



26 (XGBoost). XGboost was noted as the best performer with the AUC score of 0.98 and accuracy  
27 of 95.67%.

28 **Results:** The findings revealed that XGBoost was highly effective in predicting chronic kidney  
29 disease, achieving an accuracy of 95.76% and an AUC score of 0.98. Given that the dataset was  
30 collected from the local population, this study confirms that machine learning algorithms can  
31 assist clinicians in Rwanda in diagnosing chronic kidney disease in its early stages.

32 **Conclusion:** This study demonstrates the potential of machine learning models in predicting  
33 chronic kidney disease (CKD) in diabetic patients, highlighting the importance of local datasets  
34 for optimizing model performance in specific populations. These findings suggest that machine  
35 learning can effectively assist existing medical techniques in the early diagnosis of CKD in  
36 Rwanda.

37 **Keywords:** Chronic Kidney diseases; Diabetes; Machine learning; XGBoost

38

39 **Author summary.**

40 In this study, we trained machine learning model to predict the risk of chronic kidney disease  
41 (CKD) in patients with diabetes, using a dataset collected in Rwanda. Early detection of CKD is  
42 crucial, as it allows healthcare providers to intervene sooner, improving patient outcomes,  
43 potentially reducing financial, and health burden on the patients. We processed the data, by  
44 handling different available data issues and statistically created new features such as anemia  
45 status and length of hospital stay to improve the model's predictions. The final model, XGBoost  
46 provides insights that it can help health providers to identify high-risk patients and plan  
47 personalized care more effectively.

48 This study highlights how data-driven solutions can support healthcare delivery in resource-  
49 limited settings, by enhancing early diagnosis especially at primary healthcare level. By  
50 integrating this predictive tool into routine clinical workflows of Electronic Medical Record,  
51 healthcare institutions can make better clinical decisions that improve patient care and  
52 outcomes. This project contributes to the growing field of health informatics in Africa and  
53 shows the potential of applying advanced analytics to solve local health challenges.

54 **Introduction.**

55 While increasing attention is paid to the rising prevalence of chronic diseases in Africa, there is  
56 little focus on chronic kidney disease (CKD) though it directly resulted in an estimated 1.23  
57 million deaths worldwide in 2017 and 1046 Deaths in Rwanda in the same year with an  
58 estimated prevalence of 10% of adults aged more that 18years old (1) (2). According to the

59 International society of nephrology (ISN), worldwide in 2020 about 850 Million People were  
60 suffering chronic kidney disease at different stages, and its prevalence has a rising trend (3).  
61 CKD is one of the leading causes of cardiovascular diseases, which are then the most common  
62 cause of death worldwide (4).

63 CKD is a progressive health condition where the kidneys gradually lose their ability to perform  
64 its functions properly. It affects people of all ages and all backgrounds but it is most common in  
65 older people and those with underlying medical conditions like diabetes and/or hypertension.  
66 For the people who has CKD, there is a decline in overall kidney function leading to losing the  
67 ability to filter out waste products and maintaining fluid and acid balance (5) .

68 As kidney progresses, it leads to more serious complications that needs critical medical  
69 attention and these complications affect multiple organ systems and significantly affect the  
70 overall status of health. CKD often causes or worsen blood pressure which further the damage  
71 on the kidneys and increase the risk of heart diseases like heart failure, coronary artery diseases  
72 and stroke (6).

73 Currently, in Rwanda to diagnose Chronic Kidney diseases, a urine examination is a must  
74 irrespective the level of the healthcare institution with the aim of dipstick test, microspic  
75 analysis and identifying the quantity of protein that the kidney is able to excrete (7). In  
76 microscopic analysis, the target to look for the presence of essential blood cells in the urine.  
77 The presence of Red Blood Cells (RBC) in the urine, known as hematuria, suggest glomerular  
78 damage, hence the cells leak in urine (8).

79 Instead of dipstick, Albumin – to- creatinine ratio (ACR), a more accurate technique is used in  
80 some cases as it is considered to be the best measure of proteinuria levels (9). It gives  
81 standardized measure of proteinuria making it easier to compare results across different  
82 settings with different equipment's. It is also believed to be more sensitive than dipstick in  
83 detecting micro albuminuria which is an early sign of CKD (10). However, ACR can be  
84 influenced by the flow rate which result to not always accurately reflect proteinuria. Also, in  
85 cases with severe proteinuria, ACR is prone to failing to accurately reflect the total protein  
86 excretion (11).

87 eGFR is another test kidney functioning test. it is actually a measure to know how the kidney is  
88 able to filter waste from the blood in the process of making urine though it can be slightly  
89 affected by other conditions like dehydration or existing acid-base disorders (12). It is calculated  
90 based on blood creatinine levels, age, gender and in some cases race (13). High eGFR, >90ml,  
91 indicate well-functioning kidneys while low eGFR indicates severely damaged kidneys; when it  
92 is <15ml/kg the suffering patient have to depend only on renal replacement therapy like dialysis  
93 for lifetime or surgical kidney transplant (14).

94 While the mentioned tests are valuable for evaluating kidney function, they often fail to detect  
95 chronic kidney disease (CKD) in its early stages, when the kidneys are only minimally damaged.  
96 In these initial stages, the kidneys may still compensate for damage, leading to subtle changes  
97 in laboratory results that might not be apparent to measurement equipment (15). Additionally,  
98 factors such as age and gender can influence the interpretation of these results, increasing the  
99 risk of missed diagnoses (16).

100 With the rapid advancement of technology, clinical researchers, healthcare professionals, and  
101 ICT engineers are collaborating to explore the use of machine learning algorithms in healthcare  
102 settings, particularly for diagnosing non-communicable diseases including CKD, which often  
103 presents with no clinical symptoms at its early stage. Availability of electronic medical record  
104 (EMR) data is a great asset for not only healthcare administrators to help in administrative  
105 decision making, but also the health informaticians benefit this data for training ML algorithms  
106 that are able to assist in clinical decision making including early diagnoses (17).

107 Machine learning was proven as an important, evolving tool for accurately predicting the risk of  
108 CKD in patients with Diabetes, Hypertension, or other non-communicable diseases (18) (19).

109 With relevant, well-prepared data, ML algorithms are able to learn relationships between risk  
110 factors of a disease so that it can forecast a trend for unseen data. For CKD, the main factors  
111 associated are age, hypertension, diabetes, Glycated Hemoglobin (Hb1Ac), serum creatinine  
112 levels, proteinuria, and Estimated Glomerular Filtration Rate (eGFR), and others, from that, ML  
113 can predict CKD risk in new patients based on individual characteristics (20).

114 Machine learning has been widely used to predict CKD progression using algorithms like  
115 Random forest, support vector machines, and neural networks. A systematic review comparing  
116 these models demonstrated that RF and Gradient boosting to have excellent performance  
117 metrics results in handling non-linear data like eGFR and albuminuria levels (21) (22).

118 Also ML models have proven effectiveness in risk stratification of CKD patients which is crucial  
119 for interventions targeted on specific group. The study used XGBoost, Logistic Regression and  
120 decision trees to classify patients based on both clinical and demographic data. Here, XGBoost

121 provided best performance with an accuracy of 92% and precision of 90% outperforming other  
122 models in identifying patients at risk of CKD (23).

123 A study applied SVM and RF to clinical datasets, using clinical features like Blood pressure,  
124 diabetes status and serum creatinine levels. RF achieved a promising highest accuracy of 93%. It  
125 was able to achieve this because of strength in managing high dimensional data, and proved its  
126 potential use in routine clinical practice (24).

127 Deep learning (DL) techniques like convolutional neural networks (CNN) and recurrent neural  
128 networks (RNN) are also great in disease diagnosis. These models were applied on a large a  
129 clinical dataset including time-series la results and showed better performance compared to  
130 previous studies which used standard ML algorithms. There was improvement of 5-10%  
131 improvement over standard algorithms (25).

## 132 **Methodology.**

133 In this research study, we utilized a comprehensive dataset extracted from the OpenClinic  
134 system of the University Teaching Hospital of Kigali. This dataset specifically focuses on patients  
135 diagnosed with diabetes who were managed at this facility. A range of clinical data,  
136 demographic information, and laboratory results were collected. The study retrospectively  
137 covers a period of ten years, from 2013 to 2023.

138 To train the model, the dataset was split in a 75:25 ratio, meaning that 75% of the entire dataset  
139 was used for model training, while the remaining 25% was reserved for model testing. Four  
140 machine learning classifiers—Random Forest, Extra Gradient Boosting (XGBoost), Decision Tree,

141 and Logistic Regression—were trained and evaluated based on the Accuracy and Area Under the  
142 Curve (AUC) of the Receiver Operating Characteristic (ROC).

### 143 **Dataset Description.**

144 At the end of data collection, we compiled a dataset containing 6,900 instances and 29 features. Among  
145 these, 10 were numerical variables and 19 were categorical variables. To enhance the comprehensiveness  
146 of the dataset, five additional features were statistically engineered: anemia status, derived from  
147 hemoglobin levels; days of stay, representing the number of days the patient spent in the hospital; and  
148 medical insurance status, indicating whether the patient had insurance coverage. Additionally, 10 features  
149 that were not relevant to the model's training were excluded, resulting in 24 features being retained. Each  
150 individual feature is illustrated in **Figure 1**.

151 *Table 1. Description of the dataset*

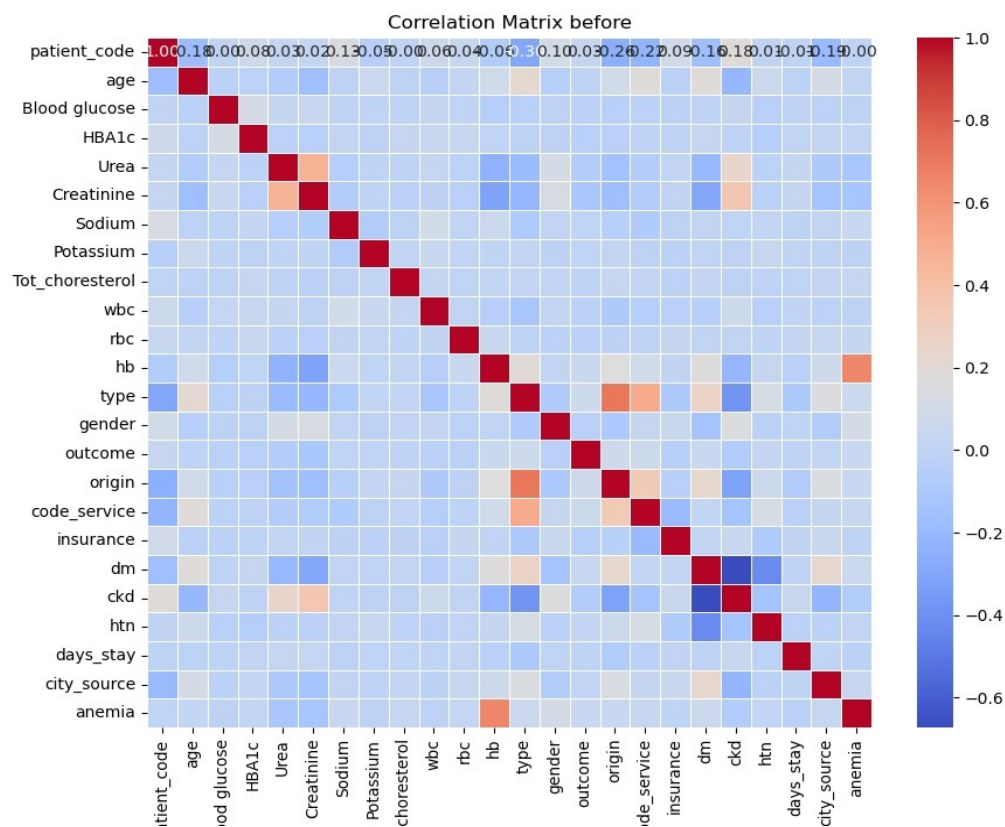
Description	Values
Data source	Kigali University teaching Hospital
Period	2013 to 2023
No of instances	2670
Number of features	26
Have CKD	446
Do not have CKD	2204

152  
153 The dataset was thoroughly examined for data quality issues. No duplicates were identified, and  
154 no feature exhibited missing data exceeding 35%. To address the remaining missing values, the



155 K-Nearest Neighbors (KNN) imputation method was employed, ensuring appropriate data filling.  
156 Additionally, proper data formatting was conducted to ensure correct interpretation by Python.  
157 An assessment for multicollinearity indicated there was minimal correlation issue, as illustrated  
158 in Figure 3. Grid Search was used to select and tune the best hyperparameters.

159 *Figure 1. Correlational matrix of numerical features.*



160

161

162 **Machine learning classifiers.**

163 **Logistic regression.**

164 Logistic regression function is a function that was initially developed by Pierre Francois Verihulst,  
165 a Belgian Medical Doctor and Mathematician, in 1840s based on research on modelling  
166 population growth (26). Many subsequent researchers continued to work on it improving it until  
167 1960 where it become how we see it today. Logistic regression is applied in different fields from  
168 social sciences, public health and medicine, to machine learning. This function is also used in  
169 engineering to predict the success or failure of a system or model (27).

170 Logistic regression models the probability of a binary outcome by mapping input variables to a  
171 value between 0 and 1 using the logistic (sigmoid) function. Its counterpart, linear regression  
172 predicts a continuous dependent variable based on a linear relationship between independent  
173 variables (27).

174 Mathematically Logistic regression for  $P: R \rightarrow (0,1)$  can be represented as follows:

$$175 \quad p = \frac{1}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n)}}$$

### 176 **Random Forest.**

177 Random forest also known as a random decision forest is used mostly in classification and  
178 regression tasks that are built on combining different decision trees through out training time.  
179 For classification tasks like this study, the output of decision tree is the class that have been  
180 selected most times (28).

### 181 **Decision Tree classifier.**

182 A decision tree is a visual and intuitive model used for both classification and regression tasks in  
183 machine learning. It represents decisions and their possible consequences in a tree-like

184 structure, where each internal node signifies a feature or attribute, each branch represents a  
185 decision rule, and each leaf node represents an outcome or final decision. The model splits the  
186 dataset into subsets based on feature values, aiming to improve prediction accuracy with each  
187 split. Decision trees are favored for their simplicity, ease of interpretation, and ability to handle  
188 both numerical and categorical data (29).

189 **Gini impurity** is calculated for each split to evaluate the impurity of the resulting subsets:

190 
$$Gini(D) = 1 - \sum_{i=1}^c (\rho_i)^2$$

191 **Entropy** measures the uncertainty in the dataset and is another criterion used to evaluate  
192 splits:

193  
194 
$$entropy = \sum_{i=1}^c -\rho_i \log(\rho_i)$$

195  
196 **XGBoost**

197 XGBoost (Extreme Gradient Boosting) is an advanced implementation of gradient boosting  
198 designed for speed and performance in machine learning tasks, particularly for classification and  
199 regression. It builds an ensemble of decision trees in a sequential manner, where each new tree  
200 corrects errors made by the previously built trees. The model minimizes a loss function, such as

201 mean squared error for regression, and incorporates regularization to prevent overfitting. The  
202 core of XGBoost optimization process can be captured by its objective function:

$$\text{Obj}(\theta) = \sum_{i=1}^N \text{Loss}(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

203

## 204 **Results and discussion.**

205 In this study, data issues were handled, hyperparameters tuned and 4 machine learning  
206 classifiers trained. Each algorithm used was described in methodology part, above.

207 To evaluate the performance of each model, a confusion matrix was generated. A confusion  
208 matrix is a structure like a table which is outputted to show the key performance metrics like:  
209 accuracy, Recall, Precision, F1-score and specificity.

210 *Table 2. Confusion matrix metrics with formulas.*

<b>Evaluation Metric</b>	<b>Formula</b>
Precision	Precision = TP / TP + FP
Recall	Recall = TP / TP + FN (30),
F1 Score	F = (Precision*recall / precision*recall)*2 (31)
Accuracy	Accuracy = TP + TN / TP + FP + TN + FN

211

212 In this study, True Positives (TP) refer to instances where the model correctly predicts the  
213 presence CKD. True Negatives (TN) indicate when the model accurately predicts its absence.  
214 False Positives (FP) occur when the model incorrectly identifies CKD when it is not present,

215 potentially causing unnecessary anxiety. False Negatives (FN) happen when the model fails to  
216 detect CKD when it is actually present, risking untreated conditions (26).

217 To obtain the classification report, various constituent variables are used to calculate the  
218 parameters essential for determining the accuracy score. Recall, or True Positive Rate,  
219 commonly referred to as sensitivity in epidemiology, is calculated using the formula provided in  
220 Table 1. The precision/positive predictive value (PPV) is calculated using the formula in Table1,  
221 while F1 score also known as tradition f-measure is the balance mean between the sensitivity  
222 and precision (26). The detailed description of classification report got by this study from all  
223 trained models is described in Table 2.

224 *Table 3. Classification results achieved by this study.*

<b>Algorithm</b>	<b>AUC score</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Accuracy achieved</b>
Random Forest	0.94	0.98	0.95	0.96	0.96
XGboost	<b>0.98</b>	<b>0.96</b>	<b>0.96</b>	<b>0.96</b>	<b>0.95</b>
Decision tree	0.89	0.84	0.84	0.84	0.83
Logistic regression	0.98	0.95	0.95	0.95	0.95

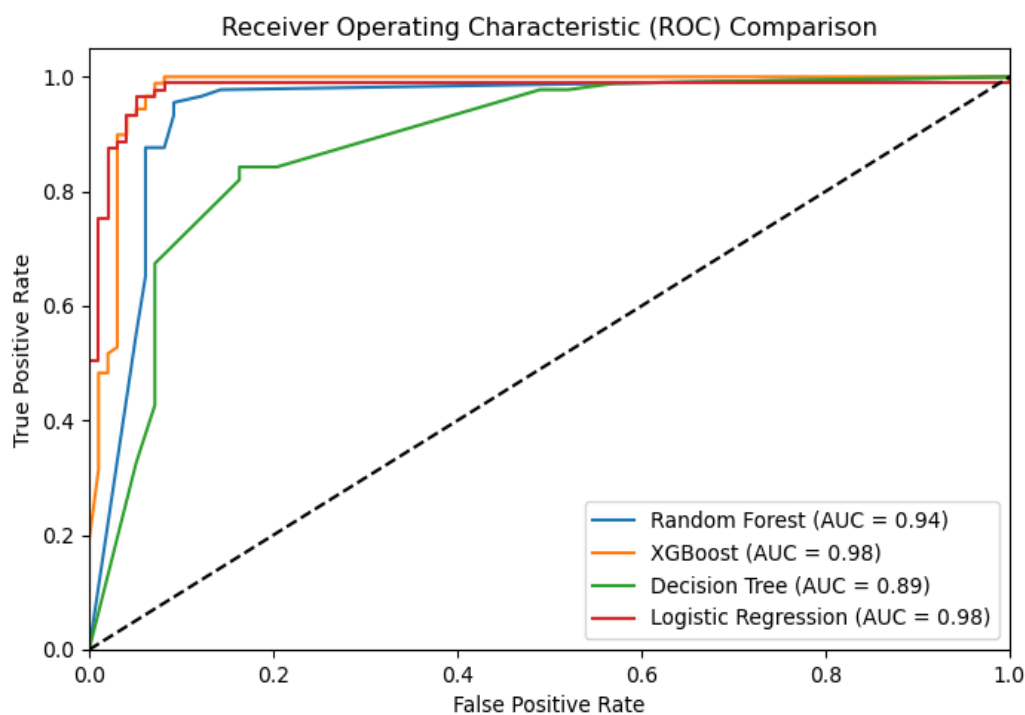
225

---

226 To further evaluate our models' performance, we plotted the ROC curves and calculated the  
227 AUC scores, as shown in Figure 1. The results indicate that XGBoost and Logistic Regression  
228 achieved the highest AUC score of 0.98, while the Decision Tree model had the lowest AUC  
229 score of 0.89.

230 Upon comparing model performance, both Logistic Regression (LR) and XGBoost achieved the  
231 same AUC and accuracy scores. However, they differed in their F1 scores, with LR scoring 0.95  
232 and XGBoost scoring 0.96. Given these results, XGBoost was selected as the best model due to  
233 its superior F1 score, indicating better balance between precision and recall.

234 *Figure 2. comparison of achieved AUC ROC results.*



235

236

237

238

239

240 *Table 4. comparison of our results with previous studies*

Author	Model	Accuracy	AUC
Wang et al.	Random forest	87.50%	0.92
Liu et al.	XGBoost	92.30%	0.96
Yashfi et al.	Random Forest	97.12%	0.98
Zhang et al.	Random Forest	89.4%	0.93
<b>This current study</b>	<b>XGBoost</b>	<b>95.76%</b>	<b>0.98</b>

241

242 The results of our study demonstrate the efficacy of the XGBoost model in predicting chronic  
243 kidney disease, achieving an accuracy of 95.76% and an AUC of 0.98. These findings position our  
244 model competitively within the existing literature.

245 When comparing our results to those of previous studies, we note that Yashfi et al. reported a  
246 remarkable accuracy of 97.12% with a Random Forest model and an AUC of 0.98 (27). While  
247 our accuracy is slightly lower, the high AUC indicates that both models are similarly effective in  
248 distinguishing between classes. Liu et al. also achieved a commendable accuracy of 92.30% with  
249 XGBoost and an AUC of 0.96, reinforcing the strength of this algorithm in this context (28).

250 In contrast, Wang et al. and Zhang et al. reported lower accuracy rates of 87.50% and 89.4%,  
251 with corresponding AUC values of 0.92 and 0.93, respectively (29) (30). These findings suggest  
252 that while Random Forest can be effective, our XGBoost model outperforms these earlier  
253 results, demonstrating enhanced predictive capability.

254 The consistent high performance of XGBoost across studies highlights its robustness as a  
255 machine learning tool for medical predictions, particularly for chronic kidney disease.

256 Additionally, the lack of statistically significant differences in accuracy between our results and  
257 those of Yashfi et al. underscores the reliability of our findings and their potential applicability  
258 in clinical settings.

259 The promising results of this study proves the transformative potential of machine learning in  
260 the diagnosis of chronic kidney diseases. As healthcare currently adopting advanced analytics,  
261 this developed model, XGBoost can significantly enhance early diagnostic accuracy and  
262 efficiency when it is deployed in EMR systems. This capability not only helps in early stages  
263 diseases detection but also supports personalized treatment strategies prescribed to individual  
264 patient profiles. The scalability of XGBoost further facilitates its implementation in diverse  
265 clinical settings, improving patient outcomes and optimizing resource allocation in healthcare  
266 systems. Using ML in CKD diagnosis presents an opportunity to revolutionize patient care,  
267 making it more proactive and data-driven.

#### 268 **Limitations.**

269 This study was limited by its failure to identify and analyze chemical biomarkers in the blood of  
270 CKD patients. Research has consistently demonstrated the significance of these biomarkers in  
271 understanding disease progression and developing targeted treatments. Additionally, the study  
272 did not investigate genetic predisposition, which also play role in developing CKD.

#### 273 **Conclusion.**

274 In this study, we utilized a dataset from the University Teaching Hospital of Kigali's electronic  
275 medical records (EMR), which included a range of demographic and clinical variables. The  
276 dataset comprised various data types, including categorical variables. From this data, we



277 developed a predictive model for chronic kidney disease using four machine learning  
278 algorithms: Random Forest, XGBoost, Decision Tree, and Logistic Regression.  
279 To evaluate the models' performance on unseen data, we generated a classification report and  
280 calculated the AUC. These metrics allowed us to assess the accuracy, precision, recall, and  
281 overall effectiveness of each model in predicting chronic kidney disease. XGBoost was noted as  
282 the best performing model with AUC, accuracy, F1score of 0.98, 0.95, 0.96 respectively. In  
283 contrast, the Decision Tree model exhibited lower performance, with an AUC of 0.89, an  
284 accuracy of 0.83, and an F1 score of 0.84. These results highlight the superiority of XGBoost in  
285 effectively predicting chronic kidney disease compared to the other algorithms tested. By  
286 applying this in clinical settings, it can positively impact patients' health and assist medical  
287 practitioners to diagnose chronic kidney diseases in its early stages.

## 288 **Recommendations.**

289 To enhance the predictive power of the models developed in this study, it is recommended to  
290 increase the number of clinical features included in the analysis. By incorporating additional  
291 relevant variables—such as biochemical markers of CKD, and lifestyle factors—the model's  
292 ability to capture the complexity of chronic kidney disease.

293 Additionally, exploring ensemble techniques could further boost model performance.  
294 Combining predictions from multiple algorithms, such as Random Forest and XGBoost, could  
295 leverage the strengths of each approach and mitigate individual model weaknesses. Future  
296 studies should focus on this point, which are crucial to improve predictive modelling in  
297 healthcare.

298 **Acknowledgement.**

299 The authors would like to express their gratitude to the Centre of Excellence in Biomedical Engineering  
300 and e-Health at the University of Rwanda for providing access to educational resources and continuous  
301 support throughout this project

302 **References.**

- 303 1. **Purcel , Carmel, Smith, Mali and Vos , Theos.** *Global, regional, and national burden of chronic kidney*  
304 *disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017.* Washington :  
305 The Lancet, 2020.
- 306 2. **FDA, Rwanda.** *Fourth Health sector strategic Plan.* Kigali : Rwanda FDA, 2018.
- 307 3. *Assessment of Global Kidney Health Care Status.* **Bello , Aki and Levin, A.** 18, s.l. : JAMA Network,  
308 2017, Vol. 385.
- 309 4. *Chronic Kidney Disease and Cardiovascular Disease: Is there Any Relationship?* **Nathalie, Vallianou,**  
310 **Shah, Mitesh and Agathoniki, Gkogkou.** 1, 2019, Vol. 15.
- 311 5. *An Overview of Chronic Kidney Disease Pathophysiology: The Impact of Gut Dysbiosis and Oral*  
312 *Disease.* **Altamura, Serena, et al.** 11, s.l. : Pubmed Central, 2023, Vol. 11.
- 313 6. **Anupam , Agarwal and Karl , Nath.** Pathophysiology of Chronic Kidney Disease Progression: Organ  
314 and Cellular Considerations. *Chronic Renal Disease (Second Edition).* s.l. : Science Direct, 2020.
- 315 7. **Center, Rwanda Biomedical.** *National Strategy and Costed Action Plan for the Prevention and Control*  
316 *of Non-Communicable Diseases in Rwanda.* Kigali : Rwanda Biomedical Center;, 2020.
- 317 8. *Hematuria as a Risk factor for progression of chronic kidney disease and death: findings from the*  
318 *Chronic Renal Insufficiency Cohort (CRIC) Study.* **Orlandi, Paula F., et al.** 9, s.l. : BMC Nephrology, 2018,  
319 Vol. 32.
- 320 9. *Urinary albumin-to-creatinine ratio levels are associated with subclinical atherosclerosis and predict*  
321 *CVD events and all-cause deaths: a prospective analysis.* **Liu, Shanshan, et al.** 3, s.l. : PubMed Central,  
322 2021, Vol. 11. e040890.
- 323 10. *Diagnostic accuracy of urine dipstick testing for albumin-to-creatinine ratio and albuminuria: A*  
324 *systematic review and meta-analysis.* **Mejia, Jhonatan R., et al.** 11, Heliyon : s.n., 2021, Vol. 7.  
325 PMC8571083.
- 326 11. *Albuminuria, Proteinuria, and Urinary Albumin to Protein Ratio in Chronic Kidney Disease.* **Wu,**  
327 **Men-Tai, et al.** 2, s.l. : Journal of Clinical Laboratory analysis, 2012, Vol. 26. 10.1002/jcla.21487.
- 328 12. *The Glomerular Filtration Rate: From the Diagnosis of Kidney Function to a Public Health Tool.*  
329 **Cusumano, Ana Maria, Tzanno-Martins, Carmen and Rosa-Diez, Guillermo Javier.** s.l. : Frontier in  
330 Medicine, 2021, Vol. 8. 10.3389/fmed.2021.769335.

- 331 13. *Measured and estimated glomerular filtration rate: current status and future directions.* **Levey,**  
332 **Andrew S, et al.** 1, s.l. : PubMed, 2020, Vol. 16. 10.1038/s41581-019-0191-y..
- 333 14. *Comprehensive review of current management guidelines of chronic kidney disease.* **Elendu,**  
334 **Chukwuka, et al.** 23, s.l. : Wolters Kruwer Medicine, 2023, Vol. 102. 10.1097/MD.00000000000033984.
- 335 15. *How Do Kidneys Adapt to a Deficit or Loss in Nephron Number?* **Fattah, Hadi, Layton, Anita and**  
336 **Vallon, Volker.** 3, s.l. : American Physiology Society, 2019, Vol. 34. 10.1152/physiol.00052.2018.
- 337 16. *Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging.* **McMaughan,**  
338 **Darcy Jones, Oloruntoba, Oluyomi and Smith, Matthew Lee.** 230, s.l. : Frontiers in Public Health, 2020,  
339 Vol. 8. 10.3389/fpubh.2020.00231.
- 340 17. *Applications of Machine Learning Using Electronic Medical Records in clinical medicine.* **Schwartz,**  
341 **John T., et al.** 4, s.l. : PubmedCentral, 2019, Vol. 16. PMC6945000.
- 342 18. *Chronic kidney disease prediction based on machine learning algorithms.* **Islam, Ariful, Majumder,**  
343 **Ziaul Hasan and Hussein, Alomgeer.** s.l. : Elsevier Journal of Pathology Informatics, 2023, Vol. 14.  
344 10.1016/j.jpi.2023.100189.
- 345 19. *A comprehensive review of machine learning techniques on diabetes detection.* **Sharma, Toshita and**  
346 **Shahcor, Manan.** 30, 2021, Vol. 4. 10.1186/s42492-021-00097-7.
- 347 20. *Risk factors for chronic kidney disease: an update.* **Kazancioglu, Rumezya.** 4, 2013, Vol. 3.  
348 10.1038/kisup.2013.79.
- 349 21. *Predicting chronic kidney disease progression with artificial intelligence.* **Isaza-Ruget, Mario A., et al.**  
350 20, s.l. : BMC Nephrology, 2024, Vol. 148.
- 351 22. *Chronic kidney disease prediction based on machine learning algorithms.* **Islam, Ariful, Majumder,**  
352 **Ziaul Hasan and Hussein, Alomgeer.** s.l. : Elsevier Journal of Pathology informatics, 2023.  
353 10.1016/j.jpi.2023.100189.
- 354 23. *Prediction Model and Risk Stratification Tool for Survival in Patients With CKD.* **S, Alexander, et al.** 2,  
355 s.l. : Elsevier Kidney International report, 2018, Vol. 3. 10.1016/j.ekir.2017.11.010.
- 356 24. *Chronic kidney disease prediction based on machine learning algorithms.* **Islam, Ariful, Majumder,**  
357 **Ziaul Hasan and Hussein, Alomgeer.** 8, Ariful Islam; Ziaul Hasan Majumder; Alomgeer Hussein : Elsevier  
358 Journal of Pathology informatics, Vol. 14. 10.1016/j.jpi.2023.100189.
- 359 25. *Gene Prediction: Methods, Challenges, and Future Perspectives.* **Chen, Qiannan.** 6, Zhejiang : British  
360 Biomedical Bulletin, 2023, Vol. 11. 10.36648/2347-5447.11.2.4.
- 361 26. **Schober, Patrick and Vetter, Thomas R.** *Logistic Regression in Medical Research.* s.l. : PubMed  
362 Central, 2021. pp. 365-366. <https://doi.org/10.1213/ANE.0000000000005247>.
- 363 27. **Sperandei, Sandro.** *Understing Logistic regression.* New Dheli : Biochemia Medica, 2014. pp. 12-18.  
364 <https://doi.org/10.11613/BM.2014.003>.
- 365 28. **Breiman, Leo.** *Random Forests.* s.l. : SpringerLink, 2001. pp. 5-32.

- 366 29. *Decision tree methods: applications for classification and prediction*. **Song, Yan-Yan and Lu, Ying.** 2,  
367 Shanghai : Shangai Archive of Psychiatry, 2015, Vol. 27. [https://doi.org/10.11919/j.issn.1002-](https://doi.org/10.11919/j.issn.1002-0829.215044)  
368 0829.215044.
- 369 30. *A Review of Evaluation Metrics in Machine Learning Algorithms*. **Naidu, Grien, Zuva, Tranos and**  
370 **sibanda, Elias.** s.l. : Springer Link, 2023. 978-3-031-35313-0.
- 371 31. *Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and*. **Botchkarev,**  
372 **Alexei.** 5, Ontario : Semantic , 2023, Vol. 12.  
373 [https://www.bing.com/ck/a?!&&p=2023dd9629d57b82JmltdHM9MTcyOTU1NTIwMCZpZ3VpZD0yY2FmMjdjYi04YTExLTY5MDktMTNiNS0zNGNjOGJlZDY4NDImaW5zaWQ9NTE5MQ&p=3&ver=2&hsh=3&fclid=2caf27cb-8a11-6909-13b5-](https://www.bing.com/ck/a?!&&p=2023dd9629d57b82JmltdHM9MTcyOTU1NTIwMCZpZ3VpZD0yY2FmMjdjYi04YTExLTY5MDktMTNiNS0zNGNjOGJlZDY4NDImaW5zaWQ9NTE5MQ&p=3&ver=2&hsh=3&fclid=2caf27cb-8a11-6909-13b5-34cc8bed6842&psq=Performance+Metrics+(Error+Measures)+in+Machine+)  
374 [34cc8bed6842&psq=Performance+Metrics+\(Error+Measures\)+in+Machine+](https://www.bing.com/ck/a?!&&p=2023dd9629d57b82JmltdHM9MTcyOTU1NTIwMCZpZ3VpZD0yY2FmMjdjYi04YTExLTY5MDktMTNiNS0zNGNjOGJlZDY4NDImaW5zaWQ9NTE5MQ&p=3&ver=2&hsh=3&fclid=2caf27cb-8a11-6909-13b5-34cc8bed6842&psq=Performance+Metrics+(Error+Measures)+in+Machine+).  
375
- 377 32. *Classification Techniques in Machine Learning: Applications and Issue*. **Soofi, Aized Amin.** 76, s.l. :  
378 Journal of Basic & Applied Sciences, 2017, Vol. 13. 10.6000/1927-5129.2017.13.76.
- 379 33. *Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms*. **Yashfi, Shanila Yunus,**  
380 **Islam, Ashikul and Pritilata.** s.l. : Researchgate, 2020.
- 381 34. *Long-Term Performance Prediction Framework based on XGBoost Decision Tree for Pultruded FRP*  
382 *Composites exposed to Water, Humidity and Alkaline Solution*. **Liu, Xing, Liu, TianQiao and Feng, Peng.**  
383 s.l. : 284, 2022, Vol. 5.
- 384 35. *Molecular contrastive learning of representations via graph neural networks*. **Wang, Yuyang, et al.**  
385 s.l. : Nature Machin Learning, 2022, Vol. 4.
- 386 36. *Artificial intelligence in recommender systems*. **Zhang, Qian, Lu, Jie and Jin, Yaochu.** 12, s.l. :  
387 COMplex and Intelligent systems, 2020, Vol. 7.
- 388