International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

A Comparative Analysis of Supervised Classification Algorithms and Missing Data Handling for Enhancing Chronic Kidney Disease Prediction

A. Swathi[*1], Dr Golda Dilip[2], Dr A.Vani Vathsala[3]
[1]Research Scholar, Department of Computer Science and Engineering
SRM Institute of Science & Technology, Vadapalani, Chennai, Tamil Nadu, India-600026
as5087@srmist.edu.in, ORCID: https://orcid.org/0009-0005-8079-8941
[2]Associate Professor, Department of Computer Science and Engineering
SRM Institute of Science & Technology ,Vadapalani, Chennai, Tamil Nadu, India-600026
goldadilip@gmail.com, ORCID: https://orcid.org/0000-0001-5175-6957
[3]Professor and Head, Department of Computer Science & Engineering, CVR College of Engineering, Hyderabad, India-501510
vani_vathsala@cvr.ac.in, ORCID: https://orcid.org/0000-0003-3229-9708

Abstract- Chronic kidney disease (CKD), which is becoming a more significant public health concern, is characterized by a gradual but concerning increase in morbidity and death, particularly in its early, asymptomatic stages. Risk factors for chronic kidney disease (CKD), including genetic predisposition, obesity, diabetes, and hypertension, affect the illness's prevalence. When there are no outward signs of an illness, it is challenging to diagnose and treat it in its early stages. To tackle this pressing issue, our research does a comprehensive investigation through a comparative comparison of supervised classification techniques. In particular, we examine the prediction performance of CKD using the Random Forest, Decision Tree, and Support Vector Machine (SVM) techniques. We also look into a number of approaches to handling missing data. Our research presents a thorough evaluation of these algorithms' performance under different data cleaning methods, pointing out both their benefits and drawbacks. Ultimately, our research aims to clarify the early detection and treatment of chronic kidney disease (CKD) and pave the way for larger-scale public health initiatives to tackle this quickly escalating health emergency.

Keywords- Chronic Kidney Disease (CKD), Supervised Classification Algorithms, Missing, Data Handling, Early Diagnosis, Public Health Initiatives.

I.INTRODUCTION

Chronic Kidney Disease (CKD) has emerged as a global public health challenge with a striking increase in morbidity and mortality rates in recent years [1]. CKD is a multi factorial condition, and its progression is often insidious, characterized by a lack of symptoms in its early stages, which poses a significant obstacle to timely diagnosis and intervention. Consequently, CKD frequently remains undetected until it reaches advanced stages, resulting in adverse health outcomes and increasing healthcare burdens [2].

The risk factors associated with CKD are diverse and well-documented. Diabetes, for instance, has been identified as a prominent contributor to CKD. Diabetic nephropathy is a common cause of CKD, and its prevalence is on the rise globally. Hypertension is another recognized risk factor, as elevated blood pressure levels can exert damaging effects on the kidneys over time. A familial history of kidney disease has also been linked to an increased risk of CKD, indicating a genetic predisposition. Additionally, obesity, with its growing prevalence worldwide, has emerged as a risk factor for CKD development, further exacerbating the CKD burden [3][4].

To address the challenge of early CKD detection, machine learning and data analysis techniques have gained prominence in recent years. Supervised classification algorithms, including Support Vector Machine (SVM), Random Forest, and Decision Tree, have demonstrated potential in predicting CKD based on patient data. These algorithms offer a data-driven approach to risk assessment, enabling the identification of individuals at heightened risk of CKD, even in the absence of apparent symptoms [5][6].

However, the accuracy and reliability of these predictive models are contingent on the quality of the input data. Missing data is a common issue in clinical datasets and can introduce bias and uncertainty into CKD risk predictions. The handling of missing data is a crucial aspect of developing robust and accurate predictive models for CKD. Various data cleaning methods exist, each with its advantages and disadvantages, and their impact on predictive model performance warrants careful examination [7].

The machine learning paradigm known as "supervised learning" is designed for applications where algorithms are used to map input data to predetermined goal or output values. These tasks include regression (which predicts numerical results) and classification (which labels data into groups). To enable the system to learn and predict unknown data, it depends on labeled datasets, where each data point is matched with a target label. The algorithm minimizes differences between genuine labels and predictions by honing its internal

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

parameters through training, which makes it more capable of generalizing to new, similar data. Often used in real-world applications such as spam email classification, image recognition, sentiment analysis, and property attribute-based real estate price prediction, supervised learning algorithms include Random Forest, Decision Trees, Support Vector Machines, Linear Regression, and Neural Networks.

In this paper we present an effective predictive classification model that can accurately classify water quality based on physicochemical parameters. The primary aim is to accurately predict and classify the quality of water samples from different locations using machine learning techniques. This model has the potential to contribute to better water management practices and safeguarding the quality of water resources. Groundwater, a vital natural resource, plays an indispensable role in sustaining life on our planet. It constitutes a significant portion of the Earth's freshwater supply, serving as a primary source of drinking water for billions of people worldwide and supporting various ecosystems. As society's demands for clean and reliable water continue to escalate due to population growth, urbanization, and industrialization, the quality of groundwater becomes a subject of paramount importance. Understanding and safeguarding groundwater quality is crucial not only for human health but also for the ecological integrity of aquatic environments.

II.MOTIVATION

The two primary risk factors for chronic kidney disease (CKD), a serious health issue, are diabetes and a family history of the condition. Research indicates that approximately thirty percent of individuals with diabetes have chronic kidney disease (CKD). Additionally, studies have shown that proactive therapy and early identification of chronic kidney disease (CKD) enhance patients' quality of life. The application of machine learning prediction algorithms is one possible way to anticipate the onset of CKD and enable timely therapy. The multiple applications of machine learning algorithms for the prognosis of chronic renal illness are highlighted by a thorough assessment of the literature.

In the context of chronic kidney disease (CKD), "chronic" refers to an extended period of progressive kidney degeneration. Damage to the kidney's blood vessels, often resulting from diabetes and hypertension, is the primary cause of this ailment. It is believed that 10% of people worldwide have chronic kidney disease or CKD for short. According to estimates from the World Health Organization (WHO), chronic illnesses caused 35 million deaths worldwide in 2019. Over 58 million people have died worldwide as a result of these tragedies. It is estimated that between the ages of 65 and 74, one in five men and one in four women worldwide may develop chronic kidney disease (CKD). The diagnosis of chronic kidney disease (CKD) frequently involves a biopsy, laboratory testing, clinical data, imaging scans, and, in certain cases, laboratory testing. However, the purpose of this work is to propose machine learning techniques as a readily available, affordable, and user-friendly means of identifying chronic kidney disease (CKD) and offering substitute tests [8].

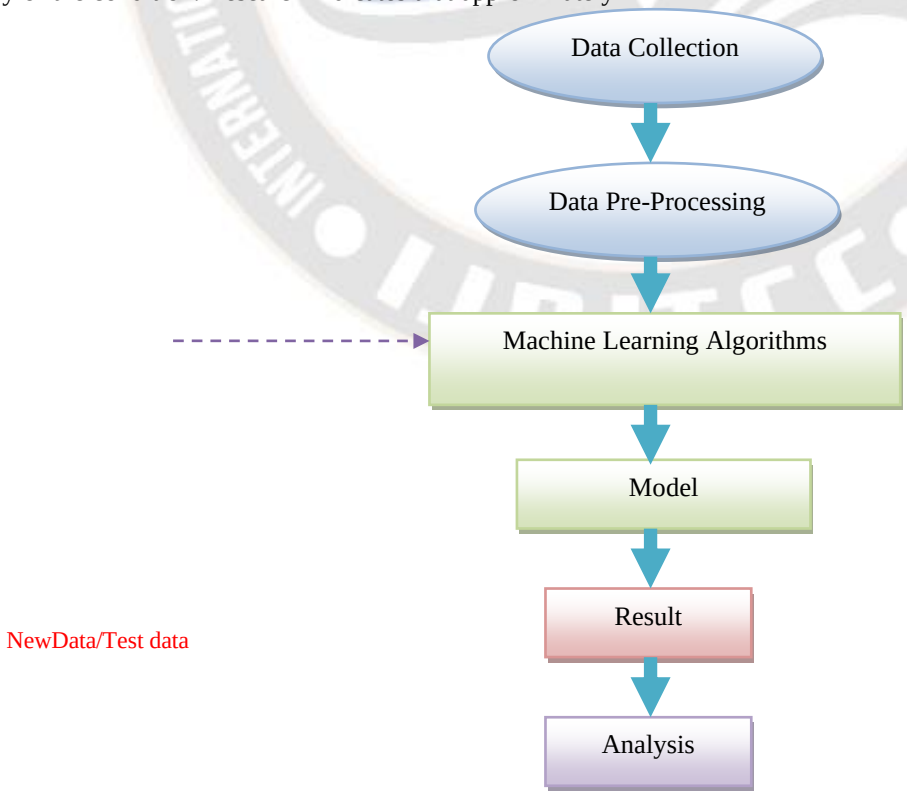

Figure 1: Stages of the Knowledge Data Discovery process

In order to tackle this urgent problem, this research study compares and contrasts several supervised classification algorithms, such as SVM, Random Forest, and Decision Tree, and uses several data cleaning techniques to deal with missing values. The goal of the study is to evaluate how well these algorithms estimate the risk of CKD under different data quality scenarios, highlighting the advantages and disadvantages of each. In the end, the research's findings aim to offer insightful information on the early detection and management of CKD, which could help shape more potent public health initiatives to counter the rapidly worsening CKD issue.

III. LITERATURE REVIEW

A thorough summary of earlier studies on renal disease prediction using machine learning classification algorithms is given in the literature review. Numerous investigations have looked into various methods and algorithms to improve the accuracy of chronic kidney disease (CKD) predictions. Liu et al. (2015) used the Random Forest algorithm to address the problem of missing data in a national health survey. In healthcare datasets, missing data is a frequent problem that can skew the results. The authors suggested a unique method for predicting and impute missing values that makes use of Random Forest, a robust and adaptable machine learning technology. Their study attempted to improve the accuracy of missing values in order to improve the quality of health data. The importance of sophisticated data imputation techniques in enhancing the dependability and usefulness of healthcare datasets is highlighted in this study [9].

Manogaran et al. conducted a comprehensive survey on big data architectures and machine learning algorithms in the healthcare sector in 2017. With the advent of big data, data analysis and management in healthcare have undergone a revolutionary upheaval. The authors provided incisive analysis on how technological advancements are changing the healthcare industry through an examination of various big data architectures and machine learning algorithms employed in healthcare settings. This study highlights the possibilities for improving patient care, diagnosis, and treatment. It's a helpful resource for understanding how big data and machine learning are being integrated in the healthcare sector.

A ground-breaking work in the area of genetics and nephrology was carried out by Sampson et al. (2015). They employed integrative genomics to find new connections in their study of Black people' APOL1 risk genotypes. The hereditary components of kidney disease risk are clarified by this study, with a focus on Black communities, who are disproportionately impacted by chronic kidney disease. The results could lead to more specialized and individualized methods to kidney disease diagnosis and therapy. They also have important implications for our understanding of the genetic foundations of renal illnesses [11].

Wickramasinghe et al. (2017) presented a plan for nutrition-based care of Chronic Kidney Disease (CKD). Multiclass neural networks, multiclass decision forests, multiclass jungles, and multiclass logistic regression were among the classification algorithms used. Based on the predicted potassium zone—which was determined by measuring the potassium levels in the patients' blood—these classifiers determined a tolerable potassium zone and suggested suitable meals [12]. Wibawa et al. (2018) evaluated and proposed the Kernel-based Extreme Learning Machine (ELM) for chronic renal ailment prediction. Four kernel-based ELM models (RBF-ELM, Linear-ELM, Polynomial-ELM, and Wavelet-ELM) were compared to the conventional ELM in terms of performance. RBF-ELM demonstrated better prediction accuracy, which may provide new information on the prognosis of CKD [13].

Dulhare and Ayesha (2016) used a Naïve Bayes classifier with One R attribute selector to predict the course of CKD in addition to creating action rules based on phases. Their goal was to stop the progression of chronic renal disease into more severe stages. Zhang et al. (2018) looked at the prediction of patients' survival rates with chronic kidney disease (CKD) using artificial neural network (ANN) models. It was observed from [14] to [15] that early identification and appropriate treatment are critical for enhancing the prognosis of CKD patients.

Aljaaf et al. (2018) emphasize the importance of state-of-the-art techniques for illness prognosis and assert that the application of predictive analytics and machine learning algorithms holds potential for early CKD prediction. Arif-Ul-Islam and Ripon (2019) investigated the efficacy of CKD detection using boosting algorithms like AdaBoost and LogitBoost. They created rules that define attribute associations within CKD using Decision Trees and Ant-Miner machine learning. Kaur and Sharma (2017) employed a range of data mining approaches in a Hadoop environment to forecast chronic kidney disease. Utilizing classifiers such as Support Vector Machine (SVM) and K-Nearest Neighbour (KNN), their research highlights the critical function big data plays in improving the prognosis and diagnosis of chronic kidney disease [16] [17] 18].

Problem Statement

Chronic Kidney Disease (CKD) presents a significant and growing public health challenge due to its insidious onset, often remaining asymptomatic in its early stages. The disease's prevalence is increasing, contributing to a substantial rise in morbidity and mortality. Several risk factors; including obesity, diabetes, hypertension, and genetic predisposition, makes early identification and treatment of CKD crucial.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

Contributions

This research addresses the urgent need to enhance CKD prediction accuracy by conducting a comprehensive investigation into the efficacy of supervised classification algorithms. Specifically, it focuses on the evaluation of Support Vector Machine (SVM), Random Forest, and Decision Tree algorithms. Furthermore, the study explores various data cleaning techniques to address the issue of missing data, shedding light on their respective advantages and disadvantages.

The primary contributions of this research lie in:

a) Improved CKD Prediction: By rigorously assessing the selected classification algorithms, this study aims to advance the state of CKD prediction, potentially leading to earlier diagnoses and more effective treatments.

b) Data Cleaning Insights: The research provides valuable insights into handling missing data, a common challenge in clinical datasets, thereby improving the robustness of predictive models.

c) Enhancing Public Health Initiatives: The findings of this study hold the potential to inform more resilient public health initiatives aimed at addressing the escalating CKD crisis by enabling earlier detection and intervention.

IV. METHODOLOGY

The primary aim of this study is to predict Chronic Kidney Disease (CKD) using machine learning algorithms. CKD is characterized by gradual kidney damage, leading to impaired blood filtration. It is often asymptomatic in its early stages, making early detection critical. CKD is associated with risk factors such as obesity, diabetes, hypertension, and genetic susceptibility. This research focuses on improving the accuracy of CKD prediction to enable timely medical intervention.

The primary objective of this research is to predict Chronic Kidney Disease (CKD) using machine learning algorithms. CKD is characterized by kidney damage that impairs the blood filtering process over an extended period, leading to the buildup of waste products and associated health issues. It is a significant global health concern, affecting 10% of the world's population, with millions succumbing to the disease due to late diagnosis.

Proposed System

The research introduces an automated system for CKD prediction. It's designed as a real-world web-based application applicable in various healthcare settings, particularly hospitals. The system operates under the assumption that its variables are independent. It employs several machine learning algorithms, including Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbor (KNN), and a Voting Classifier. These algorithms are employed to construct predictive models.

Algorithm Implementation

The algorithmic approach consists of the following steps:

1. Data Input: - Open the Chronic Kidney Disease File.

2. Pre-processing of data:

- *Use one-hot encoding, for example, to convert category values to numerical ones.*
- *Use the mean of the corresponding columns to fill in any numerical missing values.*
- *Replace any missing values in categories with the corresponding column's mode.*

3. Putting Classifier Models Together:

- *Create empty initialization models for a voting classifier, SVM, Random Forest, Decision Tree, and KNN.*

For every model in Random Forest, SVM, KNN, Voting Classifier, and Decision Tree:

- *Created training and testing sets from the dataset.*
- *Use the training data to train the model.*
- *Use testing data to forecast CKD status.*

4. Evaluating Model Precision:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

With every model:

- *Determine the confusion matrix, taking into account false positives, false negatives, true positives, and true negatives.*
- *Determine the F1-score, accuracy, precision, and recall by utilizing the confusion matrix.*

5. Choosing the Ideal Model:

- *Evaluate each model's performance metrics, such as accuracy.*
- *Determine which model is the best at predicting chronic kidney disease (CKD) based on accuracy.*

6. Product:

- *Show the top model together with its F1-score, recall, accuracy, and precision.*
- *Preserve the optimal model for CKD forecasts in the future, if desired.*

7. End

Significance

a) **Decision Trees:** Decision Trees are effective for classifying and predicting outcomes based on intuitive decision rules. Their interpretability allows for insight into the factors influencing CKD prediction, making them valuable in understanding the disease's complexities.

b) **Random Forests:** Random Forests, an ensemble approach, is employed to improve model accuracy by reducing correlations between decision trees. This is particularly useful for complex datasets with many features.

c) **Support Vector Machines (SVM):** SVMs are introduced to classify both linear and non-linear data. They aim to find an optimal hyper plane for data separation and are suitable for high-dimensional datasets.

d) **K-Nearest Neighbor (KNN):** KNN provides a simple yet versatile approach for classification problems. It's employed in scenarios such as credit ratings and loan disbursement to predict safe or risky outcomes.

e) **Logistic Regression:** Logistic Regression, based on the logistic function, is used to map values between 0 and 1. It finds applications in various fields, including finance and healthcare.

f) **Voting Classifier:** The Voting Classifier combines the strengths of multiple models to predict output based on the majority vote of individual classifiers, enhancing overall accuracy.

In essence, this research aims to automate the prediction of CKD by harnessing the capabilities of diverse machine learning algorithms and data pre-processing methodologies. By doing so, it holds the promise of not only improving early diagnosis but also advancing the treatment of CKD. This endeavor addresses the pressing requirement for more effective healthcare management in the midst of the mounting CKD crisis and the health risks it entails.

V. IMPLEMENTATION & RESULTS

For our study, data was sourced from 400 clinical records available through the UCI Machine Learning Repository. Following thorough data cleansing and missing value removal, we retained 220 complete records for analysis. This dataset comprises 25 numerical attributes, including packed cell volume, hemoglobin, sodium, potassium, blood pressure, blood glucose random, and blood urea. Additionally, it features nominal attributes such as specific gravity, albumin, and sugar. Other characteristics encompass pus cell clusters, bacteria, hypertension, anemia, thirst, hunger, heart disease, diabetes mellitus, and pedal edema. These 400 instances span the NOTCKD and CKD classes, collectively constituting 25 attributes. Within this dataset, 63% of individuals have CKD, while 37% do not.

To conduct our analysis and generate a prediction framework, we employed the widely-used R Programming Data Analytics tool. The subsequent table provides a glimpse of selected attributes and their values from the Chronic Kidney Disease Dataset, offering a glimpse of the rich and diverse information encompassed within this valuable dataset.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

Table 1: Sample Table Values for Chronic Kidney Disease Dataset

Age	BP	Glucose	Urea	Sodium	Potassium	Hemoglobin	RBC Count	Class
48	80	121	36	136	4.7	15.4	4.9	CKD
57	70	150	41	140	4.9	14.7	4.8	CKD
63	90	135	38	139	5.0	12.3	3.9	CKD
45	70	111	48	144	4.7	11.5	3.3	NOTCKD
50	80	118	39	137	5.1	13.8	4.0	NOTCKD

In our comprehensive analysis, we employed Python's scikit-learn library to construct various machine learning models for predicting Chronic Kidney Disease (CKD). Our dataset encompassed 400 records, each with 25 distinct attributes. The primary objective was to gauge the effectiveness of these models under diverse data imputation techniques, ultimately enabling us to identify the most robust and accurate predictive model. By thoroughly evaluating their performance, we aimed to enhance our understanding of CKD prediction, potentially contributing to more effective medical diagnoses and treatments. This rigorous approach signifies our commitment to leveraging data science and machine learning for the betterment of healthcare outcomes.

a) Decision Tree (DT)
b) K-Nearest Neighbors (KNN)
c) Support Vector Machines (SVM)
d) Random Forest (RF)
e) Naïve Bayes (NB)

We applied the following data imputation methods:

a) KNN Imputation
b) Mean and Mode Imputation
c) Forward Filling
d) Backward Filling

The table's significance extends beyond its simplicity; it offers a vital reference point encapsulating the training and testing accuracies for each machine learning model and the respective data imputation methods. This comprehensive data compilation plays a pivotal role in our CKD prediction study, enabling researchers and practitioners to swiftly gauge the effectiveness of various approaches. It serves as a compass, guiding us towards informed decision-making in model selection and data imputation strategy, ultimately aiming to enhance the accuracy of early CKD detection and, by extension, the efficacy of medical interventions in addressing this pressing healthcare concern.

Table 2: The performances of the models with different Imputation methods

Models	**KNN Imputation**		**Mean and Mode**		**Forward Filling and Backward Filling**	
	Training	Testing	Training	Testing	Training	Testing
Decision Tree	100%	97.5%	100%	99.1%	100%	99.16%
K-Nearest	78.21%	63.35%	94.64%	93.33%	100%	96%
SVM	96.25%	95%	100%	100%	100%	100%
Random Forest	100%	95%	100%	100%	100%	100%
Naïve Bayes	96.42%	94.16%	96.45%	95%	98.21%	99.16%

Summary Table 2 offers a comprehensive overview of the predictive performance of various machine learning models in the context of Chronic Kidney Disease (CKD) prediction, with a focus on different data imputation techniques. The models under scrutiny encompass Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest (RF), and Naïve Bayes, showcasing their versatility and applicability. The table meticulously reports accuracy percentages for both training and testing datasets, considering four distinct data imputation strategies:

a) **KNN Imputation:** This approach harnesses the power of K-Nearest Neighbors to impute missing values, capitalizing on similarity measures to replace absent data points.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

b) **Mean & Mode Imputation:** In this method, the imputation process relies on the statistical metrics of the dataset. Missing values are substituted with the mean for numerical attributes and the mode for nominal attributes, ensuring data completeness.

c) **Forward Filling:** Here, the strategy for filling in missing values involves replacing them with the most recent valid observation in the dataset, facilitating the retention of temporal patterns.

d) **Backward Filling:** Conversely, the Backward Filling method addresses missing values by replacing them with the subsequent valid value in the dataset, optimizing the utilization of available information.

The comprehensive table, which encapsulates training and testing accuracies in the context of various data imputation methods and machine learning models, is an indispensable asset for researchers and healthcare professionals working in the domain of Chronic Kidney Disease (CKD) prediction. It transcends the mere presentation of numerical results; it provides a deeper understanding of the intricate relationship between data imputation strategies and the performance of predictive models. This multifaceted perspective empowers decision-makers, including researchers, healthcare practitioners, and data scientists, to discern the strengths and weaknesses of different approaches. Such insight is instrumental in selecting the most appropriate method for CKD prediction, underlining the pivotal connection between data quality and the accuracy of predictions. By doing so, it contributes significantly to enhancing the precision and dependability of CKD diagnoses and treatments, ultimately improving healthcare outcomes in this critical field of medicine.

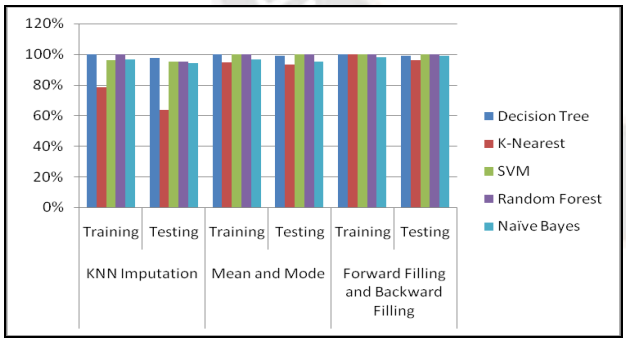

Figure 2: Accuracy comparison related different models

The table provides the accuracy percentages for every model and imputation method combination. For instance, the Decision Tree model with KNN Imputation has a 100% training accuracy and a 97.5% testing accuracy. Selecting the optimal model-imputation combination for CKD prediction is made possible by evaluating several models and techniques in a similar manner.

Decision Tree and Random Forest obtained 100% accuracy during training, whereas Decision Tree and Mean and Mode demonstrated high testing accuracy (99.10% and 99.16%) in Figure 2's model accuracy comparison. While K-Nearest Neighbours had a good testing accuracy of 93.33%, SVM and Random Forest likewise achieved flawless training accuracy. Naïve Bayes demonstrated good performance, with training accuracy of 96.42% and testing accuracy of 95% for Mean and Mode. In testing, Forward Filling outperformed Backward Filling with a 96% accuracy rate, but both methods were 100% accurate in training. K-Nearest Neighbours and Naïve Bayes demonstrated marginally lower testing accuracy, but Decision Tree and SVM demonstrated especially good performance overall.

VI. CONCLUSION AND FUTURE SCOPE

This paper addresses the challenging issue of chronic kidney disease (CKD), a covert pandemic characterized by rising rates of morbidity and death, particularly in the early, asymptomatic stages. Genetic predisposition, diabetes, obesity, and hypertension are a few of the variables that fuel the CKD epidemic and make it more challenging to detect and treat the disease early. By thoroughly comparing supervised classification techniques including Support Vector Machine (SVM), Random Forest, and Decision Tree, we aimed to increase the accuracy of CKD predictions. Additionally, we investigated several strategies for handling missing data and carefully evaluated their efficacy in comparison to various data cleaning techniques. Our results demonstrate the outstanding performance of the Random Forest and SVM models, particularly when considering the full dataset. Furthermore, we offer a comprehensive study of accuracy percentages for each model-imputation coupling, enabling the selection of optimal pairings for CKD prediction. In order to improve predictive accuracy, future studies on Chronic Kidney Disease (CKD) prediction could investigate ensemble models and use meta-heuristic algorithms for hyper parameter optimization. Real-time healthcare apps for early intervention should be developed, and robust feature engineering methodologies should be further explored. Working together with medical professionals can yield domain-specific insights, and extensive implementation in clinical environments can confirm usefulness. Ethical considerations and patient data privacy protection are important factors to take into account. Furthermore, prospects for thorough CKD risk assessment and management are provided by longitudinal data analysis and the incorporation of novel data sources as genomics and wearable's.

References

[1] Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., & Yang, C. W. (2013). Chronic kidney disease: Global dimension and perspectives. The Lancet, 382(9888), 260-272.

[2]. Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. The Lancet, 379(9811), 165-180.

[3]. Thomas, M. C., Cooper, M. E., Zimmet, P., & P. Shaw, J. (2015). Global burden of kidney disease and risk factors for type 2 diabetes. In Global Perspectives on Diabetes: Looking Beyond the Numbers (pp. 109-124). Springer.

[4].Sarafidis, P. A., Ruilope, L. M., Insua, J. T., Safar, M. E., & Ruilope, L. M. (2012). Blood pressure reduction in dialysis: Role of salt and volume management. Nephrology Dialysis Transplantation, 27(6), 2136-2139.

[5].Selvakumar, C., Sathik, M. M. S., & Bose, I. (2019). A novel framework for classifying chronic kidney disease stages using probabilistic neural network. Soft Computing, 23(9), 2939-2948.

[6].Ahmed, M., Mahmood, A. N., & Hu, J. (2018). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.

[7].Ejerblad, E., Fored, C. M., Lindblad, P., Fryzek, J., McLaughlin, J. K., Nyrén, O., & Elinder, C. G. (2006). Obesity and risk for chronic renal failure. Journal of the American Society of Nephrology, 17(6), 1695-1702.

[8].Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O'Callaghan, C. A., Lasserson, D. S., & Hobbs, F. D. (2016). Global prevalence of chronic kidney disease - A systematic review and meta-analysis. PLoS One, 11(7), e0158765.

[9]. Liu, Y., Wang, J., Zhong, W., & He, J. (2015). Using random forest to predict missing data in a national health survey. Computational and Mathematical Methods in Medicine, 2015, 1-10.

[10]. Manogaran, G., Lopez, D., & Thiagarajan, K. (2017). A survey of big data architectures and machine learning algorithms in healthcare. Journal of King Saud University-Computer and Information Sciences.

[11]. Sampson, M. G., Robertson, C. C., Martini, S., Mariani, L. H., Lemley, K. V., Gillies, C. E., ... & Kretzler, M. (2015). Integrative genomics identifies novel associations with APOL1 risk genotypes in Black NEPTUNE subjects. Journal of the American Society of Nephrology, 26(4), 814-825.

[12]. M. P. N. M. Wickramasinghe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300-303.

[13]. H. A. Wibawa, I. Malik and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease," 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2018, pp. 1-4

[14]. U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve bayes classifier," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5.

[15]. H. Zhang, C. Hung, W. C. Chu, P. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 1351-1356.

[16]. J. Aljaafet al., "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, 2018, pp. 1-9.

[17]. Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6.

[18] CH, R. ., Batra, I. ., & Malik, A.(2021) Comparative Analysis on Blockchain Technology Considering Security Breaches, Proceedings of Trends in Electronics and Health Informatics. Lecture Notes in Networks and Systems,vol.376.Springer,Singapore. \

[19]. G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 973-979.

[20]. Kamalesh, M. D. (2016). Predicting the risk of diabetes mellitus to subpopulations using association rule mining. In Proceedings of the International Conference of Soft Computing Systems, Advances in Intelligent Systems and Computing (Vol. 397). Springer.

[21] CH, R. ., Batra, I. , & Malik, A. . (2021) Combining Blockchain Multi Authority and Botnet to Create a Hybrid Adaptive Crypto Cloud Framework, ICCS-2021(IEEE Xplore Digital Library).

[22]. Revathy, B., Parvathavarthini, B., & Shiny Caroline. (2016). Decision Theory: An Unprecedented Validation Scheme for Rough-Fuzzy Clustering. International Journal on Artificial Intelligence Tools, 25(2), World Scientific Publishing Company.