Prediction Early Chronic Kidney Disease and its incidence using machine learning and neural networks Algorithm

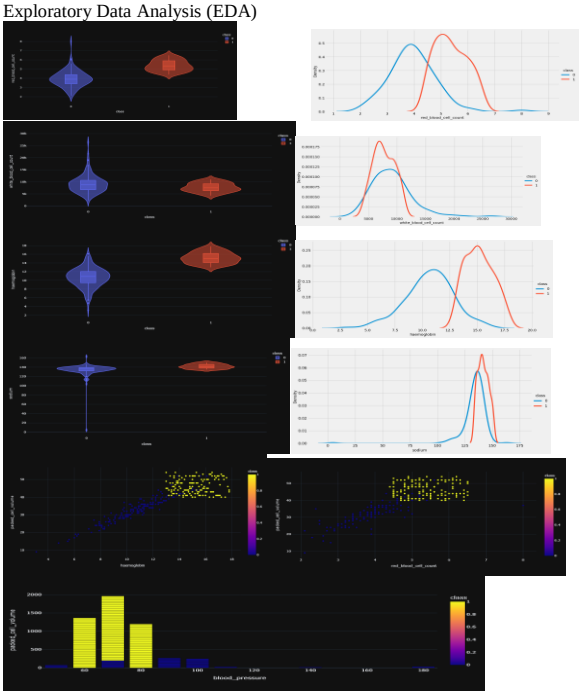
Dawit Shibabaw (PhD student at Bahir Dar University, Department of Data science) Email: dawitshibabaw14@gmail.com, Bahir dar Ethiopia

Abstract:-Chronic kidney disease is a general term for describing any disorders that lead to the gradual loss of kidney function or structure. According to a report from the World Health Organization (WHO) since 2022, 68 million peoples die due to chronic kidney disease, and forecast in 2030 it leads to 70 million people dying. As researcher's knowledge, related works show that there is some limitation/gaps that attempt to predict chronic kidney disease using machine learning and neural network Algorithms, we develop the predictive model for chronic kidney disease. The data were pre-processed to get quality data that are suitable for a machine/deep-learning algorithm to develop a model that predicts chronic kidney disease.

Introduction: According to a report from the World Health Organization (WHO) since 2022, 68 million peoples die due to chronic kidney disease, and forecast in 2030 it leads to 70 million people dying. In spite of that, people in developing countries are being affected by CKD. Among the major developing countries, Ethiopia is a highly problem faced country [1][2]. In Ethiopia, the emerging burden of non-communicable diseases becomes a serious public health concern which had a great impact on CKD. This is accompanied by poor knowledge of the early diagnosis of CKD. Besides its economic and medical impact, CKD becomes one of the reasons for psychological affections like depression in Ethiopia. Kidney disease is a serious public health problem in Ethiopia affecting hundreds of thousands of people irrespective of age, and sex [3]. To this end, different research was conducted on chronic kidney disease detection and predicting its incidence in different directions of the world and Ethiopia. For example N. C. Hodel et al. [4], D. A. Debal and T. M. Sitote [5], J. Snegha et al. [6], S. Y. Yashfi et al. [7], Rady and Anwar [8], However researchers concluded that deep learning scored better than statistical and classical machine learning compared with previous research that is conducted by statistical and classical machine learning methods. Directly applying research conducted in other the country is not applicable the chronic kidney disease because we do have a different dataset nature, the genetic makeup of patients, weather condition, and type of the food that we have.

Method

Exploratory Data Analysis (EDA)



Data Pre Processing

Filling for null values: researcher used to fill missed value filling null values, we will use two methods, random sampling for higher null values and # mean/mode sampling for lower null values and KNN imputation methods.

Feature Encoding: To encode the feature researcher used Label Encoder.

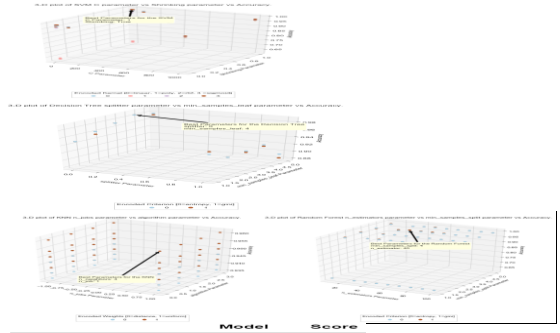
Outliers Handling: We identify the outlier using a boxplot and correct by using the Interquartile Range (IQR)

Class imbalance handling: researcher checked class imbalance problem and handle the SMOTE and TOMIC methods.

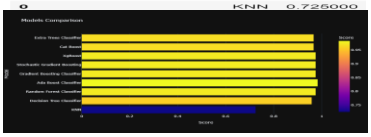
Feature selection: researcher applied both Filter (mi_best and chi_best) and wrapper (Step forward and Step backward) based on the Baseline fitness measure, so step backward selection techniques scored best having 15 features and tried all attributes for experimentation, while having all attributes scored batter than selected features.

Result and discussion

Hypermeter tuning using grid search



	Model	Score
3	Ada Boost Classifier	0.983333
2	Random Forest Classifier	0.975000
4	Gradient Boosting Classifier	0.975000
5	Stochastic Gradient Boosting	0.975000
6	XgBoost	0.975000
7	Cat Boost	0.966667
8	Extra Trees Classifier	0.966667
1	Decision Tree Classifier	0.958333
0	KNN	0.725000



Conclusion

This paper deals with the prediction of CKD in people. The 24 attributes present the best attributes are taken for prediction. Prediction is done using the Deep learning technique, ANN. In this classification problem

Acknowledge: - University of Gondar for financial support and Data source, advisory, classmate.

References

[1] S. Xie, Z. Yu, and Z. Lv, "Multi-disease prediction based on deep learning: A survey," C. - Comput. Model. Eng. Sci., vol. 127, no. 3, 2021, doi: 10.32604/CMES.2021.016728.

[2] J. Rashid et al., "An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction," Front. Public Heal., vol. 10, no. March, pp. 1–20, 2022, doi: 10.3389/fpubh.2022.860396.

[3] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," J. Big Data, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00657-5.

[4] N. C. Hodel et al., "The epidemiology of chronic kidney disease and the association with non-communicable and communicable disorders in a population of sub-Saharan Africa," PLoS One, vol. 13, no. 10, pp. 1–17, 2018, doi: 10.1371/journal.pone.0205326.

[5] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," J. Big Data, vol. 9, no. 1, pp. 137–140, 2022, doi: 10.1186/s40537-022-00657-5.

[6] J. Snegha, V. Tharani, S. D. Preetha, R. Charanya, and S. Bhavani, "Chronic Kidney Disease Prediction Using Data Mining," Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020, pp. 1–5, 2020, doi: 10.1109/ic-ETITE47903.2020.482.

[7] S. Y. Yashfi et al., "Risk Prediction of Chronic Kidney Disease Using Machine Learning Algorithms," 2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020, 2020, doi: 10.1109/ICCCNT49239.2020.9225548.

[8] E. H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," Informatics Med. Unlocked, vol. 15, no. December 2018, p. 100178, 2019, doi: 10.1016/j.imu.2019.100178.