PREDICTING THE PROGRESSION OF DIABETIC KIDNEY DISEASE USING MACHINE LEARNING

Adebanwo Habeeb-Lahi Adetola

Introduction

The study focuses on the significant impact of diabetes mellitus on renal morbidity and mortality, with diabetic nephropathy being the primary cause of chronic kidney failure in the United States[1]. Various renal lesions, especially involving the glomeruli, are associated with kidney disease in diabetic patients[2]. Chronic kidney disease (CKD) is a common complication in type 2 diabetes, leading to excess morbidity and mortality, including cardiovascular issues[3-6].

Microalbuminuria is an early sign of diabetic nephropathy, and without intervention, many diabetic patients progress to overt nephropathy over time[7]. Persistent hyperglycemia is considered responsible for long-term complications, including diabetic nephropathy, but other factors like hypertension and obesity also play a role[7].

CKD poses economic challenges due to the resources required for renal replacement therapy. Therefore identifying the contributing factors, early diagnosis and medical interventions halt CKD progression and improve the well being of diabetic patient[8].

The study aims to develop a predictive model for end-stage renal disease in type 2 diabetic patients with kidney diseases. It explores various multi-label classification models like RandomForestClassifier, XGBClassifier, LGBMClassifier, and CatBoostClassifier, comparing them with standard and state-of-the-art methods to identify factors contributing to diabetic kidney disease progression.

Experimental Data

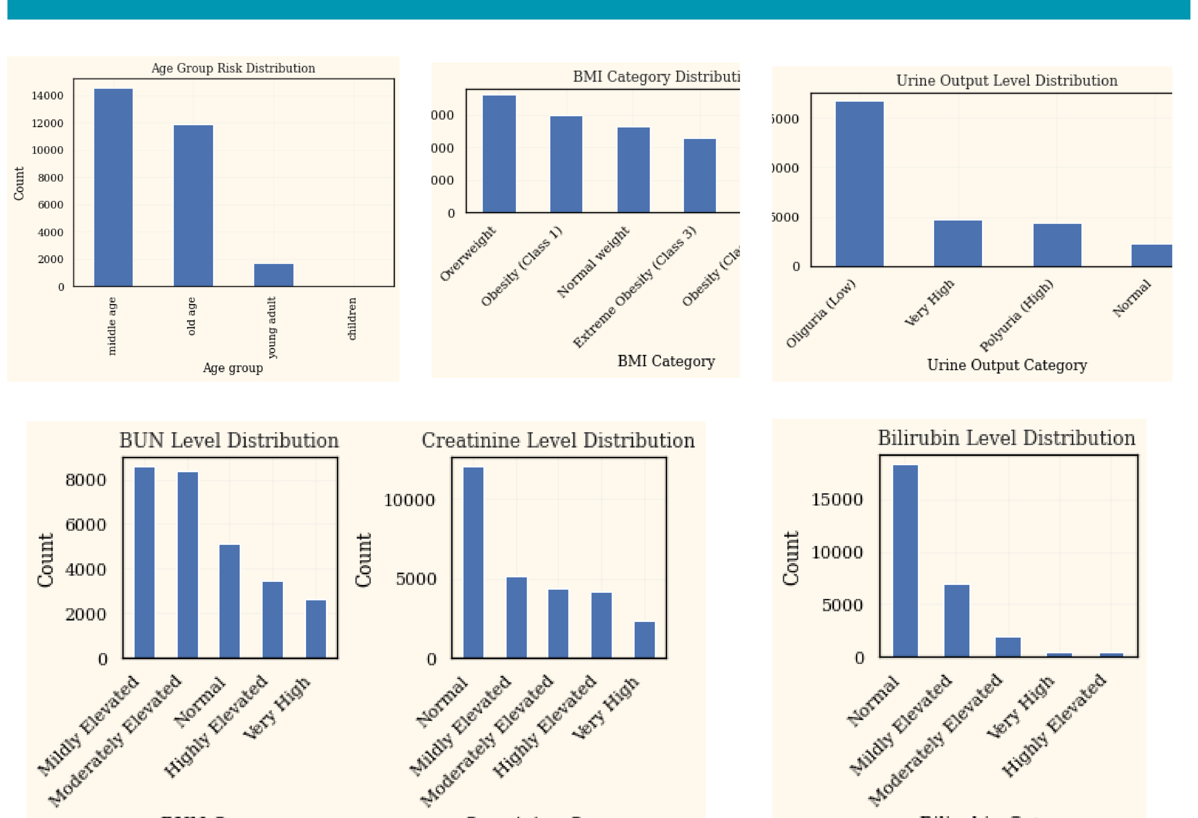
This approach involves utilizing the dataset available from the UCI Machine Learning Repository, consisting of 214 attributes and 130,157 instances with 9 distinct labels encompassing medical conditions like diabetes mellitus, renal failure, aids, cirrhosis, hepatic failure, immunosuppression, leukemia, lymphoma, and solid tumor with metastasis.

Data Processing

Prior to the training of machine learning algorithms for prediction, the dataset underwent preprocessing, including filtering out diabetic patients, as they were the subject of interest in this study. Relevant features were then extracted, leading to a final dataset with 26 attributes, encompassing the 9 labels and representing a total of 28,151 instances. All attributes in this dataset are numerical.

For the training of machine learning algorithms, the entire dataset with 28,151 instances was utilized, serving as the foundation for the predictive modeling process.
Data cleaning, which include handling missing and duplicate data was done using Pandas library. SMOTE was used to handle the imbalance dataset.

Data Distribution



Methodology

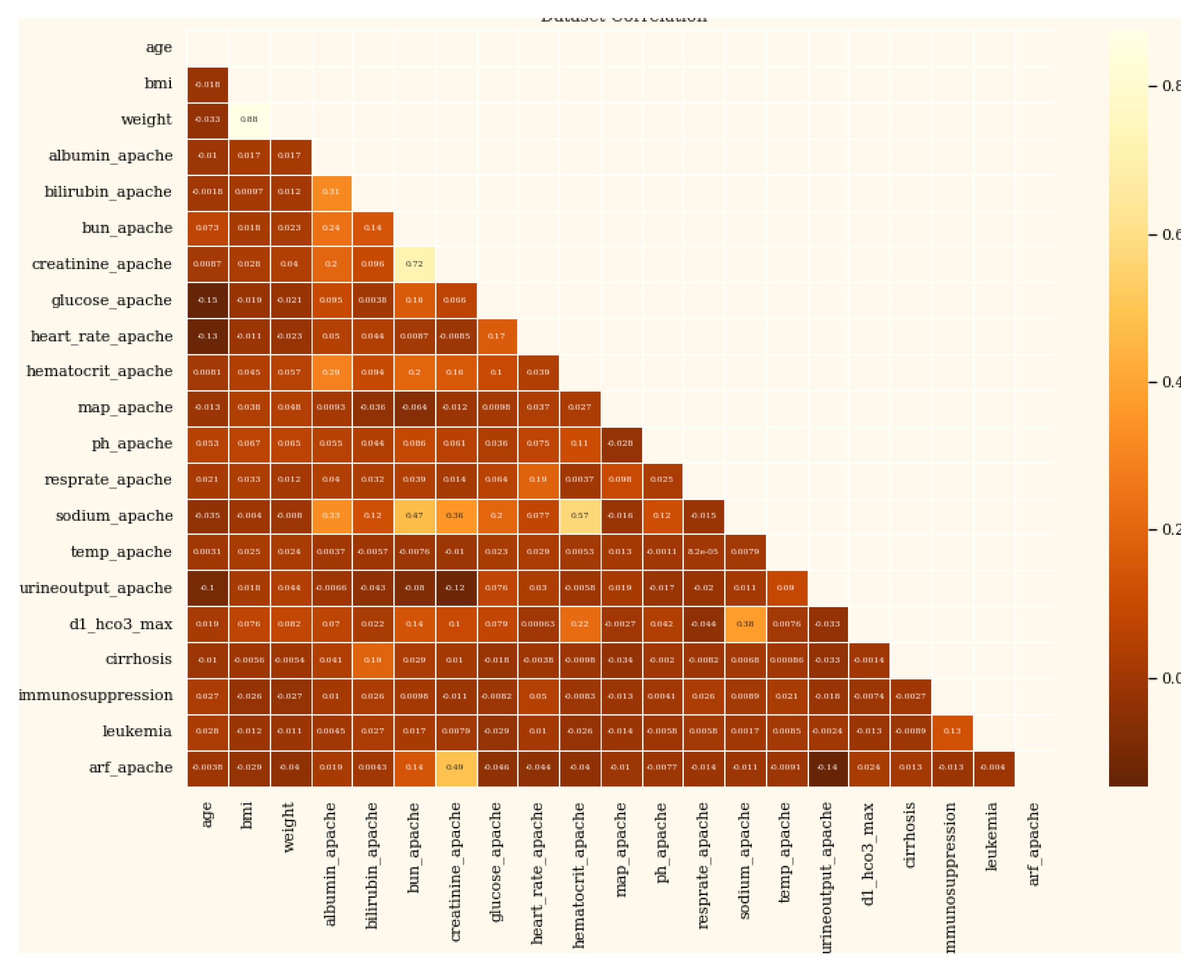
The dataset instances were divided into an 80:20 ratio for training and testing purposes, and this validation process was iterated 10 times.

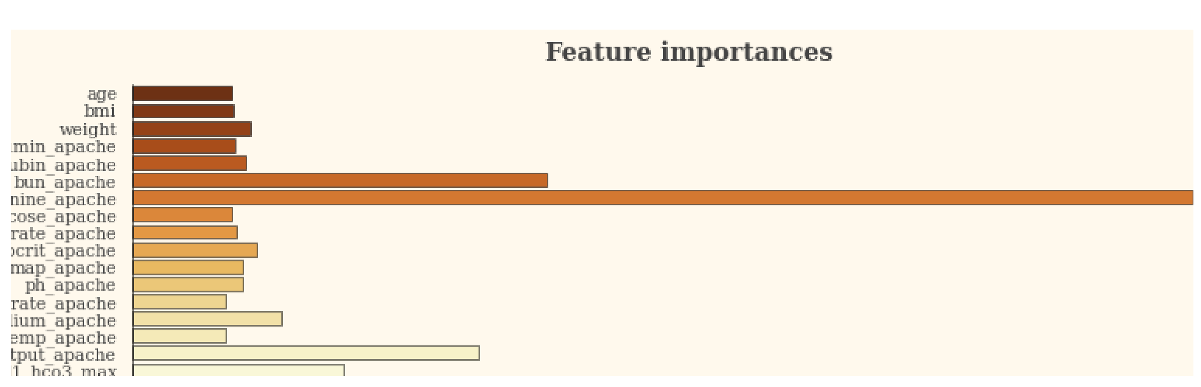
The level of multi-label complexity was assessed using two metrics: Label Cardinality and Label Density.

The assessment measures considered in this analysis include confusion matrix, area under the curve (AUC), and accuracy.

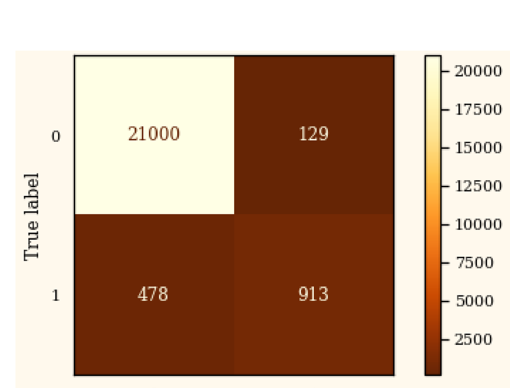
To analyze the dataset effectively, a series of experiments were conducted using diverse machine learning algorithms, such as RandomForestClassifier, XGBClassifier, LGBMClassifier, and CatBoostClassifier.

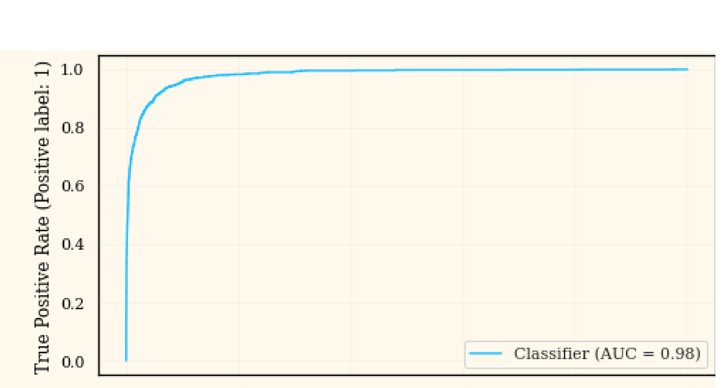
Results





Model	Training Accuracy	Test Accuracy	AUC
RF	100	96.09	0.9587
XGB	100	95.98	0.9568
LGM	100	96.03	0.9564
CatBoost	99.72	95.94	0.9562





Discussion

1. Exceptional Performance: The models showed exceptional performance with an accuracy rate exceeding 95% in accurately detecting Renal Failure in diabetic patients.
2. Comprehensive Parameters: The models relied on a comprehensive set of parameters, including age, BMI, albumin, bilirubin, blood glucose level, serum creatinine, blood urea nitrogen, hematocrit, mean arterial pressure, pH, respiratory rate, sodium level, urine output, and bicarbonate levels.
3. Promising Potential: The application of this predictive model holds promising potential for the medical community as it facilitates early diagnosis of kidney disease by effectively identifying predisposing factors.
4. Risk Prediction: The model helps medical professionals predict the likelihood of developing kidney diseases in diabetic patients by identifying individuals at risk and assigning them risk scores.
5. Efficiency: The method is efficient in predicting renal failure, requiring significantly less time for the prediction process, enabling early treatment initiation and classification of a broader population of patients within a shorter timeframe.
6. Dataset Limitations: The study acknowledges limitations in the dataset used, particularly the scarcity of datasets with the same comprehensive set of features, making meaningful comparisons with other datasets challenging. The dataset's parameters may not be sufficient for diagnosing comorbidities associated with kidney disease, requiring additional datasets and information for a more comprehensive assessment.

Conclusion

This research focused on developing an early-stage renal failure diagnosis model tailored for diabetic patients using various machine learning algorithms. The models were trained and validated using the discussed features. To enhance model efficiency, associations between different features were examined, identifying key attributes.

Filter feature selection showed that blood urea nitrogen, urine output, bilirubin, and bicarbonate were crucial predictors for renal failure. The study employed diverse machine learning algorithms to enable early kidney disease diagnosis in diabetic patients. Accuracy was the primary criterion for evaluating algorithm performance, aiming to provide healthcare professionals with a reliable tool for timely detection and improved management of diabetic kidney disease.

References

1. Robbins and Cotran Pathologic Basis of Disease. Glomerular lesions associated with systemic diseases ; Diabetic Nephropathy.
2. Ibrahim HN, Hostetter TH : Diabetic Nephropathy. J Am Soc Nephrol 8: 487, 1997.
3. Afkarian M, Sachs MC, Kestenbaum B, et al. Kidney disease and increased mortality risk in type 2 diabetes. JASN 2013;24:302–8.doi:10.1681/ASN.2012070718Abstract/FREE Full TextGoogle Scholar
4. Thomas MC, Brownlee M, Susztak K, et al. Diabetic kidney disease. Nat Rev Dis Primers 2015;1:1–20.doi:10.1038/nrdp.2015.18Google Scholar
5. Persson F, Rossing P. Diagnosis of diabetic kidney disease: state of the art and future perspective. Kidney International Supplements 2018;8:2–7.doi:10.1016/j.kisu.2017.10.003Google Scholar
6. Webster AC, Nagler EV, Morton RL, et al..Chronic kidney disease. Lancet2017;389:1238–52.doi:10.1016/S0140-6736(16)32064-5CrossRefPubMedGoogle Scholar
7. Robbins and Cotran Pathologic Basis of Disease. Clinical features of Diabetes.
8. Francesco Locatelli , Lucia Del Vecchio, Pietro Pozzoni. The importance of early detection of chronic kidney disease

Affliation: Olabisi Onabanjo University.
Email: adetolaprince@gmail.com
Phone:+2348169824649