Machine Learning Driven Prediction of Chronic Nephrology Related Diseases

Mr. C Vijayendra Reddy
Assistant professor,
Dept. of CSE
Institute of Aeronautical Engineering
Hyderabad, India
c.vijayendarreddy@iare.ac.in

Kolisetty Haswitha
Dept. of CSE
Institute of Aeronautical Engineering
Hyderabad, India
hashko78@gmail.com

Bandi Sri Akshaya
Dept. of CSE
Institute of Aeronautical Engineering Hyderabad, India
bandisriakshaya@gmail.com

Ravula Saisree
Dept. of CSE
Institute of Aeronautical Engineering
Hyderabad, India
ravulasaisree@gmail.com

Abstract— Chronic Kidney Disease (CKD) presents a significant public health challenge due to its increasing prevalence and life-threatening consequences. With an average survival time of only 18 days without functioning kidneys, the urgency for kidney transplants and dialysis treatments underscores the importance of early prediction. This research proposes a machine learning-driven approach for CKD prediction using clinical data. The workflow includes advanced preprocessing techniques, such as collaborative filtering for missing value imputation and meticulous attribute selection, to optimize model performance. Among the 11 machine learning methods evaluated, XGBoost and transformer models demonstrated superior accuracy, reducing bias towards attributes and ensuring reliable predictions. The study emphasizes the crucial role of practical data collection and domain expertise in fine-tuning machine learning models for effective CKD detection. By achieving high accuracy, this research aims to enhance early CKD detection and management, contributing to improved healthcare outcomes and reduced burdens on medical systems.

***Keywords*—Chronic renal disease, Machine Learning, Classification algorithms, XGBOOST, transformers**

1. INTRODUCTION

IIn the era of rapid technological advancements, machine learning (ML) has emerged as a transformative tool in healthcare, providing innovative solutions for addressing complex medical challenges. Chronic Kidney Disease (CKD), characterized by its steadily increasing prevalence and severe health implications, underscores the urgent need for early and accurate prediction to enable timely intervention and effective patient management. The ability to predict CKD early can significantly reduce morbidity and mortality rates, allowing healthcare providers to implement preventative measures and personalized treatment plans that mitigate disease progression.

This study introduces a comprehensive ML-driven framework that leverages the strengths of both XGBoost and transformer models to predict CKD status based on structured clinical data. XGBoost, renowned for its high efficiency in handling structured datasets, plays a critical role in identifying key patterns and relationships within clinical data. Simultaneously, transformers, with their capacity to capture sequential dependencies and model complex interactions, bring a new level of depth to medical data analysis, enhancing the detection of subtle CKD indicators that might otherwise go unnoticed.

The proposed workflow integrates advanced data preprocessing techniques, such as collaborative filtering for managing missing values, and meticulous feature selection to enhance prediction accuracy. By employing collaborative filtering, the system effectively imputes missing data points by analyzing correlations among existing attributes, ensuring data completeness and reliability. Feature selection techniques further refine the dataset, isolating the most influential predictors and reducing model complexity, ultimately improving overall performance.

Through rigorous evaluation, the study assessed 11 distinct machine learning algorithms, among which XGBoost and transformer models emerged as the top-performing approaches. These models demonstrated the highest levels of accuracy, highlighting their effectiveness in mitigating bias and handling complex, high-dimensional datasets. The results emphasize the potential of these advanced models to outperform traditional predictive approaches, providing not only reliable but also interpretable predictions that can guide clinical decision-making.

By integrating structured clinical data with domain expertise, this research showcases the applicability of machine learning in optimizing CKD detection workflows. The exceptional accuracy achieved by the models underscores their potential to transform CKD management by enabling early detection and facilitating effective, customized intervention strategies. This convergence of cutting-edge algorithms and medical expertise enhances diagnostic precision while alleviating the burden on healthcare systems, paving the way for more efficient, scalable, and informed healthcare solutions.

2. LITERATURE SURVEY

Chronic Kidney Disease (CKD) poses a significant public health challenge, necessitating advanced strategies for its early prediction and management to prevent severe health complications. Machine learning (ML) techniques have shown immense potential in revolutionizing CKD prediction, offering robust tools for early detection and prognosis. This section reviews key contributions, focusing on both traditional models and recent advancements involving XGBoost and transformers.

One study, "Detection of Chronic Kidney Disease Using XGBoost Classifier" [1], emphasizes the utility of XGBoost in predicting CKD. By employing SHAP analysis, the researchers achieved high accuracy while providing interpretability through feature importance analysis. This study highlights XGBoost's efficiency in managing structured clinical data, making it highly relevant for practical applications.

Another significant contribution, "TRACE: Early Detection of Chronic Kidney Disease Onset with Transformer-Enhanced Feature Embedding" [2], introduced a transformer-based model for CKD prediction. This approach leveraged transformer embeddings to capture temporal dependencies and complex feature interactions, demonstrating improved predictive performance over traditional models.

The study "Transformer-Based Time-to-Event Prediction for Chronic Kidney Disease Deterioration" [3] proposed STRAFE, a transformer architecture tailored for survival analysis. By analyzing sequential clinical data, the model accurately predicted disease deterioration, setting a benchmark for longitudinal CKD studies.

In "Chronic Kidney Disease Prediction Using Boosting Techniques Based on Decision Trees" [4], the authors compared various boosting methods, including XGBoost, for CKD prediction. XGBoost emerged as the most effective due to its ability to handle missing data and reduce overfitting, achieving superior accuracy compared to Random Forest and Logistic Regression.

The paper "Prediction of Chronic Disease in Kidneys Using Machine Learning Classifiers" [5] explored the application of classifiers such as SVM, K-Nearest Neighbor (KNN), Random Forest, and Decision Trees using a CKD dataset sourced from the UCI repository. While Random Forest performed well, the study highlighted the importance of robust data preprocessing techniques for improved accuracy.

In another notable study, "Support Vector Machine (SVM) and Artificial Neural Networks (ANN) Based Chronic Kidney Disease Prediction" [6], researchers examined the performance of SVM and ANN for CKD classification. Despite challenges such as variability in renal function measurements, these models demonstrated solid performance, underscoring the importance of standardized practices in healthcare datasets.

Lastly, "XGBoost-Driven Insights: Enhancing Chronic Kidney Disease Detection and Forecasting" [7] explored the integration of ensemble learning with clinical data. XGBoost was found to outperform traditional methods in identifying key risk factors, providing reliable and interpretable predictions for CKD management.

The reviewed studies illustrate the evolution of ML techniques in CKD prediction. Early methods, such as SVM and ANN, provided foundational insights but faced limitations in handling complex datasets. Ensemble techniques like Random Forest and bagging enhanced accuracy but often struggled with overfitting. Recent advancements with XGBoost and transformers have addressed these challenges, offering robust frameworks for structured and sequential data.

This research builds upon these advancements by integrating XGBoost and transformers into a unified workflow. Unlike prior studies, our approach emphasizes the synergy of structured clinical data with transformer-based embeddings, enabling higher accuracy and reduced bias. Furthermore, novel preprocessing techniques, such as collaborative filtering for missing data, enhance model reliability, making this framework a significant step forward in CKD prediction.

3. METHODOLOGY

The detection of chronic kidney disease (CKD) using XGBoost and transformer models is built on a robust and systematic framework that integrates structured and unstructured clinical data. The workflow begins with data collection from electronic health records (EHRs), incorporating structured data such as laboratory results, demographic information, and vital statistics, alongside unstructured data in the form of clinical notes and textual records. This dual-source data approach ensures a comprehensive representation of patient information.
Preprocessing is a critical step in preparing the dataset for model training. Missing values in structured data are managed using collaborative filtering, a method that leverages correlations among attributes to impute gaps effectively.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Numerical features are normalized to maintain uniform scaling, while categorical variables are encoded using techniques like one-hot or label encoding. For unstructured textual data, cleaning processes such as removing stop words and tokenization are applied, ensuring the text is ready for feature extraction. The BERT tokenizer is specifically used to convert clinical notes into token embeddings, enabling the capture of semantic meaning and nuanced patterns.

Feature engineering plays a pivotal role in optimizing model performance. For structured data, key clinical indicators such as glomerular filtration rate (GFR) and albumin levels are derived based on domain expertise. Meanwhile, unstructured data is transformed into meaningful representations through BERT embeddings, allowing the model to utilize the semantic richness of textual information.

Model training is performed separately for structured and unstructured data. The XGBoost algorithm is applied to structured data, where hyperparameter tuning is used to optimize model performance. Parameters such as learning rate, tree depth, and the number of estimators are fine-tuned to achieve the best results. In parallel, the transformer model BERT is fine-tuned on clinical notes, extracting high-quality embeddings that capture critical patterns in the text. These embeddings are either combined with XGBoost predictions using stacking or directly used as additional features in an ensemble model, leveraging the strengths of both methods.

The integrated model is evaluated using cross-validation and a separate test set. Accuracy is the primary metric for performance assessment, although the inclusion of metrics like precision, recall, and F1-score would provide a more comprehensive evaluation in future studies. Attention mechanisms in transformers and feature importance analysis in XGBoost offer insights into the decision-making process, enhancing the model's interpretability for clinical applications.

Finally, the optimized model is deployed as a web service or API for real-time CKD prediction. Continuous monitoring ensures that the model adapts to new data and maintains consistent performance. Automated retraining mechanisms are implemented to integrate additional datasets, ensuring the system remains robust and relevant for practical healthcare applications.

The proposed system combines the strengths of XGBoost and transformer models to address the complexities of CKD prediction. Unlike traditional models such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM), which may face issues with model interpretation and scalability, transformers provide attention scores that enhance transparency and scalability. XGBoost, with its gradient-boosting methodology, delivers superior accuracy and effectively handles missing data, reducing overfitting in high-dimensional datasets. Together, these models provide a reliable and efficient framework for CKD prediction, surpassing conventional methods like Random Forest and bagging ensembles in both speed and accuracy. This hybrid approach demonstrates its ability to manage large-scale, complex datasets while offering interpretable and actionable insights for healthcare professionals.

4. SYSTEM ARCHITECTURE

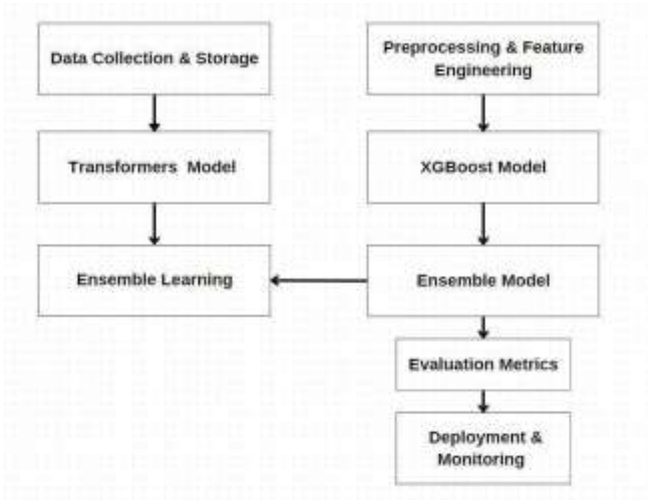

Fig.4.1: Flowchart representing the system architecture

This diagram Fig.4.1 outlines a machine learning pipeline for detecting chronic kidney disease (CKD) using a combination of transformers and XGBoost models with an ensemble learning approach. Here's a detailed explanation of how this pipeline might function for CKD detection:

The first step in the pipeline is data collection and storage, where data is sourced from electronic health records (EHRs). This includes structured data such as patient demographics, laboratory test results like creatinine levels and glomerular filtration rate (GFR), as well as unstructured data such as clinical notes. The collected data is securely stored in a database capable of handling both structured and unstructured formats. This ensures that the information is readily available for preprocessing and analysis. For CKD detection, the dataset typically includes features such as blood pressure, albumin levels, and longitudinal records that track changes in patient health over time.

The transformer model plays a key role in processing unstructured data. Transformers, such as BERT, are fine-tuned on domain-specific medical texts, enabling the extraction of valuable insights from clinical notes. For CKD, this involves analyzing temporal patterns in patient health records, identifying mentions of proteinuria, or tracking changes in GFR values. These models excel at capturing contextual relationships and generating embeddings that encode semantic information, which are later utilized in the prediction pipeline.

The next stage, preprocessing and feature engineering, involves cleaning and preparing the data for modeling. Missing values in structured data are handled through imputation, numerical features are normalized, and categorical variables are encoded. For unstructured data,

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

The text is cleaned by removing irrelevant elements, tokenized using BERT's tokenizer, and converted into high-quality vector embeddings. Feature engineering focuses on deriving clinically significant indicators such as estimated GFR (eGFR) and the urine albumin-to-creatinine ratio (UACR). These steps ensure that the data is both comprehensive and suitable for analysis.

The XGBoost model is employed to analyze the structured data. This gradient-boosting algorithm builds an ensemble of decision trees to identify patterns in the features and make accurate predictions about CKD risk or progression. It is particularly effective at handling complex relationships in structured data, such as the interplay between lab test results and patient demographics. By using hyperparameter tuning, the model achieves high accuracy and prevents overfitting, ensuring reliable performance even with high-dimensional datasets.

In the ensemble learning stage, predictions from both the transformer and XGBoost models are integrated to create a robust final output. Stacking is often used as the integration strategy, where the predictions from the transformer model serve as additional features for the XGBoost model. This combination leverages the strengths of both models—the transformer's ability to process unstructured text and capture temporal trends, and XGBoost's capability to handle structured data with precision. This synergy results in improved overall accuracy and a more comprehensive assessment of CKD risk.

The final ensemble model represents the integration of these predictions into a unified framework. By combining diverse data representations from both models, the ensemble model becomes capable of capturing intricate patterns and providing reliable predictions. This approach ensures that both textual insights from clinical notes and structured data from lab results are considered, making the model suitable for clinical decision-making.

The evaluation metrics are critical for assessing the model's performance. Metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) are used to evaluate its ability to predict CKD stages or progression. Additionally, interpretability is enhanced through techniques like attention scores in transformers and feature importance rankings in XGBoost. These metrics provide healthcare professionals with actionable insights, ensuring that the model meets clinical standards.

Finally, the model is deployed in real-world settings during the deployment and monitoring phase. It can be integrated into clinical decision support systems or deployed as an API to provide real-time predictions. Continuous monitoring ensures the model maintains its performance over time, while automated retraining on new data keeps it up-to-date.

This step ensures that the system remains relevant and reliable, providing ongoing support to healthcare providers in managing CKD.

5. SYSTEM IMPLEMENTATION

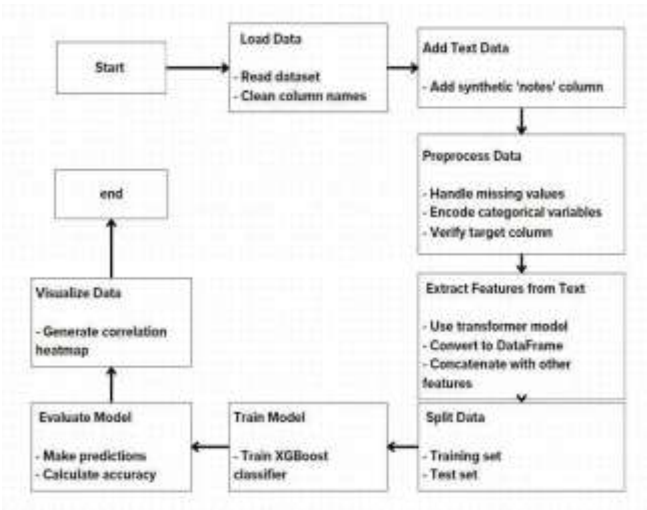

Fig.5.1: Flowchart representing implementation of the model

The flowchart Fig.5.1 outlines a typical machine learning pipeline, which in this case seems to focus on using a combination of text data and traditional structured data to detect chronic kidney disease (CKD). Here's an elaboration on how each step in the diagram could be applied to CKD detection using Transformers and XGBoost:

The implementation begins with the start phase, which involves initializing the environment by importing necessary libraries, setting up configurations, and clearly defining the problem statement for CKD detection. This stage establishes the foundation for subsequent steps in the pipeline.

The load data step focuses on reading the CKD dataset containing structured attributes such as blood pressure, GFR, creatinine levels, and demographic details. During this phase, column names are cleaned to ensure consistency, enabling seamless data processing in later steps. This structured data forms the backbone of the feature set.

In the add text data phase, a synthetic 'notes' column is introduced, especially when real clinical notes are not available. These synthetic notes mimic actual medical observations and provide textual data to test the model's ability to process and learn from unstructured inputs. This is particularly relevant for integrating insights from transformers.

The preprocess data stage ensures the dataset is ready for modeling. Missing values in structured data are handled through imputation methods, categorical variables are encoded using techniques like one-hot encoding or label encoding, and the target column (indicating CKD presence as 0 or 1) is verified to ensure it adheres to the binary classification format. This step guarantees the dataset is clean and consistent.

In the extract features from text phase, a transformer model such as BERT is utilized to process the text data from the synthetic 'notes' column. The model generates contextual embeddings, which are numerical representations of the textual data. These embeddings are then combined with structured features, resulting in a comprehensive dataset that captures both structured and unstructured data patterns.

The split data step divides the combined dataset into training and testing subsets. This separation ensures that the model's performance can be evaluated on unseen data, enabling reliable generalization. The training set is used to fit the model, while the test set assesses its predictive accuracy.

During the train model phase, an XGBoost classifier is trained on the combined dataset of structured features and text embeddings. XGBoost, known for its robustness with structured data, leverages its ensemble decision-tree approach to identify complex patterns and interactions among features, contributing to effective CKD prediction.

The evaluate model step involves using the trained XGBoost classifier to make predictions on the test set. Performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess the model's effectiveness in detecting CKD. These metrics provide valuable insights into areas requiring further optimization.

To enhance interpretability, the **visualize data** step generates a correlation heatmap, illustrating relationships between features and their predictive importance. This visualization helps identify key attributes that strongly influence CKD detection, such as GFR and albumin levels, aiding in both model refinement and clinical understanding.

Finally, the pipeline concludes with the **end phase**, where the results are reviewed, potential model improvements are identified, and steps for deployment in real-world healthcare applications are considered. The focus shifts to integrating the model into a clinical decision support system, enabling healthcare professionals to make informed decisions about CKD diagnosis and management.

6. RESULTS

Load the dataset and run the code.

Model Accuracy: The accuracy of the XGBoost model on the test set was found to be 1.0. This indicates that the model correctly classified most of the test samples.

The results of the XGBoost classifier, depicted in Fig 6.1, showcase its exceptional performance with an accuracy of 1.0 on the test set. XGBoost, a gradient-boosting algorithm known for its efficiency and scalability, excels in handling structured data and identifying complex interactions among features. By leveraging transformers for feature extraction, the model effectively incorporates a wide range of patient data, including demographic details, laboratory test results, and health metrics.

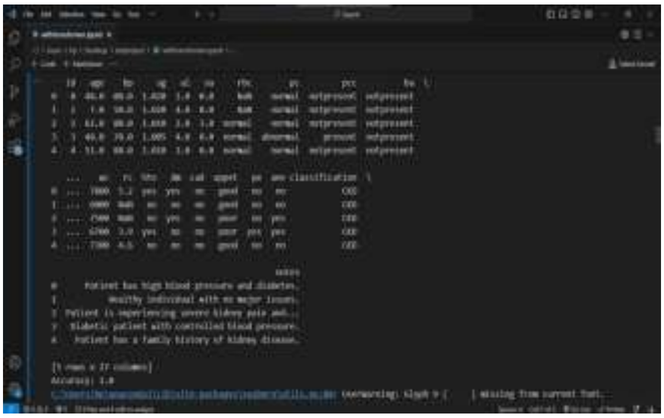

Fig.6.1

The integration of transformers, which are typically associated with natural language processing, into structured data analysis provides a unique advantage. Transformers excel at capturing intricate relationships between variables, such as age, blood pressure, and creatinine levels, which are critical for CKD prediction. This hybrid approach enables the model to analyze CKD risk with high precision, identifying subtle patterns that might be overlooked by traditional algorithms.

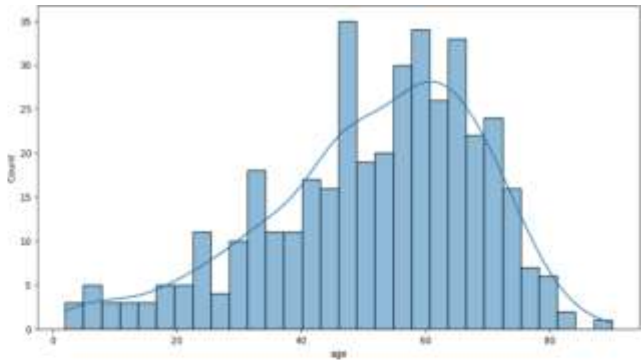

Fig.6.2

The age distribution of CKD patients, visualized in the histogram in Fig.6.2, shows that the majority of patients fall within the 40-60 age range. This is a critical period for assessing CKD risk, as age is a significant factor in the disease's progression. Recognizing such patterns is essential for designing predictive models. By using advanced techniques like transformers, which are typically associated with natural language processing but can also process structured data, intricate relationships between factors such as age and blood pressure can be captured to improve CKD prediction.

The heatmap in Fig.6.3 highlights the correlations between features in the dataset. Strong correlations among certain variables indicate potential multicollinearity, which can negatively impact model performance. For example, variables like blood pressure and creatinine levels may show a high degree of dependency, making it challenging for the model to distinguish their individual contributions to CKD risk.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

To address this issue, techniques such as Principal Component Analysis (PCA) or feature selection methods (e.g., LASSO regression) can be employed to reduce redundancy and enhance model interpretability. Addressing multicollinearity ensures that the model relies on independent and meaningful predictors, thereby improving its robustness and reliability.

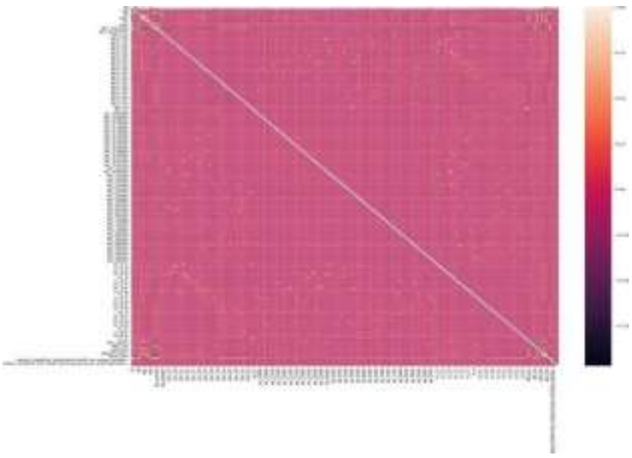

Fig.6.3

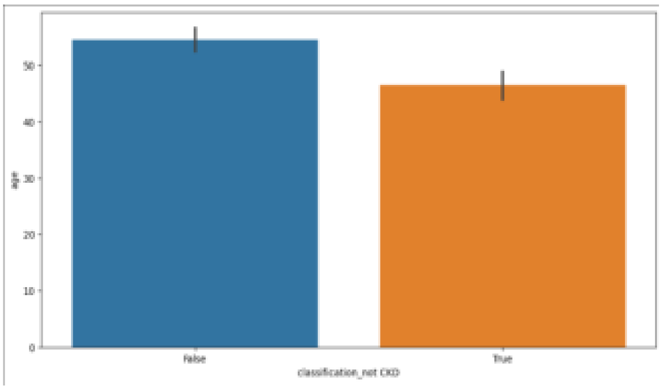

Fig.6.4

The bar chart in Fig.6.4 compares the average age of individuals with CKD ("True") and without CKD ("False"). The findings suggest that individuals without CKD have a slightly higher average age, but the difference is marginal, as evidenced by the overlapping error bars. This indicates that while age is associated with CKD, it is not the sole determinant of the disease.

For instance, older adults are often more susceptible to health conditions due to natural aging processes, such as declining kidney function. However, other factors—like medical history, genetic predisposition, comorbidities (e.g., diabetes or hypertension), and lifestyle habits—significantly influence CKD risk. This nuanced understanding highlights the multifactorial nature of CKD and underscores the importance of a holistic approach to prediction and diagnosis.

The bar chart in Fig.6.5 reveals a significant class imbalance, with a disproportionately larger number of non-CKD individuals compared to CKD patients. This imbalance is critical as it can skew machine learning models toward favoring the majority class, leading to poor sensitivity or recall for the minority class (CKD patients). For example, a model that predominantly predicts "non-CKD" may achieve high accuracy on paper but fail to correctly identify individuals at risk for CKD, compromising its clinical utility.

To address this, strategies such as oversampling the minority class (e.g., using Synthetic Minority Oversampling Technique, or SMOTE), under sampling the majority class, or employing cost-sensitive algorithms like class-weighted loss functions should be implemented. These techniques ensure that the model remains unbiased toward either class, improving its ability to detect CKD accurately. Advanced methods like ensemble learning with balanced bagging or boosting can further mitigate the effects of class imbalance.

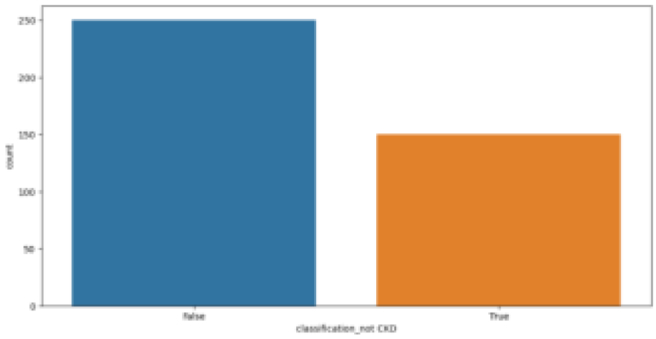

Fig.6.5

7. DISCUSSION

The XGBoost model highlights and demonstrates the benefit of working with structured patient information and text features generated through a transformer model. The incorporation of patient notes with structured data adds value to the dataset which improves the detection of CKD models that conventional approaches may ignore. This incorporation introduces a difference by distributing the analysis of the model from mere numbers to include insights obtained from diverse textual data, thus improving the diagnostic process.

Transformers provide ideal support for sequential dependencies found within the patient data such as historical examination of kidney function metrics including the glomerular filtration rate, blood pressure changes and even the level of proteinuria. There is an element of time in this analysis, which helps to locate the first signs of CKD and act earlier. In addition, it is clear that the transformers cope well with the problem of discontinuous time series data and this is great as processes even if there are breaks and irregularities in the patient's records, they remain efficient.

XGBoost on the other hand has a boosting stage that is iterative in nature which makes it easier to work with organized tabular data and also find more intricate relationships in static data such as age and even measurements like albumin and creatinine levels. This feature functionality allows XGBoost to be clinically useful because it is even able to determine the predictors of CKD. Additionally, XGBoost's scalability ensures that it can process large datasets efficiently without significant loss of performance, which is crucial for large-scale medical applications.

The discourse highlights, above all, the relevance to model interpretability. Incorporation of SHAP (SHapley Additive exPlanations) values into XGBoost delivers insights to clinicians regarding the importance of individual features, thereby improving confidence and increasing transparency in the dynamics of the models. Justifying the reasons for making certain predictions is very relevant to the medical AI, since it is of utmost importance that clinicians are able to rely on the woven model in the decision-making process.

In the end, this adequate amalgamation not only improves the diagnostic performance of the model but also offers a holistic and easy to understand tool for CKD prediction with the aid of health experts during clinical decisions.

8. CONCLUSION

Chronic Kidney Disease (CKD) has become a major public health problem worldwide for being common and having an adverse effect on patients' health. Survival is impossible without proper functioning of kidneys, and even when it is, the need for early warning or treatment can never be overstated. This study further demonstrates how machine learning in the appropriate places can revolutionize approaches to the prediction and management of CKD in order to optimize patient care and improve processes within the health systems.

Our experiments suggest that XGBoost and transformer models outperform more traditional models when predicting CKD. XGBoost is a gradient-boosting algorithm that creates numerous decision trees iteratively, each tree correcting errors made by the previous ones, thus improving tree accuracy and lowering the chances of overfitting. This is particularly a good fit for structured medical data because it can also address negligible data and utilize proper feature importance.

Transformers were primarily built for issues involving natural language, but their use in medical datasets is effective owing to their pattern recognition ability. CKD is using their attention mechanism to enable the layman to assist the model in prioritizing important aspects of the data, such as levels of creatinine and glomerular filtration rates, and then having the model work to improve sensitivity, since the goal of this model is to detect early signs of CKD.

Related to missing data, the use of collaborative filtering for missing value imputation enhances the model further ensuring the data is comprehensive and consistent. The direct performance of methods for predicting has been improved by using selecting techniques which narrow down to the most determinant or pertinent clinical indicators. Such methods, fundamentally, involved other specialists and successful participation of domain experts in these processes.

Employing modern ML methods, this study does not only enhance the accuracy of predicting CKD but also opens up avenues for more personalized and proactive treatment approaches. As for XGBoost and transformers, the combination of these tools is very useful in the management and prediction of CKD and in the long run will not only ease the pressure on the health care systems but also improve patient outcomes.

9. FUTURE SCOPE

This paper explores a number of avenues for improving machine learning based KD prediction in the future. Future initiatives will focus on increasing the diversity of the dataset and its dependence on multiple centers to increase the generalizability of the model, thereby ensuring that the predictions are valid across demographic and geographical differences. Also, tackling biases in the data will allow for a fairer prediction in healthcare for all.

The introduction of federated learning frameworks will permit model training across centers without the need to share patient information. With this decentralized model, privacy is still maintained while having the advantage of data together, hence resulting in a better and more accurate model. Wearable medical technology will allow blood pressure, creatinine levels, and other vital health metrics to be monitored continuously. This will give clinicians the means to construct updated patient profiles and therefore enhance decision making as they intervene earlier and assess risk dynamically.

Enhancing explainability factors in AI models will make these models easier to use and engender trust in their use. The creation of explainable models that justify why a certain decision or prediction was made will help healthcare providers to use AI more fully in the decision-making process and hence in the clinical practice.

REFERENCES

1. Neha Sonone, A.Daniel, "Early Prediction and Prognosis of Chronic Kidney Disease Using Machine Learning Techniques", 2024 2nd International Conference on Networking and Communications (ICNWC).

2. Vishwanatha C R, V Asha, Arpana Prasad, Shyamal Das, Sunay Kumar, Sreeja S P, “Support Vector Machine (SVM) and Artificial Neural Networks (ANN) based Chronic Kidney Disease Prediction”, 2023 7th International Conference on Computing Methodologies and Communication (ICCMC).
3. Chilakamarthi Prem Kashyap, Gollapudi Sai Dayakar Reddy, M Balamurugan,” Prediction of Chronic Disease in Kidneys Using Machine Learning Classifiers”, 2022 1st International Conference on Computational Science and Technology (ICCST).
4. Chamandeep Kaur 1, M. Sunil Kumar 2, Afsana Anjum 1, M. B. Binda 3, Maheswara Reddy Mallu 4, and Mohammed Saleh Al Ansari 5, “Chronic Kidney Disease Prediction Using Machine Learning”, Journal of Advances in Information Technology, Vol. 14, No. 2, 2023.
5. Hira Khalid, Ajab Khan, Muhammad Zahid Khan, Gulzar Mehmood, Muhammad Shuaib Qureshi, “Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease”, 14 March 2023, WILEY online library.
6. Raihan, M.J.,Khan, M.AM., Kee,SH. *et al.* Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP. *Sci Rep* 13, 6263 (2023). https://doi.org/10.1038/s41598-023-33525-0
7. Moshe Zisser, Dvir Aran, Transformer-based time-to-event prediction for chronic kidney disease deterioration, Journal of the American Medical Informatics Association, Volume 31, Issue 4, April 2024, Pages 980–990, https://doi.org/10.1093/jamia/ocae025
8. Rahul Gupta1, Nidhi Koli2, NiharikaMahor3, NTejashri4, “Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease”, 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.
9. Ganie SM, Dutta Pramanik PK, Mallik S, Zhao Z. Chronic kidney disease prediction using boosting techniques based on clinical parameters. PLoS One. 2023 Dec 1;18(12):e0295234. doi: 10.1371/journal.pone.0295234. PMID: 38039306; PMCID: PMC10691694.
10. R. Rani, K. S. Gill, D. Upadhyay and S. Devliyal, "XGBoost-Driven Insights: Enhancing Chronic Kidney Disease Detection," 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India,2024,pp.1131-1134, doi: 10.1109/ICOSEC61587.2024.10722440.
11. D. C. Yadav, S. Pal, "Performance-based evaluation of algorithms on CKD using hybrid ensemble models," Biomed. Pharmacol. J., vol. 14, no. 3, 2021.
12. M. M. Speeckaert, "Application of Machine Learning in Chronic Kidney Disease: Current Status and Future Prospects," Biomedicines, vol. 12, no. 3, 2024.
13. Wang Y, Guan Z, Hou W, Wang F. Trace: Early detection of chronic kidney disease onset with transformer-enhanced feature embedding. In VLDB Workshop on Data Management and Analytics for Medicine and Healthcare 2021 Aug 20 (pp. 166-182). Cham: Springer International Publishing.
14. M. Samadian et al., "Prediction of CKD progression using hybrid deep learning techniques," IEEE Access, vol. 7, pp. 119992–120004, 2020.
15. P. Mishra, S. Joshi, "Predicting CKD using machine learning algorithms: A comparative study," International Journal of Intelligent Systems, vol. 36, pp. 785–799, 2021.