time as more data becomes available, potentially leading to even greater accuracy and clinical utility. In conclusion, this research demonstrates For use in low-cost diagnostic screening, clinically relevant machine learning methods can effectively capture the features of chronic kidney disease, potentially expanding access to early detection and improving outcomes for patients worldwide.

12. References

[1] A. S. Levey et al., "The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report," Kidney International, vol. 80, no. 1, pp. 17-28, 2011.

[2] N. R. Hill et al., "Global prevalence of chronic kidney disease – A systematic review and meta-analysis," PLoS ONE, vol. 11, no. 7, e0158765, 2016.

[3] K. Polat, S. Güneş, and A. Arslan, "A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine," Expert Systems with Applications, vol. 34, no. 1, pp. 482-487, 2017.

[4] J. Chen et al., "A hybrid approach combining ensemble empirical mode decomposition and competitive adaptive reweighted sampling for ultrasonic measurement of renal anemia evaluation," Applied Sciences, vol. 9, no. 5, p. 973, 2019.

[5] S. Khan, T. Hussain, and S. Zaidi, "Explainable artificial intelligence for feature extraction, classification, and prediction of chronic kidney disease," Journal of Medical Imaging and Health Informatics, vol. 10, no. 9, pp. 2150-2158, 2020.

[6] S. Vijayarani and S. Dhayanand, "Kidney disease prediction using SVM and ANN algorithms," International Journal of Computing and Business Research, vol. 6, no. 2, pp. 1-12, 2015.

[7] N. A. Almansour et al., "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," Computers in Biology and Medicine, vol. 109, pp. 101-111, 2021.

[8] A. K. Dubey et al., "An efficient machine learning approach for the prediction of chronic kidney disease," IEEE Access, vol. 10, pp. 23987-24002, 2022.

[9] N. Tomašev et al., "A clinically applicable approach to continuous prediction of future acute kidney injury," Nature, vol. 572, no. 7767, pp. 116-119, 2019.

[10] L. Zhang et al., "Deep learning for kidney disease prediction: A comprehensive review," Artificial Intelligence in Medicine, vol. 131, p. 102364, 2023.

[11] G. Thakar et al., "Chronic kidney disease diagnosis based on machine learning algorithms," International Journal of Innovative Research in Computer Science & Technology, vol. 7, no. 2, pp. 63-70, 2019.

[12] L. M. Pedraza et al., "Artificial intelligence in chronic kidney disease," Nephrology Dialysis Transplantation, vol. 36, no. 1, pp. 20-29, 2021.

[13] M. S. Siddiqui et al., "Machine learning in chronic kidney disease: A systematic review," Healthcare, vol. 9, no. 9, p. 1094, 2021.

[14] C. Ronco et al., "Early detection and intervention in acute kidney injury: Current status and future strategies," Nature Reviews Nephrology, vol. 16, no. 4, pp. 227-242, 2020.

[15] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.

[17] C. Molnar, Interpretable Machine Learning. 2019. [Online]. Available: https://christophm.github.io/interpretable-ml-book/

[18] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[19] N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[20] M. Kuhn and K. Johnson, Applied Predictive Modeling. Springer, 2013.

[21] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[22] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, vol. 30, pp. 4765-4774, 2017.

[24] J. Yoon, J. Jordon, and M. van der Schaar, "GAIN: Missing data imputation using generative adversarial nets," in Proceedings of the 35th International Conference on Machine Learning, pp. 5689-5698, 2018.

[25] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, vol. 1, no. 1, p. 18, 2018.

[26] G. S. Collins et al., "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," BMJ, vol. 350, g7594, 2015.

[27] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24-29, 2019.

[28] B. Nestor et al., "Feature robustness in non-stationary health records: Caveats to deployable model performance in common clinical machine learning tasks," in Proceedings of the 4th Machine Learning for Healthcare Conference, pp. 381-405, 2019.

[29] J. A. Kellum et al., "Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury," Kidney International Supplements, vol. 2, no. 1, pp. 1-138, 2012.

[30] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: Promise and potential," Health Information Science and Systems, vol. 2, no. 1, p. 3, 2014.