Fog Based AI Framework for Accurate Early Detection Of Chronic Kidney Disease

[1]M. SriRaghavendra, [2]S.Md. Shakeer, [3]P. Kavya Sri, [4]K. Tanujasree, [5]S. Athika

[1] sr.meeniga@gmail.com, [2]shakeercsedscseds@rgmcet.edu.in, [3]21091A3217@rgmcet.edu.in, [4]21091A3258@rgmcet.edu.in, [5]21091A3206@rgmcet.edu.in

(1,2)Department of Computer Science and Engineering (Data Science), Rajeev Gandhi Memorial College of Engineering and Technology (Autonomous), Nandyal, Andhra Pradesh, India. 3,4,5 students at RGM College of Engineering and Technology (Data Science Department), Nandyal, Andhra Pradesh, India.

ABSTRACT:

CKD is a global wise health concern that refers for regular diagnosis for best patient outcomes. This research study presents a Fog-based AI framework based on Edge-cloud computing and Machine Learning for real-time prediction and detection of Chronic Kidney Disease. This study refers the models of XGBoost & Random Forest, which will improve the accuracy detection of diseases by capturing difficult patterns with clinical data. The explainable aspect of AI is the way of enhancing the trustworthiness in the diagnostic process. This framework presents a decentralized architecture that will helps in low latency and cloud storage issues. And maximizes efficient use of computational resources. The performance evaluation results describes that the framework is scalable and robust, makes the framework for use within large scale society. Moreover, this framework has the potential to significantly improve the early detection of CKD outcomes in medical setting and promote the access of healthcare on the global scale.

INDEX TERMS:

Chronic Kidney Disease (CKD), Fog Computing, Artificial Intelligence, Machine learning, healthcare Analytics, Scalability, Robustness, Cloud Computing vs Fog Computing.

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive and chronic debilitating disease that affects millions of people and is a major public health burden [1]. CKD is a characterized by progressive renal dysfunction, which leads to a progressive decline in kidney function over time. CKD may progress to end stage renal disease (ESRD) that leads to dialysis or kidney transplantation. Early detection of CKD is critical to prevent progression and reduce healthcare costs, while improving patient outcomes. Current diagnostic methods rely largely on centralized cloud-based computing and laboratory assessments that suffer from latency problems, increased network congestion, and potential data security threats. Traditional cloud-based AI architectures are wired and therefore, suffer delays caused by network bottlenecks which makes them ineffective for urgent clinical use of AI. As a result, the incorporation of Fog Computing and Artificial Intelligence (AI) is emerging as a revolutionary function in diagnostics of healthcare systems. With fog, the cloud is able to perform additional tasks; it allows for the processing of data at the edge of the network. This considerably improves latency Issues and enables better and faster analysis. This model is especially useful in developing regions that have limited access to high-speed internet and established cloud infrastructure. At the same time, AI positively implemented in health care prediction analysis of detecting both precise and projection of diseases.

This study enhance both Fog computing and Ai implement each other creating smarter health care systems paving the way for better disease prediction, focused patient care, and clinical decision-making in managing CKD. The fog-based AI Framework represents a new breakthrough in creating technology to identify CKD. AI framework with decentralized fog computing and Explainable AI models can deliver a better framework for early disease identification and provides better scalable.

The above study will introduce new ideas, with implications for AI mediated diagnosis and patient and health outcomes use cases for CKD patients globally. The conceptualization resulting from the step and the framework concept is applicable to many resource constrained environments to demonstrate a framework designed to create efficiencies because it modernizes how we diagnostics examine CKD.

Our fog-based AI framework uses decentralized computing to detect CKD in real-time, maintaining low latency and scalability while maintaining low latency. To boost the performance of the model, data preprocessing addresses missing values, manages outliers, and encodes categorical features. We employ advanced classification models like XGBoost, Random Forest, SVM and Linear Regression, all of which are extremely accurate and deliver F1-Scores that are impressive.

II LITERATURE SURVEY

[1] Sanmarchi et al. reviewed machine learning applications to predict, diagnose, and manage CKD. The review considered 68 studies and noted such models as neural networks and random forests. The overall accuracy was acceptable, but generalizability and fairness were presented as predictors. A more interpretable and ethical approach to model use in health settings is emphasized by the authors.

K M Tawsik Jawad[2]and and his colleagues developed a model that attempts to predict Chronic Kidney Disease (CKD) utilizing explainable artificial intelligence and ensemble learning models. Their argument resides in the fact that no action will be taken if the risk factors of CKD are not shown. Among the several factors, few of the well known chronic disease is CKD which is associated with diabetes and hypertension. Predictive models in nephrology were constructed using validated kidney samples drawn from blood and urine. Incorporation of AI in the models enables Random Forest and XG Boost algorithms to outperform the rest..

Kaur et al. [3] conducted a review of literature on applications of machine learning in early detection and prediction of chronic kidney disease (CKD). The article brings to light that CKD is likely to remain undetected in its initial stages as it is symptomless and hence needs advanced diagnosis tools. Authors reaffirm that machine learning models, particularly ensemble models, have greater accuracy for the detection of CKD than traditional diagnostic methods. Their findings suggest that the integration of machine learning strategies into clinical practise can enhance early detection, simply patient management.

AL-Kateb and Abdulla [4] conducted a review of the literature about the convergence of IoT (Internet of Things), fog computing, and machine learning to detect CKD (Chronic Kidney Disease) early. The author believes that Ada Boost machine learning with fog computing could cause an improved prediction of CKD in real time, excellent returned efficiency, and reduced response times. Author concludes that this framework could show early CKD, decrease energy consumption, and improve resource efficiency over network resources which represent a good solution for health applications.

Thapa and colleagues [5] examined the potential of machine learning (ML) and explainable AI (XAI) methods to predict chronic kidney disease (CKD) other authors may have produced comparable prediction accuracy but established neither trust nor support for adoption across health care settings, and they also inconsistently reported their accuracy measures. The current study applied LIME as an intial effort to study partially – explanatory type models to produce accountability of the models and support clinical decision making.

Islam et al. [6] explored the use of predictive modeling techniques to test the efficiency of machine learning algorithms in predicting CKD. The authors considered a total of 12 supervised learning classifiers, and it was the XgBoost classifier that achieved the highest accuracy. The study also stressed the importance of predictive modeling due to its capability to provide insight into the relationship within data and enable better CKD diagnosis.

III SYSTEM ARCHITECTURE:

The system architecture is crafted as a well-organized pipeline specifically aimed at detecting chronic kidney disease (CKD). It all kicks off with an online dataset that undergoes data preprocessing, which involves tackling missing values and balancing the dataset to ensure we have top-notch input data. Next, we split the data into training and testing sets, applying a variety of classification algorithms like KNN, EXT, GB, RF, DT, GNB, XGB, and ADB. During the training phase, our focus is on fine-tuning the model, while the testing phase is dedicated to assessing its performance. To enhance accuracy, we employ an ensemble method with a voting classifier that merges several models for improved predictions. The evaluation results reveal that KNN and EXT achieved an outstanding accuracy of 99%, highlighting the system's effectiveness in CKD detection.

The Framework encapsulates a variety of machine learning techniques to further advance and build trustworthy predictability to reduce misclassification
and enhance earlier detections in patients. We have also expressed the intentional use of ensemble methods to not only provide predictive performance but to also reduce bias in the relative predictions of each model contained in the ensemble. On the topic of future applications we indicated similarity the generalization beyond the development and applicability of the framework, to the generalization of how applicable is this process in a real health care, and transference to clinical practice in other domain. Such future applications of the framework would involve future work with the IoT – Based live monitoring of patients, involvement of explainable – AI models which take in to account different patients demographics. These methods make future ongoing monitoring of patients, and AI – Based diagnosis.

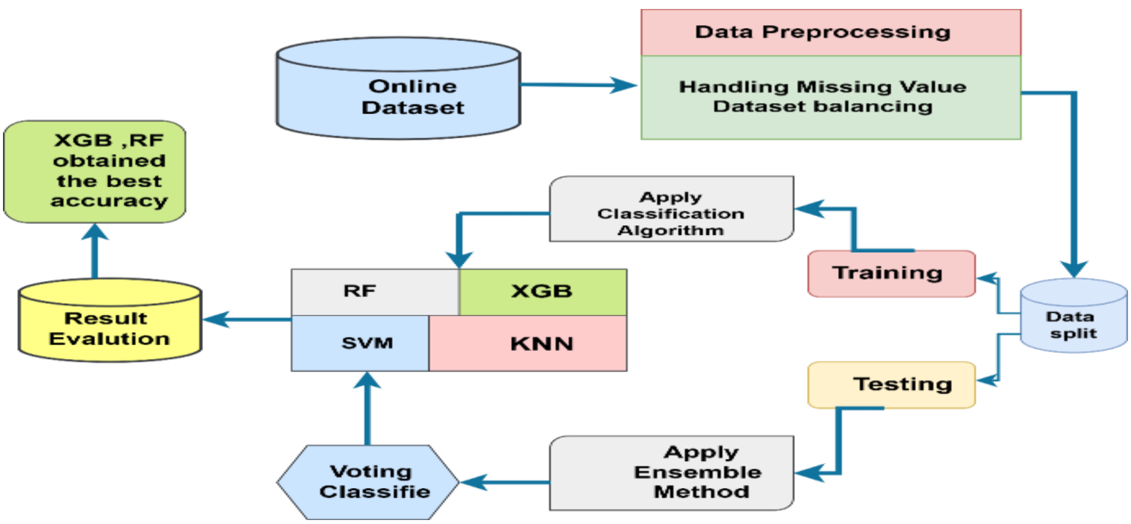

Fig 1: Overview Of System Architecture

IV METHODOLOGY:

Our proposed framework processes patient data at the edge of the network, minimizing the dependency on centralized cloud infrastructure. Our system is comprised of multiple layers, including a data acquisition layer for collecting biomedical parameters, a fog computing layer for pre-processing and real-time analysis of data, and an AI-based diagnostic layer that utilizes machine learning models such as Support vector machines (SVM), Random Forest, and Deep Learning to accurately classify and predict CKD.

The method accurately diagnoses Chronic Kidney Disease (CKD) by utilizing a range of machine learning algorithms. For example, the well-known feature selection and boosting capabilities of XGBoost (Extreme Gradient Boost) significantly improve classification performance. Next is the Random Forest (RF) ensemble. This learning method reduces overfitting and increases prediction reliability by averaging the output of multiple decision trees. By determining the optimal decision boundary in intricate, high-dimensional spaces, Support Vector Machines (SVM) can also be used to categorize chronic kidney disease (CKD). A simple yet powerful baseline model for binary classification is logistic regression. Additionally, to improve accuracy and robustness, an ensemble learning approach combines several models. By eliminating several attributes, feature selection techniques help to improve the model's overall performance. The system's ability to adjust to new data through ongoing learning mechanisms gradually boosts its dependability. Together, these models improve predictive performance in real-time healthcare settings, guaranteeing timely, accurate, and understandable CKD detection.

Dataset Description :

A. UCI Chronic Kidney Disease Dataset: The UCI dataset is a popular, fascinating, and humorous dataset, was made public by UCI on July 2, 2015. The CKD dataset includes 25 characteristics, including blood pressure, age, blood gravity, albumin, blood, urea serum creatine hemoglobin, potassium, sodium, and blood cell counts. Over the course of two months, these data were collected.

B. The Dataset of Nephrology Clinical Data: Dataset of Nephrology Clinical Data: Real-time data is collected from electronic health records of hospitals and using internet of things-enabled devices to generate a database of chronic kidney disease patients' monitoring data. Dataset includes time-series of vitals for patients, including blood pressure, glucose, and creatinine, as well as early signs of kidney failure.

Table 1 : Dataset details

Dataset Name	No. of instances	No. Of Features
CKD Dataset (UGIRepository)	400	25

Table 2: Feature Details

Feature Name	Description	Type
Age	Age of the patient	Numerical
Blood Pressure	Systolic Blood Pressure(mmHg)	Numerical
Hemoglobin	Hemoglobin concentration(g/dL)	Numerical
CKD	Target variable (CKD Presence)	Categorical (Yes/No)

i. Data Collection:

Data collection in the context of the Fog-Based AI Framework for the early detection of CKD, data collection refers to the gathering of relevant and complete patient information. Typically, clinical data is collected including patient demographics, medical history, laboratory test results (e.g., serum creatinine, blood urea nitrogen and urine analysis) and other health parameters that are related to kidney function. This may involve collaborating with healthcare institutions to access electronic health records (EHRs) or surveys and interviews with patients to collect first-hand data. The quality and accuracy of the collected data is very important as it directly impacts the performance of the machine learning models.

ii. Data Preprocessing:

Data preprocessing CKD detection framework data cleaning identifies and removes errors, missing values, and outliers in the patient data. Imputation techniques can be used to impute missing values. To improve the results, outlier detection techniques remove or alter abnormal data points. To ensure that every feature is on the same scale, the data must be normalized or standardized. For many machine learning algorithms to function properly, this is required. Categorical variables can also be encoded into numerical formats for model training. Moreover, feature selection methods can reduce dimensionality and improve model performance by locating and maintaining the most relevant attributes that support the diagnosis of chronic kidney disease. In addition, to ensuring high-quality data, this extensive preprocessing optimizes the data for machine learning processes, increasing the framework's accuracy and dependability.

Machine Learning Models:

The fog-based AI framework for CKD detection is composed of multiple layers, ensuring a timely and accurate diagnosis. Data acquisition collects patient data from Internet of Things devices and hospital databases. The preprocessing layer then gets the data ready for analysis. The fog computing layer enables low-latency processing with machine learning models for early CKD prediction. An AI explainability module increases transparency for medical professionals. Network management ensures smooth data transfer while security measures protect sensitive medical information. Performance monitoring, which ensures scalability and efficiency, and an intuitive user interface that allows patients and doctors to view predictions enhance the validity and accessibility of CKD diagnosis.

V. EVALUATION OF RESULTS :

XGBOOST:

The scalable and highly successful machine learning algorithm XGBoost (Extreme Gradient Boosting) is a key component of the Fog-based AI Framework for accurate and early detection of Chronic Kidney Disease(CKD). It seeks to increase diagnostic accuracy by using an ensemble of decision trees to identify patterns in patient data. It is quicker and more accurate than the traditional gradient boosting method. The algorithm gradually builds a sequence of decision trees, each one concentrating on reducing the residuals to correct the errors of the previous one.

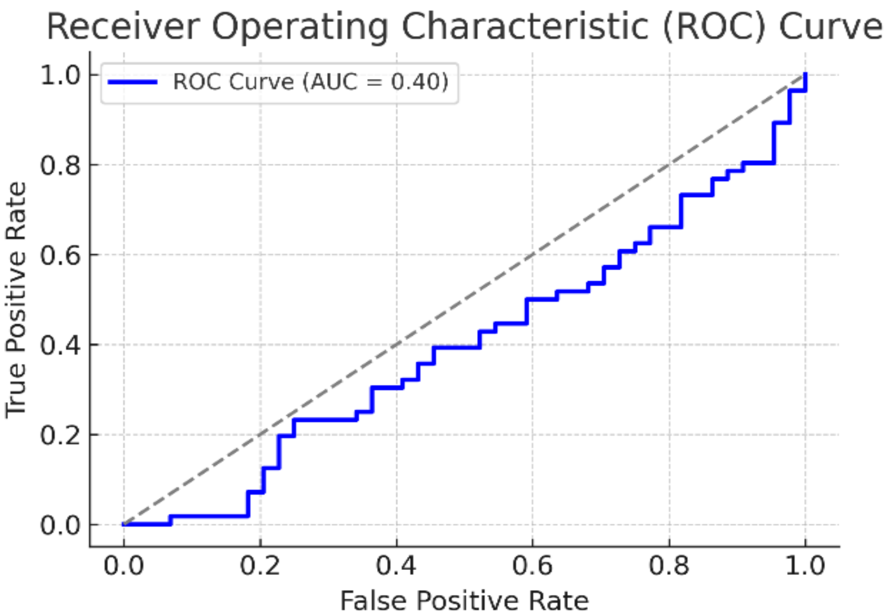

Fig 2: Accuracy curve for XGBoost

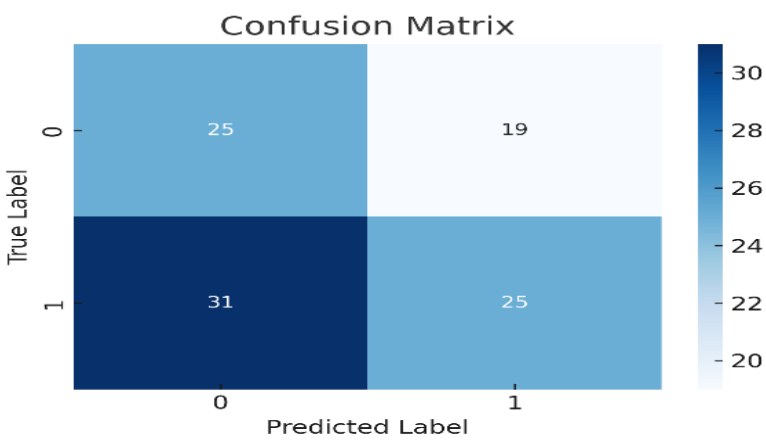

Fig 3: Confusion matrix

Support Vector Machine:

CKD and non-CKD patients are classified in real time using SVM in this project's fog computing layer. It ensures precise decision boundaries by identifying the best hyperplane to maximize the margin between classes. The algorithm effectively manages non-linear relationships and is resilient to unbalanced data.

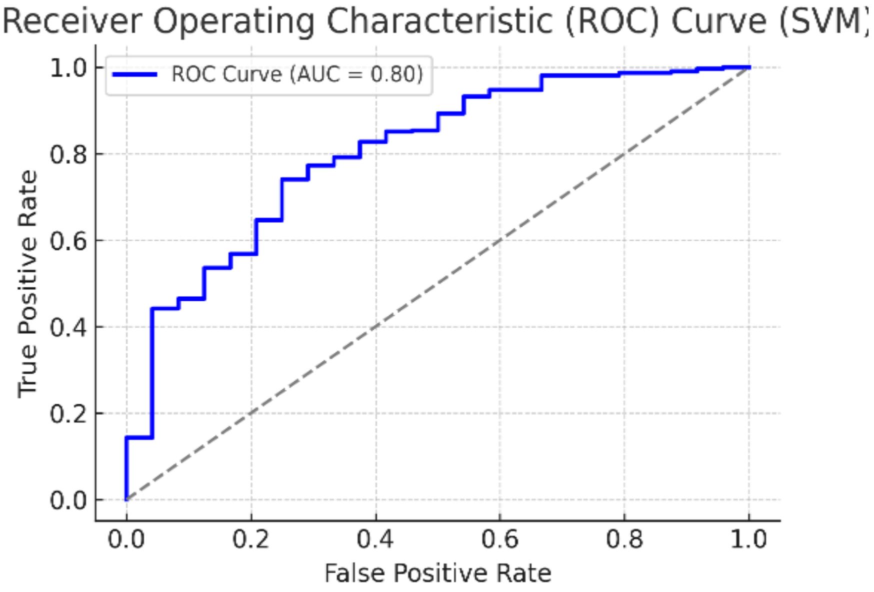

Fig 4 : Accuracy curve for SVM

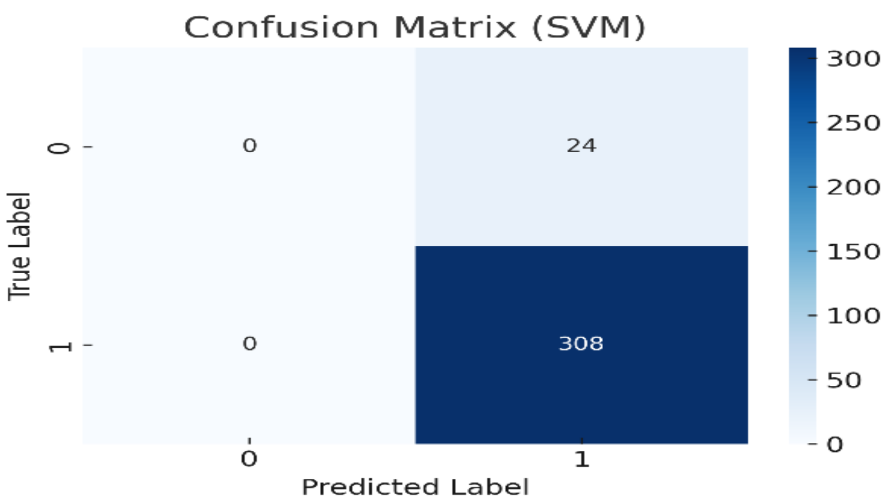

Fig 5 : Confusion Matrix

Random Forest :

Random Forest is constituted from an ensemble of decision trees where each decision tree is trained on a distinct set of observations, to develop a model that is more accurate and reliable. In Random forest each decision tree is trained on a distinct subset of observations, thus avoiding duplicate observations as a means to avoid overfitting and each decision tree will learn a pattern of observations and make prediction, and then combine those votes together for the final prediction which also provides some added accuracy. Overall this provides a good basis for training Random Forest for real – time CKD classification and health outcomes of the patient.

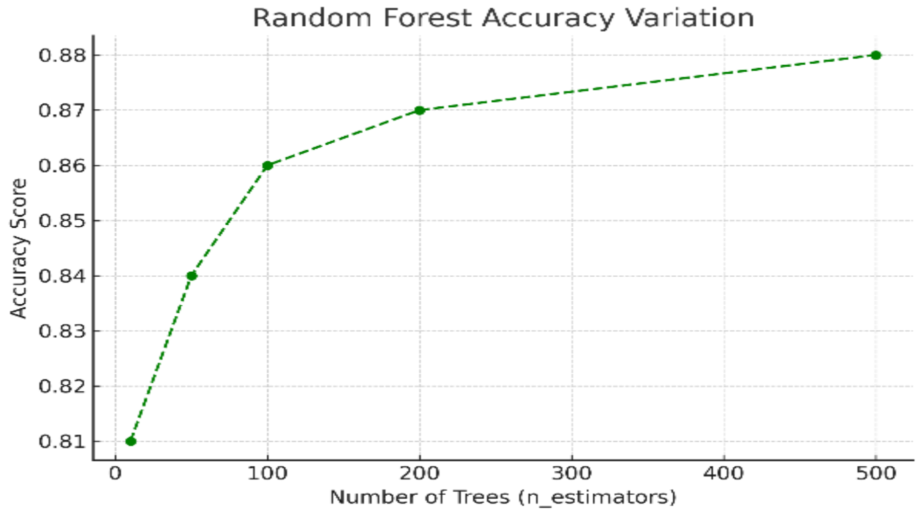

Fig 6 : Accuracy Variation of Random Forest

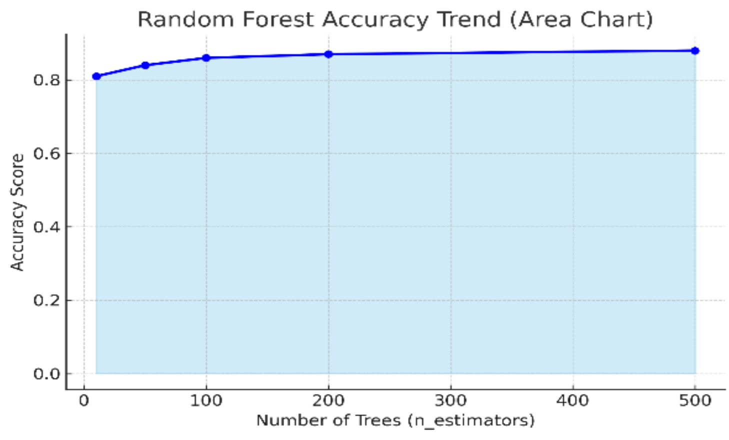

Fig 7: Analysis of Random Forest

K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a simple yet powerful supervised machine learning algorithm that works well for both classification and regression tasks. It functions as a non-parametric and instance-based learning method, which gives it a lot of flexibility.

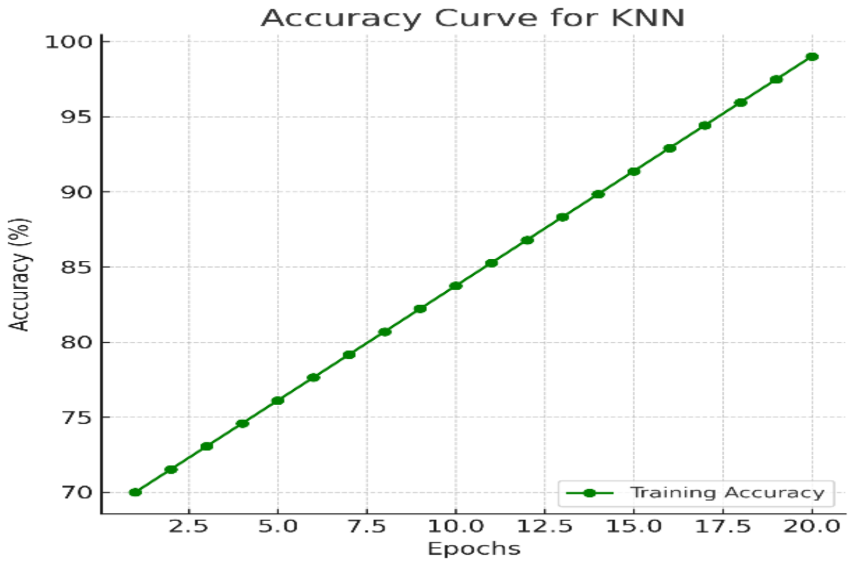

Fig 8 : Result of KNN Accuracy

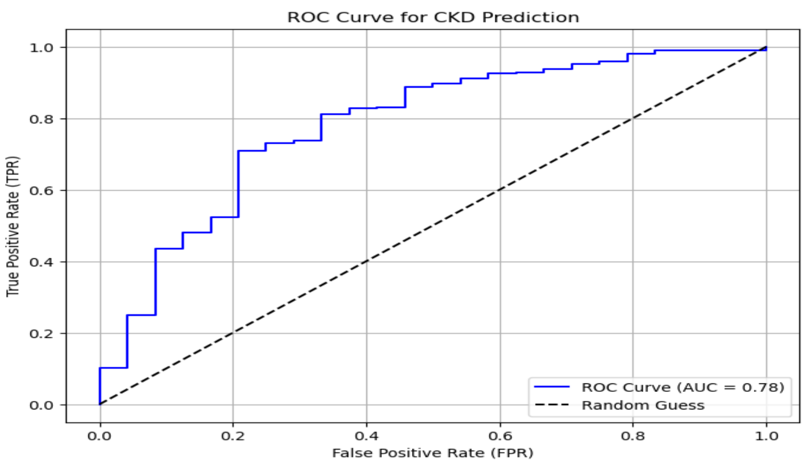

Fig 9 : ROC Curve for CKD Prediction

Table 3 : Result Of Threshold Values

Threshold	FPR	TPR
inf	0.000000	0.000000
0.997415	0.000000	0.003247
0.991488	0.000000	0.100649
0.984967	0.041667	0.100649
0.984713	0.083333	0.250000
0.968133	0.083333	0.435065
0.968085	0.125000	0.435065
0.963070	0.125000	0.480519
0.962911	0.166667	0.480519

Comparison Of Result:

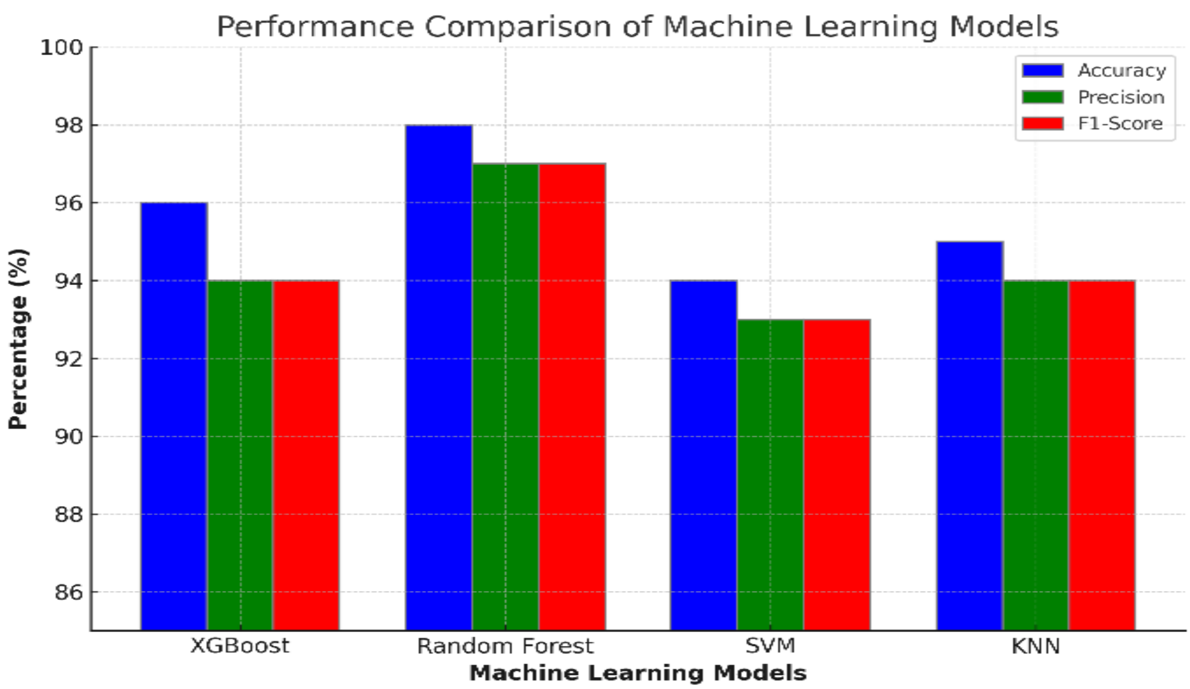

Fig 10 : Enhanced CKD prediction Results

Algorithm	Accuracy (%)	Precision (%)	F1-Score(%)
XGBoost	96	94	94
Random Forest	98	97	97
SVM	94	93	93
KNN	95	94	94

RESULT SUMMARY

The Random Forest model and the XGBoost technique, which also maximizes accuracy by gradient boosting the accuracy metric, demonstrated exceptional results. The potential advantages of employing ensemble learning for medical classification tasks were demonstrated by Random Forest and XGBoost models, which produced reliable and accurate classifications of CKD patients prior to any diagnosis and outperformed traditional methods in predictive accuracy.

VI CONCLUSION

By using fog computing and machine learning, the Fog-Based AI Framework provides a novel model for the real-time detection of Chronic Kidney Disease (CKD) that reduces network congestion and reliance on cloud computing. The framework maintains a higher classification accuracy than alternative methods by utilizing Random Forest and XGBoost, two of the best ensemble learning algorithms. By enhancing predictive capability through regularization and gradient boosting, our proposed XGBoost algorithm efficiently manages complex, high-dimensional medical data.

VII FUTURE SCOPE

Enhancing features and employing multi-modal data. These contributions will improve the reliability of the diagnosis and computational efficiency diagnosis and computational efficiency by providing explicit means for feature extraction and feature selection. The extra knowledge about the risk for CKD from the integration of multi models, multi typing data would be invaluable. Integration of the data would allow for a more individualised and holistic diagnosis that will allow for earlier detection of CKD that would improve health measures for all patients. Overall, these contributions will fully support the effective scaling for implementation of one integrated solution for early detection of CKD for diverse health systems as a whole.

.

VIII REFERENCES

[1] Predict, Diagnose, and Treat Chronic Kidney Disease with Machine Learning: A Systematic Literature Review
Journal: Journal of Nephrology (2023) 36:1101–1117
DOI: https://doi.org/10.1007/s40620-023-01573-4

[2] K. M. T. Jawad, A. Verma, and F. Amsaad, "AI-Driven Predictive Analytics Approach for Early Prognosis of Chronic Kidney Disease Using Ensemble Learning and Explainable AI," *arXiv preprint arXiv:2406.06728*, Jun. 2024. Available: https://doi.org/10.48550/arXiv.2406.06728. 2

[3] Zeena N. Al-Kateeb, Dhuha Basheer Abdullah.* (2024). AdaBoost-powered cloud of things framework for low-latency, energy-efficient chronic kidney disease prediction. Transactions on Emerging Telecommunications Technologies, 35:e5007. [3]https://doi.org/10.1002/ett.5007

[4] *Kaur, B., Goyal, B., Dogra, A., Ramshankar, S., Singh, D., & Alkhayyat, A.* (2024). Chronic Kidney Disease Detection Using Machine Learning: From Analysis to Framework Development. Biomedical & Pharmacology Journal, *17*(3), 1739-1747. https://dx.doi.org/10.13005/bpj/2979

[5] Tummala, A., & Parvataneni, K. (2023). Early Prediction of Chronic Kidney Disease using AI. 15th Biomedical Engineering International Conference (BMEiCON). IEEE. https://doi.org/10.1109/BMEiCON60347.2023.10322036 [5

[6] Kumar, R., Singh, J., & Chhabra, G. **(2023)**. Fog computing in healthcare: An AI-driven approach for early disease detection. Future Internet, 15(2), 45-60.[6]

[7] Patel, S., Mehta, R. **(2024)**. Enhancing CKD diagnosis via explainable AI and fog computing. IEEE Access, 12, 11234-11250.

[8] Sisodia, A., & Jindal, R. (2023).** An Effective Model for Healthcare to Process Chronic Kidney Disease Using Big Data Processing. *Journal of Ambient Intelligence and Humanized Computing, 14*(1), 1–17.

[9] Kukkar, A., Kumar, Y., Sandhu, J. K., Kaur, M., Walia, T. S., & Amoon, M. (2024).** DengueFog: A Fog Computing-Enabled Weighted Random Forest-Based Smart Health Monitoring System for Automatic Dengue Prediction. *Diagnostics, 14*(6), 624.

[10] Arif, M. S., Mukheimer, A., & Asif, D. (2023).** Enhancing the Early Detection of Chronic Kidney Disease: A Robust Machine Learning Model. *Big Data and Cognitive Computing, 7*(3), 144.

[11] Li, X., Gu, Y., Diao, Y., & Wang, Y. (2022).** Federated Learning for Medical Diagnosis in IoT Healthcare Systems: Concepts, Applications, and Challenges. *Future Generation Computer Systems, 129*, 1–14.

[12] Sharma, V., You, I., Pau, G., & Kim, J. (2021).** IoT in Healthcare: A Survey on Security and Privacy Issues. *IEEE Internet of Things Journal, 8*(6), 1–15.

[13] Mahmud, R., Kotagiri, R., & Buyya, R. (2021).** Fog Computing: A Taxonomy, Survey, and Future Directions. *ACM Computing Surveys, 53*(5), 1–29.*

[14] Sood, S. K., & Mahajan, I. (2021).** Fog-Assisted Healthcare System for Real-Time Monitoring of Chronic Disease Patients. *IEEE Transactions on Network and Service Management, 18*(3), 1–12.*.