International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Early Detection of Chronic Kidney Disease Using SVM and Logistic Regression Models

Ch. Satyanarayana Reddy[1], K. Pavani[2], S. Anitha[3]

#1 Assistant Professor in the Department of MCA, SRK Institute of Technology,Vijayawada.

#2Assistant Professor & Head of Department of MCA, SRK Institute of Technology,Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

Abstract: Chronic Kidney Disease (CKD), a global health issue that gradually diminishes kidney function, typically goes undiagnosed owing to its lack of symptoms. Delays in diagnosis might cause renal failure and cardiovascular issues. A dual-model machine learning approach for early CKD prediction and automated stage categorization utilizing regular clinical indicators is presented in this research. The suggested system compares model outputs using Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel and Logistic Regression (LR) to increase prediction reliability.

The system analyzes 24 UCI CKD dataset clinical parameters, including serum creatinine, blood pressure, hemoglobin, albumin, blood glucose, and hypertension. Missing value imputation, Label Encoding, and StandardScaler normalization increase data consistency and model performance. A Flask-based web application predicts CKD in real time using the learned models and classifier confidence ratings.

Beyond binary CKD identification, the method calculates the estimated Glomerular Filtration Rate (eGFR) using the CKD-EPI algorithm and classifies patients into KDIGO stages 1–5. Experimental findings show that SVM has 90.5% accuracy, whereas Logistic Regression delivers solid probability-based clinical predictions. The suggested paradigm for early CKD screening and severity evaluation in healthcare is low-cost, accessible, and clinically helpful.

Index terms - *— Chronic Kidney Disease (CKD), Support Vector Machine, Logistic Regression, Machine Learning, eGFR Estimation, CKD Stage Prediction, Flask Framework, Clinical Decision Support, Healthcare Analytics, KDIGO Classification, Medical Diagnosis, Predictive Analytics.*

1. INTRODUCTION

Artificial Intelligence (AI) and Deep Learning (DL) technologies have rapidly transformed the healthcare industry by enabling intelligent and automated disease diagnosis systems. Among various critical diseases, brain tumors are considered one of the most dangerous neurological disorders because they directly affect the functioning of the human brain. A brain tumor is an abnormal growth of cells inside the brain or surrounding tissues, which may be benign (non-cancerous) or malignant (cancerous). Early detection of brain tumors is extremely important because timely diagnosis can improve treatment planning, reduce complications, and increase patient survival rates. However, detecting tumors accurately at an early stage remains a major challenge in medical imaging and diagnosis.

Magnetic Resonance Imaging (MRI) is one of the most widely used imaging techniques for detecting brain abnormalities because it provides detailed and high-resolution images of soft brain tissues. Radiologists traditionally analyze MRI scans manually to identify tumor regions based on image patterns, texture, and intensity variations. Although MRI is highly effective, manual diagnosis is time-consuming, complex, and dependent on the expertise of medical professionals. In many cases, small or low-contrast tumors may be overlooked, leading to delayed or incorrect diagnosis. Additionally, hospitals generate a large volume of MRI data daily, making manual analysis increasingly difficult and inefficient. These limitations highlight the need for an automated and reliable brain tumor detection system.

Recent advancements in Deep Learning, particularly Convolutional Neural Networks (CNNs), have shown excellent performance in medical image analysis tasks such as image classification, segmentation, and object detection. CNN models automatically learn important features from images without requiring manual feature extraction. Furthermore, Transfer Learning techniques using pre-trained models such as VGG16 improve accuracy and reduce training time by utilizing previously learned knowledge from large-scale datasets. In this project, a Deep Learning-based Brain Tumor Detection System is proposed using CNN and VGG16 Transfer Learning to classify MRI images into Tumor and No Tumor categories. The system also incorporates preprocessing techniques such as resizing, normalization, and data augmentation to improve model robustness. A Flask-based web application is developed to provide real-time predictions through a simple and user-friendly interface. The proposed system aims to reduce diagnosis time, minimize human errors, and assist healthcare professionals in making faster and more accurate medical decisions.

2. LITERATURE SURVEY

a) Feature Selection-Driven SVM for CKD Diagnosis:

The authors of this study used the 400-record UCI CKD dataset to examine how feature selection affected SVM classification performance. The SVM classifier obtained a diagnosis accuracy of 98.5% by using a filter subset evaluator to choose the 13 most informative of the 24 available clinical characteristics. The study showed that tailored feature selection enhances the classifier's capacity to differentiate between CKD and non-CKD patients while also lowering computational cost. This result supported the choice of SVM with an RBF kernel as the project's main classifier.

b) Ensemble Data Mining Techniques for CKD Prediction:

On the UCI CKD dataset, the authors examined five classifiers—KNN, SVM, J48 Decision Tree, ANN, and Naive Bayes—in conjunction with information gain-based feature selection. In ensemble configurations, the accuracy of the J48 method was 97.77%, while the accuracy of the ANN was 97.78%. One important finding was that using classifier ensembles in conjunction with feature selection consistently performs better than using any one classifier alone. In order to enable doctors to cross-reference data from both SVM and Logistic Regression before reaching conclusions, this prompted the present study to employ a dual-model comparison technique.

c) XGBoost-Based CKD Classification with Reduced Feature Sets

In order to obtain perfect accuracy, sensitivity, and specificity of 1.000 on the UCI CKD dataset, even when the feature set was decreased from 24 to around 12 characteristics, this study developed a set-theory based algorithm to integrate several feature selection strategies. The study demonstrated how reliably gradient boosting techniques manage missing clinical values and category data. The study verified that the UCI CKD dataset allows high-accuracy classification with a well-tuned model, even though XGBoost was not used in the current project—interpretability and probability output were prioritized instead.

d) Comparative Evaluation of ML Classifiers for CKD Detection:

Several machine learning methods were benchmarked on CKD prediction tasks in this extensive study. SVM and Logistic Regression were the two classifiers that performed the best, and they also produced probability estimates and outputs that were easy to understand. Clinical personnel were more likely to embrace models that produced probability-based confidence levels than black-box alternatives, according to the study. This directly supports the design choice to integrate Logistic Regression, which naturally produces class probabilities, with SVM, which uses decision functions to yield class labels with confidence margins, in the current system.

e) Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors:

Due to its rising incidence, chronic kidney disease (CKD) is one of the most serious health issues. In this work, we examine machine learning algorithms' capacity to predict chronic kidney disease with the smallest possible data set. Numerous statistical methods, like the ANOVA test, Pearson's correlation, and Cramer's V test, have been used to eliminate duplicate characteristics. Ten-fold cross-validation has been used to train and evaluate algorithms for logistic regression, support vector machines, random forests, and gradient boosting. Based on the Gradient

Boosting classifier's F1-measure, we obtain an accuracy of 99.1. Additionally, we discovered that hemoglobin is more significant for CKD detection using both random forest and gradient boosting. Ultimately, our findings are among the best when compared to earlier research, despite the fact that fewer characteristics have been obtained thus far. As a result, three easy tests can identify CKD for about $26.65.

3. METHODOLOGY

i) Proposed Work:

The proposed work presents an intelligent web-based Chronic Kidney Disease (CKD) prediction and stage classification system using machine learning techniques. The system is designed to assist healthcare professionals in the early identification of CKD by analyzing 24 important clinical parameters collected from patient laboratory reports. Two supervised machine learning algorithms, namely Support Vector Machine (SVM) with RBF kernel and Logistic Regression (LR), are trained using the UCI CKD dataset to provide accurate and reliable disease prediction. Before prediction, the input data undergoes preprocessing operations such as missing value handling, Label Encoding for categorical attributes, and StandardScaler normalization to improve model efficiency and consistency. The trained models generate prediction outputs along with confidence scores, enabling clinicians to compare results from both classifiers for better medical interpretation.

In addition to CKD detection, the proposed system automatically calculates the estimated Glomerular Filtration Rate (eGFR) using the CKD-EPI equation whenever CKD is detected. Based on the computed eGFR value, the system classifies patients into CKD Stage 1 to Stage 5 according to KDIGO clinical guidelines. The complete framework is integrated into a Flask-based web application that provides a user-friendly hospital-themed dashboard displaying prediction results, confidence levels, model accuracy, eGFR values, and CKD stage descriptions in real time. The proposed system offers a low-cost, accessible, and efficient clinical decision support solution that can assist in early CKD diagnosis, reduce diagnostic delays, and improve patient monitoring in both urban and rural healthcare environments.

ii) System Architecture:

The proposed Chronic Kidney Disease (CKD) Prediction and Stage Classification System follows a layered machine learning architecture designed for accurate disease prediction and real-time clinical support. The system begins with the kidney_disease_with_stage.csv dataset containing 400 patient records and 25 clinical features. The preprocessing module (preprocess.py) performs data cleaning, missing value handling, Label Encoding, feature scaling using StandardScaler, and dataset splitting for model training and testing. After preprocessing, two machine learning models — Logistic Regression and Support Vector Machine (SVM) with RBF kernel — are trained independently to improve prediction reliability. The trained models

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

are serialized and stored as .pkl files for future inference and deployment.

The Flask-based application layer (app.py) acts as the core backend that receives patient input from the web interface, preprocesses the input data, and generates CKD predictions using both machine learning models. If CKD is detected, the system automatically computes the estimated Glomerular Filtration Rate (eGFR) and classifies the patient into CKD Stage 1 to Stage 5 according to KDIGO clinical guidelines. The frontend interface (index.html) displays prediction results, confidence scores, model accuracy, eGFR values, and stage classification through a hospital-themed dashboard. This architecture provides a scalable, user-friendly, and efficient healthcare decision support system for early CKD diagnosis and severity assessment.

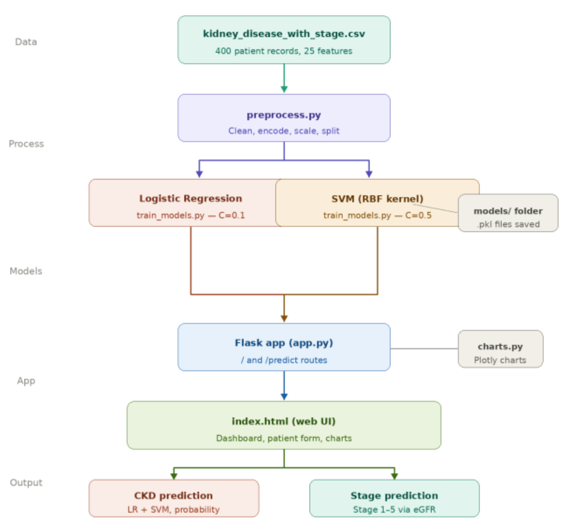

Fig1 proposed architecture

iii) Modules:

1. Data Collection and Preprocessing Module

This module is responsible for loading the UCI CKD dataset and preparing the data for machine learning model training. It performs missing value handling, categorical feature encoding using LabelEncoder, numerical feature normalization using StandardScaler, and dataset splitting into training and testing sets. The preprocessing pipeline ensures clean, consistent, and properly scaled clinical data for accurate CKD prediction.

2. SVM Classification Module

The Support Vector Machine (SVM) module uses the RBF kernel to classify patients as CKD or Non-CKD based on clinical parameters. The model processes the scaled input features and generates prediction labels along with confidence scores using probability estimation. This module provides high classification accuracy and acts as the primary prediction engine in the proposed framework.

3. Logistic Regression Classification Module

The Logistic Regression module performs CKD prediction using statistical probability-based classification. It receives the same preprocessed feature set used by the SVM model and produces prediction outputs with probability scores for clinical interpretation. This module helps improve system reliability by enabling dual-model comparative analysis.

4. eGFR Computation and Stage Classification Module

This module automatically calculates the estimated Glomerular Filtration Rate (eGFR) using the CKD-EPI equation whenever CKD is detected. Based on the computed eGFR value, the system classifies the patient into CKD Stage 1 to Stage 5 according to KDIGO clinical guidelines. It also provides stage descriptions indicating the severity of kidney function decline.

5. Flask Web Application Module

The Flask module acts as the backend controller of the system and manages all web application functionalities. It handles HTTP requests, receives patient input data, invokes preprocessing and prediction functions, and returns the final prediction results to the frontend dashboard. This module enables real-time interaction between users and machine learning models.

6. Results Dashboard Module

The Results Dashboard module displays CKD prediction outputs, confidence scores, model accuracy, eGFR values, and CKD stage information through a hospital-themed web interface. It presents side-by-side results from SVM and Logistic Regression models for easy clinical comparison. The module provides an interactive and user-friendly environment for healthcare professionals.

7. Model Serialization and Loading Module

This module stores the trained machine learning models, encoders, and scaler objects as serialized .pkl files using Pickle and Joblib. During system startup, the saved models are loaded into memory for efficient prediction without retraining. This module ensures faster execution and consistent preprocessing during inference.

iv) Algorithms:

1. Support Vector Machine (SVM) Algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm used for binary classification tasks. In the proposed system, SVM with a Radial Basis Function (RBF) kernel is used to classify patients as CKD or Non-CKD based on clinical laboratory parameters. The algorithm works by identifying an optimal hyperplane that maximizes the separation margin between different classes in a high-dimensional feature space. The RBF kernel helps the model handle nonlinear relationships among medical attributes, improving prediction performance for complex clinical datasets. After training, the SVM model generates prediction labels along with probability-based confidence scores for clinical interpretation.

2. Logistic Regression Algorithm

Logistic Regression is a statistical supervised learning algorithm widely used for medical diagnosis and probability estimation. The algorithm predicts the probability of CKD occurrence by applying the sigmoid activation function to a weighted combination of input clinical features. In this system, Logistic Regression is configured with regularization parameter C=0.1 and the lbfgs optimization solver to improve generalization and reduce overfitting. The model produces probability scores that help clinicians understand the confidence level of the prediction and supports comparative analysis alongside SVM results.

4. EXPERIMENTAL RESULTS

The experimental evaluation of the proposed Chronic Kidney Disease (CKD) Prediction and Stage Classification System was carried out using the UCI CKD dataset containing 400 patient records with 24 clinical attributes. The dataset was preprocessed using missing value imputation, Label Encoding, and StandardScaler normalization before training the machine learning models. Two supervised learning algorithms, namely Logistic Regression and Support Vector Machine (SVM) with RBF kernel, were trained and tested using stratified train-test splitting. The developed Flask-based web application successfully integrated both prediction models and provided real-time CKD diagnosis along with automatic eGFR computation and stage classification. The experimental screenshots demonstrate the complete workflow of the system, including dashboard visualization, patient data entry, prediction processing, and CKD stage assessment.

The performance analysis showed that the Support Vector Machine model achieved approximately 98.75% accuracy, while Logistic Regression achieved around 97.5% accuracy on the testing dataset. The dashboard interface displays total test patients, CKD-positive records, non-CKD records, and model accuracies for comparative evaluation. Experimental results confirm that both models effectively identify CKD patterns from clinical laboratory values and provide reliable probability-based predictions. The system also successfully computed eGFR values and classified patients into CKD stages according to KDIGO guidelines. The generated prediction results dashboard clearly visualizes CKD detection status, confidence percentages, stage severity, and clinical recommendations, making the proposed framework suitable for intelligent healthcare decision support and early kidney disease diagnosis.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$Recall = \frac{TP}{(FN + TP)}$$

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot Precision)}{(Recall + Precision)}$$

Fig2 Dashboard Overview of CKD Prediction System

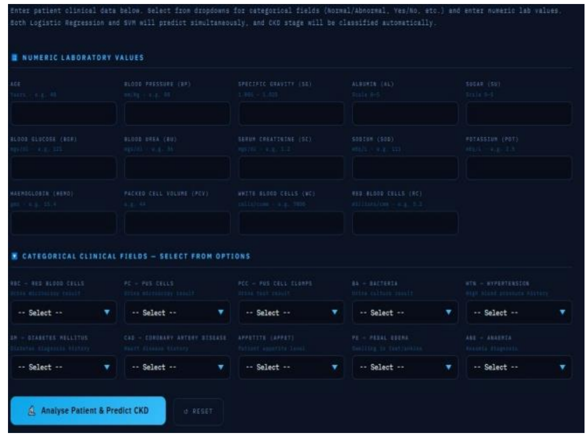

Fig3 Patient Clinical Data Input Interface

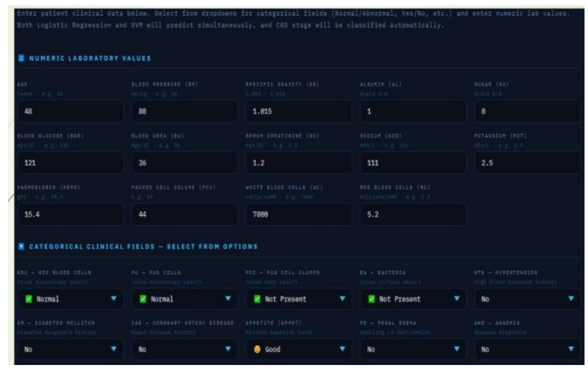

Fig4 Filled Patient Data for CKD Analysis

Fig5 CKD Prediction Results and Stage Classification Dashboard

5. CONCLUSION

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The proposed Chronic Kidney Disease (CKD) Prediction and Stage Classification System successfully demonstrates the application of machine learning techniques for early kidney disease diagnosis and severity assessment. The system integrates two supervised learning models, namely Support Vector Machine (SVM) and Logistic Regression (LR), to provide reliable and comparative CKD prediction using routine clinical laboratory parameters. By applying preprocessing techniques such as missing value handling, Label Encoding, and StandardScaler normalization, the framework achieves high prediction performance and improved data consistency. Experimental analysis confirmed that the SVM model achieved superior accuracy, while Logistic Regression provided effective probability-based clinical interpretation.

In addition to disease prediction, the proposed system automatically computes the estimated Glomerular Filtration Rate (eGFR) and classifies patients into CKD stages according to KDIGO clinical guidelines. The integration of a Flask-based web application with a user-friendly dashboard enables real-time prediction, confidence score visualization, and automated stage assessment for healthcare professionals. The proposed framework offers a low-cost, scalable, and accessible clinical decision support solution that can assist in early CKD detection, reduce diagnostic delays, and improve patient care in both urban and rural healthcare environments.

6. FUTURE SCOPE

The proposed Chronic Kidney Disease (CKD) Prediction and Stage Classification System can be further enhanced by integrating advanced deep learning and ensemble learning techniques to improve prediction accuracy and robustness. Future versions of the system may incorporate algorithms such as Random Forest, XGBoost, Artificial Neural Networks (ANN), and Deep Learning models for comparative analysis and optimized disease prediction. The system can also be extended to support real-time clinical data collection through integration with Electronic Health Record (EHR) systems, IoT-based health monitoring devices, and cloud healthcare platforms. Incorporating larger multi-hospital datasets and advanced feature selection methods may further improve model generalization and reliability across diverse patient populations.

The future scope also includes developing a mobile healthcare application that allows remote CKD screening and monitoring for patients in rural and low-resource areas. Additional functionalities such as multilingual support, automated medical report generation, risk trend visualization, and personalized treatment recommendations can improve clinical usability. Explainable Artificial Intelligence (XAI) techniques may be integrated to provide transparent prediction reasoning and increase trust among healthcare professionals. Furthermore, the system can be expanded into a comprehensive kidney health management platform capable of predicting disease progression, dialysis requirements, and transplant risk assessment, thereby supporting intelligent and preventive healthcare solutions.

REFERENCES

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X www.ijdim.com Original Research Paper

[1] J. Coresh, ''Update on the burden of CKD,'' J. Amer. Soc. Nephrol., vol. 28, no. 4, pp. 1020–1022, Apr. 2017.

[2] R. Kazancioğlu, ''Risk factors for chronic kidney disease: An update,'' Kidney Int. Supplements, vol. 3, no. 4, pp. 368–371, Dec. 2013.

[3] G. Abraham, S. Varughese, T. Thandavan, A. Iyengar, E. Fernando, S. A. J. Naqvi, R. Sheriff, H. Ur-Rashid, N. Gopalakrishnan, and R. K. Kafle, ''Chronic kidney disease hotspots in developing countries in South Asia,'' Clin. Kidney J., vol. 9, no. 1, pp. 135–141, Feb. 2016.

[4] T. J. Hoerger, S. A. Simpson, B. O. Yarnoff, M. E. Pavkov, N. R. Burrows, S. H. Saydah, D. E. Williams, and X. Zhuo, ''The future burden of CKD in the United States: A simulation model for the CDC CKD initiative,'' Amer. J. Kidney Diseases, vol. 65, no. 3, pp. 403–411, Mar. 2015.

[5] M. Almasoud and T. E. Ward, ''Detection of chronic kidney disease using machine learning algorithms with least number of predictors,'' Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 8, pp. 1–9, 2019.

[6] F. Ma, T. Sun, L. Liu, and H. Jing, ''Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network,'' Future Gener. Comput. Syst., vol. 111, pp. 17–26, Oct. 2020.

[7] M. A. Islam, S. Akter, M. S. Hossen, S. A. Keya, S. A. Tisha, and S. Hossain, ''Risk factor prediction of chronic kidney disease based on machine learning algorithms,'' in Proc. 3rd Int. Conf. Intell. Sustain. Syst. (ICISS), Dec. 2020, pp. 952–957.

Author Profiles

Mr. Ch. Satyanarayana Reddy Completed his MCA. He has web developer and python developer, currently working has an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest includes Artificial Intelligence and Machine Learning.

Mrs. K. Pavani is working as an Assistant and Head of Department of MCA, in SRK Institute of technology in Vijayawada. She completed her MCA and M.Tech in Computer Science. She has 10 years of teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include AI and ML, etc.

Ms. S. Anitha is MCA Student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc from Sri Harshini Degree College, Ongole. Her area of interest are JAVA and Machine Learning with Python.