Explainable AI for Cardiorenal Insights: Linking ECG Features to Kidney Function Decline in a Heart Failure Cohort

Sona M. Al younis[1], Namareq Widatalla[1], Amnaa Samjeed[1], and Ahsan H. Khandoker[1]

[1]Khalifa University of Science and Technology, Abu Dhabi, UAE

Abstract

The pathological pathways involved in the cardiorenal syndrome remains unknown and more research is needed to uncover this relationship. Here, we investigate the correlation of electrocardiogram (ECG) features in chronic heart failure (CHF) patients with renal function measured by estimated glomerular filtration rate (eGFR) We used the Sudden Cardiac Death in Chronic Heart Failure (MUSIC) data from Physionet database, to classify CHF patients into two classes (eGFR < 66 and eGFR > 66). Different machine learning (ML) models were employed, among which, the logistic regression (LR) model demonstrated the best performance, achieving an accuracy of 69.35%, a sensitivity of 70.71%, and an area under the receiver operating characteristic curve (ROC-AUC) of 0.752. Furthermore, explainable ML analysis revealed that Minimum RR Interval (RRI), QRS Duration, RR Range, and corrected (QTc) were identified as significant predictors of eGFR classes. These results pave the way for understanding the cardiorenal syndrome.

1. Introduction

Chronic kidney disease (CKD) is a major medical condition that affects around 8% - 16% of the population and it is characterized by a decline in kidney function [1]. Kidney function is estimated by glomerular filtration rate (GFR), however, measurement of GFR is time-consuming and expensive, hence, estimated GFR (eGFR) is clinically used for kidney function assessment [2]. Clinically, a persistent eGFR below 60 mL/min/1.73 m^2 for 3 months is considered abnormal [3]. CKD patients are prone to cardiovascular diseases (CVDs), and vice versa, CKD patients are prone to CVD and this is known as the cardiorenal syndrome [4]. The bidirectional relationship between the heart and kidney was confirmed previously [5]. However, the specific pathways leading to such deterioration is not fully understood and more research is needed to understand this interaction. Previously, Homstorm et al. [6] developed a Deep Learning model to predict CKD from electrocardiogram (ECG) records that were taken 1-year prior to CKD diagnosis. Another study by Kwon et al. used deep learning to detect renal impairment by using ECG [7]. Both studies demonstrated that ECG was linked to CKD. In our previous study, we demonstrated by using regression that heart rate (HR) measured from a 2-minute of ECG signal was associated with eGFR [8]. However, there is still a limited understanding of the intricate connection between ECG-derived parameters and CKD. Moreover, there is an increasing demand for noninvasive, less expensive tools to assess in the detection of CKD. Therefore, the aims of this study are as follows: 1) To report, the significance of ECG-derived parameters as clinical markers in early prediction of CKD. 2) To explore the potential use of different ML to categories CHF patients into two classes (eGFR < 66 mL/min/1.73m^2 and eGFR > 66 mL/min/1.73m^2) based on ECG features.

2. Methods

2.1. Data description

Data from the Sudden Cardiac Death in Chronic Heart Failure (MUSIC) available online in physionet were used [7, 8]. The full details of data collection and patient recruitment are explained in detail in [9]. Briefly, a total of 992 patients with chronic heart failure (CHF) enrolled at the specialized heart failure clinics of eight University Spanish Hospitals were recruited between April 2003 and December 2004. A 12-lead ECG, and blood samples were collected at enrollment. Patients were excluded if they had recent acute coronary syndrome or severe valvular disease amenable to surgical repair. |

Table 1: Medical records of the utilized dataset

Feature	Overall (992)	Class 0 (496)	Class 1 (496)	p-value
eGFR(mL/min/1.73$m^{2)}$	66.8±6.2	49.3±11.1	83.8±12.1	-
Gender (M/F)	718/274	327/169	391/105	<0.001
BMI (kg/m^2)	28.5± 4.5	28.6± 4.4	28.4 ± 4.5	0.473
DBP (mmHg)	74.2± 12.0	73.1 ± 11.4	75.4 ± 12.4	0.002
SBP (mmHg)	126.9± 22	126.1 ± 21	127.8 ± 22	0.223

Diabetes (Y/N)	356/636	203/293	153/343	<0.001
Hypertension	565/427	315/181	250/246	<0.001
PID (Yes/No)	34/958	25/471	9/487	0.001
ECG-Derived Features (ms)				
PR Interval	411.8±372	455.9±388	367.6±350	<0.001
QRS Duration	125.4±35.1	130.0±36	120.8±33	<0.001
QT Interval	412.4±48	416.8±51	407.9±45	0.004
QT Corrected	444.1±45	448.9±47	439.2±43	<0.001
Minimum RR	515.6±110	529.9±117	501.2±102	<0.001
Average RR	856.3±139	868.1±151	844.5±125	0.007
RR Range	727.9±405	714.7±409	741.0±401	0.307
Medication (Yes/No)				
Diabetes	283/709	159/337	124/372	0.013
Amiodarone	105/887	77/419	28/468	<0.001
Betablockers	675/317	301/195	473/122	<0.001
Digoxin	298/694	151/345	147/349	0.782
Loop Diuretics	721/271	416/80	305/191	<0.001
Hidralazina	28/964	24/472	4/492	<0.001
ACEInhibitor	734/258	354/142	380/116	0.060
Nitrovasodilator	282/710	169/327	113/383	<0.001

Variables epresented in (Mean±Std), BMI, Body mass index, DBP, Diastolic blood pressure, SBP, Systolic blood pressure, PID, Prior Implantable Device.

The MUSIC study complied with ethical standards, with the study protocol approved by institutional review boards. Informed consent was obtained from all participants, and the MUSIC database has been de-identified following General Data Protection Regulation (GDPR) guidelines to ensure the protection of patient health information (PHI) [7, 8]. In this study, we employed a binary classification approach to investigate the relationship between eGFR and ECG parameters. eGFR was calculated by using the creatinine-based CKD-Epidemiology Collaboration (EPI) (2021) equation [10].

To ensure a balanced study, we used a median-based threshold to classify the dataset into two classes: patients with eGFR values above the median (Class 1, eGFR > 66 mL/min/1.73m^2) and those below the median (Class 0, eGFR < 66 mL/min/1.73m^2). A threshold of 66 is close to 60 which is clinically used to diagnose CKD stage 3. The patient cohort comprised 274 females and 718 males. To investigate cardio-renal balance in CHF patients, we used the ECG-derived features that were provided online along with demographic and blood biochemistry. Table 1 summarizes the mean, standard deviation and counts of the features.

We used an independent two-sample t-test to evaluate whether there were significant differences between the two groups. For categorical variables, such as the gender feature, we used a chi-square test of independence to determine whether the distribution of the categories (e.g., Male and Female) differed significantly between the two groups. For each feature, the resulting smaller p-value (e.g., less than 0.05) indicates that the feature's distribution differs significantly between the low and high eGFR samples. Data processing and analysis were performed using Python, utilizing pandas for data management, scipy for statistical testing, and Seaborn and Matplotlib for generating visualizations such as feature importance plots. All data were utilized as inputs for the classification models.

Table 2. Classification model performance

Metrics	**DT**	**ANN**	**RF**	**LR**	**XGBoost**
Best hyperparameters combination	Max depth: [20] Min. samples leaf: [10] Criterion: [gini]	Hidden layer sizes: [50] Activation: [tanh] Solver: [adam], Alpha: [0.001]	Max depth: [10] Min samples split: [5] Estimators: [200]	Solver: [liblinear] Penalty: [l1] C: [10.0]	Max depth: [5] Estimators: [60] Learning rate: [0.01] Lambda: [0.5]
Accuracy (%)	54.27	52.26	68.84	**69.35**	65.33
Precision (%)	54.26	51.72	69.07	**69.13**	66.30
Sensitivity (%)	51.52	60.61	67.68	**70.71**	61.62
F1 Score (%)	52.85	55.81	68.37	**69.65**	63.87
MCC	0.08	0.0467	0.378	**0.387**	0.307
ROC AUC	0.558	0.523	0.688	**0.752**	0.702

Maximum values in each row are bold

2.2. Proposed Classification Models

To compare the effectiveness of various ML models, five models were implemented for the classification of CKD patients [10], [11], including Decision Tree (DT), Artificial Neural Network (ANN), Random Forest (RF), Logistic Regression (LR), and Extreme Gradient Boosting (XGBoost). The dataset was divided into an 80% training set and a 20% testing set. Hyperparameter optimization was conducted using the GridSearchCV utility in Python, employing layered cross-validation via the Scikit-learn library. All models were trained and evaluated on a system equipped with a GeForce NVIDIA MX350 graphics card (2 GB) and an Intel Core i5-1135G7 (11th Gen) processor. Key performance metrics, including Accuracy, Precision, Sensitivity, F1 Score, Matthews Correlation Coefficient (MCC), and the Area Under the Receiver Operating Characteristic Curve (ROC AUC), were computed to evaluate the performance of the classification models. A 5-fold crossvalidation technique was employed for all models to ensure the reliability of the results. This approach partitions the dataset into five subsets, using four for training and the remaining one for validation. The process is repeated five times, with each subset serving as the validation set once.

This method provides a robust estimate of model performance and minimizes the risk of overfitting. Furthermore, the SHapley Additive exPlanations (SHAP) analysis and the Local Interpretable Model-agnostic Explanations (LIME) algorithm were applied to enhance the global and local interpretability of models, utilizing ECG-derived features for CKD prediction in the LR model, which demonstrated the best performance [12].

3. Results and Discussion

3.1. HF Classification Performance

Table 2 displays the results of the performance of the developed classification models; the models achieved accuracy ranging from 52.26% to 69.35% for distinguishing CHF individuals with lower eGFR values. Among the evaluated models, LR (i.e., liblinear Solver, Penalty 11, and C 10.0) achieved the highest accuracy (69.35%), outperforming the other models in Precision (69.13%), Sensitivity (70.71%), and F1 Score (69.65%). It also attained the highest MCC (0.387) and a notable ROC AUC (0.752), indicating a strong discriminative ability in classifying CKD patients. RF closely followed with an accuracy of 68.84%, demonstrating a balanced trade-off between Precision (69.07%), Sensitivity (67.68%), and an MCC of 0.378. Its ROC AUC score of 0.688 suggests reliable classification capability. While XGBoost though slightly lower in performance, still provided reasonable classification capability. Decision Tree and ANN exhibited lower predictive strength.

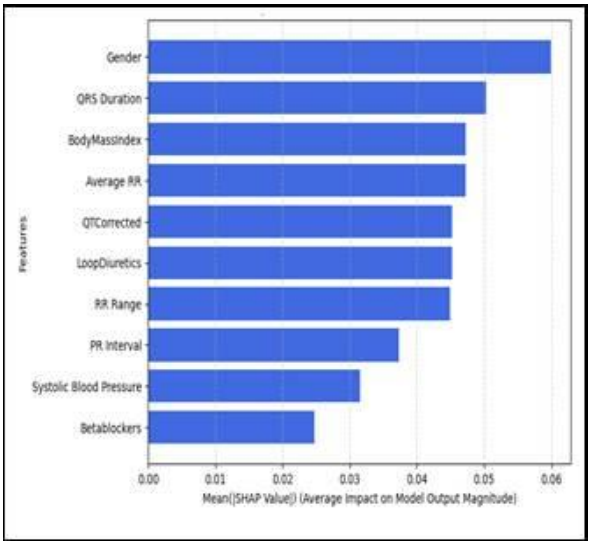

Figure 1. Ranking of top 10 feature importance indicated by SHAP algorithm for predicting eGFR classes in CHF patients.

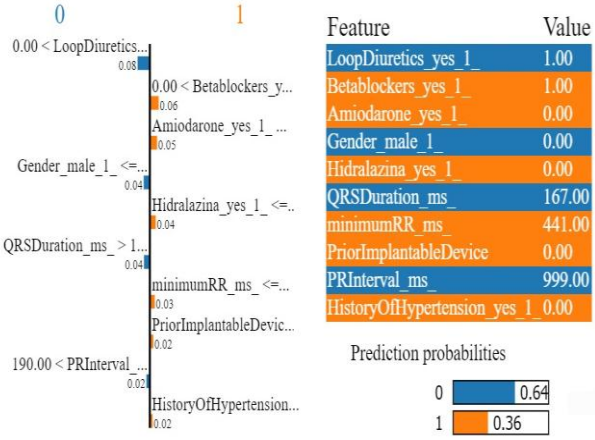

Figure 2. A visualization of LIME model scores for patient ID 3 using the LR model.

3.2. Explainable ML

Figure 1 shows the top 10 SHAP values; the results revealed that gender had the most significant impact, followed by QRS duration, BMI, RR range, and corrected (QTc) interval. Additionally, medication-related variables, such as Loop Diuretics and Beta Blockers, also played a role, emphasizing the influence of cardiovascular treatments on CKD classification. Furthermore, LIME was utilized to analyze and interpret predictions at the individual level. This approach was demonstrated by examining a specific patient case, as illustrated in Figure 2. The figure highlights the key 10 features that influenced

the classification of individuals into eGFR > 66 mL/min/1.73m^2 (orange) or eGFR < 66 mL/min/1.73m^2 (blue).

The respective contributions of these features and their specific values are detailed in the figure, providing a clear visualization of their impact on the model's decision-making process. The prediction outcome for patient ID 3 confidently suggests that this specific patient is having eGFR < 66 mL/min/1.73m^2, with a prediction confidence of 64%. QRS duration of 167 ms is noted, which may indicate prolonged ventricular depolarization, potentially impacting the classification. The minimum RRI 441ms suggests variability in HR, which could be a contributing factor. A PR interval of 999 ms is significantly prolonged, indicating potential atrioventricular conduction delays, which may also influence the prediction of lower eGFR. The results of this study suggest that cardiac function assessed by means of ECG is related to kidney function. This is consistent with earlier studies [4, 6, 13]. In this study, among the ECG features, QRS duration, average RRI and QTc interval were found to have the most significant correlation with eGFR (Figure 1). We expect these correlations are related to electrolyte regulation performed by the kidney and reflected in ECG [4]. Our findings demonstrate that noninvasive assessment of ECG is potentially helpful to monitor kidney function. The main limitation of this study is the absence of control subjects. Here, we included CHF patients only. Therefore, it is unknown how the model's performance will change in the ECG of control subjects.

Acknowledgment

The work in this paper has been supported by CIRA grant 8474000408 (CIRA 2021-051), 8474000376 (KU-KAU project) and 8474000132 (HEIC) CIRA grant (2019-023) awarded to Ahsan Khandoker by Khalifa University Abu Dhabi.

References

[1] S. Altamura, D. Pietropaoli, F. Lombardi, R. Pinto and C. Ferri, "An Overview of Chronic Kidney Disease Pathophysiology: The Impact of Gut Dysbiosis and Oral Disease," Biomedicines., vol. 11, no. 11, 2023.

[2] T. Chen, D. Knicely and M. Grams, "Chronic Kidney Disease Diagnosis and Management," JAMA, vol. 322, no. 13, pp. 1294-1304, 2020.

[3] L. Inker and S. Titan, "Measurement and Estimation of GFR for Use in Clinical Practice: Core Curriculum 2021," AJKD, vol. 78, no. 5, pp. 736-749, 2021.

[4] L. Holmstorm et al. "Deep Learning-Based Electrocardiographic Screening for Chronic Kidney Disease," Commun Med (Lond), vol. 3, no. 1, 2023.

[5] J. Kwon, et al, "Artificial Intelligence Assessment for Early Detection and Prediction of Renal Impairment using Electrocardiography," Int Urol Nephrol. , vol. 54, no. 10, p. 2733–2744, 2022.

[6] N. Widatalla, P. Buenaventura, S. Ghosh and A. Khandoker, "Investigating the Association between Resting Heart Rate and Estimated Glomerular Filtration Rate : A UK BioBank Study," in 13th ESGCO Meeting International conference of the European Study Group on Cardiovascular Oscillations, Zaragoza, Spain, 2024.

[7] A. Martin-Yebra, J. Martinez and P. Laguna, "MUSIC (Sudden Cardiac Death in Chronic Heart Failure)," PhysioNet, 2024.

[8] A. Mart´ın-Yebra, A. Luna, R. Vazquez, P. Caminal, P. Laguna and J. Mart´ınez, "The MUSIC Database: Sudden Cardiac Death in Heart Failure Patients," in Computing in Cardiology, Karlsruhe, 2024.

[9] A. Lesley and e. al., "New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race," N Engl J Med, vol. 385, no. 19, pp. 1737-1749, 2021.

[10] S. M. Al Younis et al., "Prediction of heart failure patients with distinct left ventricular ejection fraction levels using circadian ECG features and machine learning," PLoS One, vol. 19, no. 5, p. e0302639, 2024.

[11] S. M. Al Younis et al., "Investigating automated regression models for estimating left ventricular ejection fraction levels in heart failure patients using circadian ECG features," PLoS One, vol.18 no. 12, 2023.

[12] C. Hu et al., "Application of Interpretable Machine Learning for Early Prediction of Prognosis in Acute Kidney Injury," Computational and Structural Biotechnology Journal, vol. 20, pp. 2861–2870, 2022. DOI: 10.1016/j.csbj.2022.06.003.

[13] Widatalla, N., Younis, S.A. and Khandoker, A. "Heart rate transition patterns reveal autonomic dysfunction in heart failure with renal function decline: a symbolic and Markov model approach". BioData Mining, 45 (2025). https://doi-org.khalifa.idm.oclc.org/10.1186/s13040-025-00460-x

Sona Al Younis, sona.alyounis@ku.ac.ae
Healthcare Engineering Innovation Group, Department of Biomedical Engineering and Biotechnology, Khalifa University of Science and Technology, P.O. Box 127788 Abu Dhabi, UAE