DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023 ***Int. J. Exp. Res. Rev.*, Vol. 45: 288-300 (2024)**

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

ISSN 2455-4855

Development of a Regression Model for Prediction of Chronic Kidney Disease Risk

Sonal Saini*, Ajay Shanker Singh and Alok Katiyar

Check for updates

School of Computer Science and Engineering, Galgotias University, Greater Noida, India

E-mail/Orcid Id:

SS, iamsonalsaini@gmail.com, https://orcid.org/0009-0007-9575-4137; *ASS,* ajay.shankersingh@galgotiasuniversity.edu.in, https://orcid.org/0000-0001-8963-6141; *AK,* alok.katiyar@galgotiasuniversity.edu.in, https://orcid.org/0000-0002-1645-6585

Article History:
Received: 05th Aug., 2024
Accepted: 25th Nov., 2024
Published: 30th Nov., 2024

Keywords:
Kidney disease, regression modeling, risk prediction, data pre-processing, mathematical modeling, optimization.

How to cite this Article:
Sonal Saini, Ajay Shanker Singh and Alok Katiyar (2024). Development of a Regression Model for Prediction of Chronic Kidney Disease Risk. *International Journal of Experimental Research and Review*, *45*, 288-300.
DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

Abstract: In recent years, chronic kidney disease (CKD) has been widespread in public health. Therefore, the early prediction of these diseases can save many lives. Keeping this fact in mind, this study presents a new way to predict CKD using regression modeling, aiming to improve early detection and save lives. For this purpose, the first authors collected the data of 104 patients, then re-arranged them in ten different parameters and calculated their scores. Thereafter, a composite CATH score is calculated as an output variable. Then, a suitable regression model will be identified based on various parameters such as R-squared, Adjusted R-squared, and PRESS values. Thereafter, to identify the significance of the selected model, the authors performed an Analysis of Variances (ANOVA) at a confidence interval of <0.05. Results revealed that the developed model has a higher degree of fitness and is suitable for prediction purposes. Finally, the authors performed parameter analysis to identify the effects of various parameters on CKD.

Introduction

In recent decades, Chronic Kidney Disease (CKD) has been a vital health concern for human beings. It's estimated that nearly 10% of the global population is affected by this disease, posing a remarkable burden on the global healthcare systems (Tangri et al., 2011; Echouffo-Tcheugui et al., 2012; Xiao et al., 2019 and Nelson et al., 2019). These diseases are characterized by a gradual loss of kidney function with time and can lead to End-Stage Renal Disease (ESRD) (Chien et al., 2010; Yang et al., 2023; Wang et al., 2024). The early detection and risk prediction of these diseases are very critical issues for preventing and improving patient outcomes (Bai et al., 2022; Dritsas & Trigka, 2022; Singh et al., 2022; Yang et al., 2023; Hassan et al., 2023; Ozcan and Peker, 2023).

The pathophysiology of CKD is complex and multifactorial, involving genetic, environmental, and metabolic factors (Hassan et al., 2023; Ozcan and Peker, 2023). Traditional risk factors such as diabetes, hypertension, and cardiovascular disease are well-established contributors to CKD development (Jiang et al., 2020). However, there is a growing recognition of the importance of other factors, including inflammation, oxidative stress and lifestyle behaviors (Matsushita et al., 2020; Schena et al., 2021). These insights underscore the need for a comprehensive risk assessment model that encompasses a wide range of variables.

In response to the limitations of existing prediction models, which often focus on a narrow set of clinical indicators, recent advances in data analytics have paved the way for more sophisticated approaches. Regression modeling has emerged as a powerful tool for identifying at-risk individuals by analyzing complex datasets with numerous potential predictors (Krishnamurthy et al., 2021; Sawhney et al., 2023; Liu et al., 2024). This approach can enhance the accuracy of risk prediction, informing targeted interventions and personalized management strategies (Hosseini Sarkhosh et al., 2023).

***Corresponding Author:** iamsonalsaini@gmail.com

Despite the potential of regression models, their adoption in clinical practice has been impeded by challenges in handling high-dimensional data and integrating diverse types of clinical information. Data pre-processing, including feature selection and dimensionality reduction, is crucial in constructing effective predictive models (Khan et al., 2023). By distilling large datasets into meaningful composite scores, researchers can improve the interpretability of the models and facilitate their application in real-world settings (Chen et al., 2023; Ningthoujam et al., 2024).

The present study seeks to contribute to this burgeoning field by proposing a novel regression modeling framework for CKD risk prediction. Our research leverages a comprehensive cardiovascular health dataset, applying data pre-processing techniques to refine the analysis and enhance the predictive power of the models. We systematically evaluate linear, 2FI, quadratic, and cubic regression models to determine their efficacy in predicting CKD risk, focusing on their respective R-squared, Adjusted R-squared and PRESS statistics as measures of performance.

The predictive models developed in this study are grounded in a solid theoretical framework that considers the biological and clinical underpinnings of CKD. Our approach builds upon previous research that has identified potential biomarkers and risk factors for kidney disease. By incorporating these factors into our regression models, we aim to provide a nuanced understanding of CKD risk that reflects the complexity of the disease.

In addition to their theoretical contributions, our findings have practical implications for public health and clinical practice. By identifying individuals at high risk for CKD, healthcare providers can prioritize interventions and allocate resources more effectively. Moreover, our research offers insights into the design of future epidemiological studies and clinical trials aimed at preventing CKD and its complications.

Thus, the present study addresses a significant gap in the literature by developing a robust regression modeling framework for CKD risk prediction. Our comprehensive analysis, supported by a data-driven approach, holds promise for enhancing the early detection of CKD and optimizing patient care.

Data Collection & Data Pre-Processing

Initially, we collected real datasets of 104 chronic kidney disease patients from a renowned hospital of Delhi-NCR, India, in 2023. Then, the data is pre-processed for further analysis. Initially, data was collected using various parameters. Then, we re-classified data based on different patient attributes like demographic, physiological, and clinical parameters. These number of variables are reduced due to a decrease in the degree of data complexity.

Data Reduction Strategy

Grouping and Formula Application

To acquire the composite score of CATH value, we grouped the various variables based on their nature by using formulas as shown in Eq. (1)-(6).

Demographic Variables

- Variables Combined: Age, Sex, Weight
- Formula: $$Demographic\ Score = \frac{Age}{100} + \frac{Weight}{200} + Sex \quad (1)$$
- Rationale: This score normalizes age and weight to a 0-1 scale and incorporates sex directly, providing a composite demographic index.

Vital Signs

- Variables Combined: Blood Pressure (BP), Pulse
- Formula: $$Vital\ Signs\ Score = \frac{BP}{300} + \frac{Pulse}{200} \quad (2)$$
- Rationale: Normalization of BP and Pulse allows for a unified score that reflects overall cardiovascular strain.

Respiratory Symptoms

- Variables Combined: Lung Rales, Dyspnea
- Formula: $$Respiratory\ Score = Lung\ Rales + Dyspnea \quad (3)$$
- Rationale: A simple additive model captures the presence of respiratory issues.

Cardiac Symptoms

- Variable Used: St Elevation
- Rationale: Used directly as a binary indicator of significant cardiac events.

Metabolic Measures

- Variables Combined: Fasting Blood Sugar (FBS), Creatinine (CREAT)
- Formula: $$Metabolic\ Score = \frac{FBS}{300} + \frac{CREAT}{20} \quad (4)$$
- Rationale: This score integrates key metabolic parameters relevant to cardiovascular health.

Blood Work

- Variables Combined: Blood Urea Nitrogen (BUN), Hemoglobin (HB), Platelets (PLT)
- Formula: $$Blood\ Work\ Score = \frac{BUN}{100} + \frac{HB}{20} + \frac{PLT}{500} \quad (5)$$
- Rationale: Aggregates critical blood parameters into a single indicative score.

Immune and Inflammatory Markers

- Variables Combined: White Blood Cells (WBC), Reduced Erythropoietin

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

- Formula: $Immune\ Score = \frac{WBC}{20} + Reduced\ Erythropoietin$ (6)
- Rationale: Reflects the inflammatory state and immune activity related to cardiovascular health.

Additional Parameters

- Variables Used: Length, Edema, Angioplasty
- Rationale: These were used directly as they are indicative of physical characteristics and medical interventions.

Implementation

The formulas were applied to the transformation of each group into its respective composite score. These scores were then appended to the dataset, replacing the original variables. This data transformation provides the simplicity of collected data with increased clarity. In this study, the CATH score is selected as the output variable on which the degree of prediction is decided later. After calculating the scores of all ten parameters, a composite

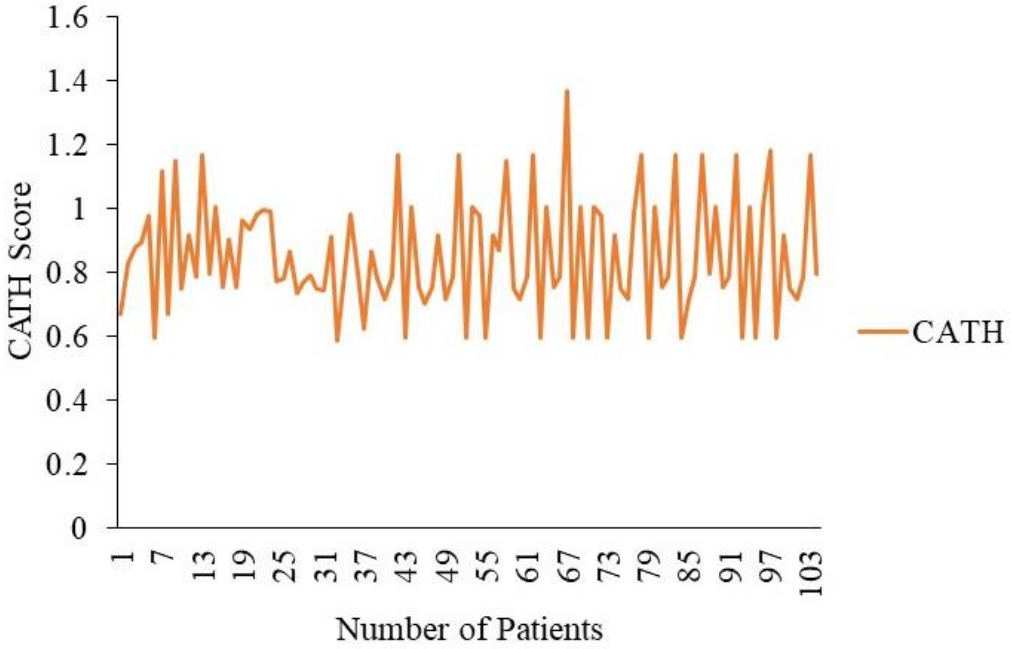

Figure 1. Calculated CATH Score variation for all patients.

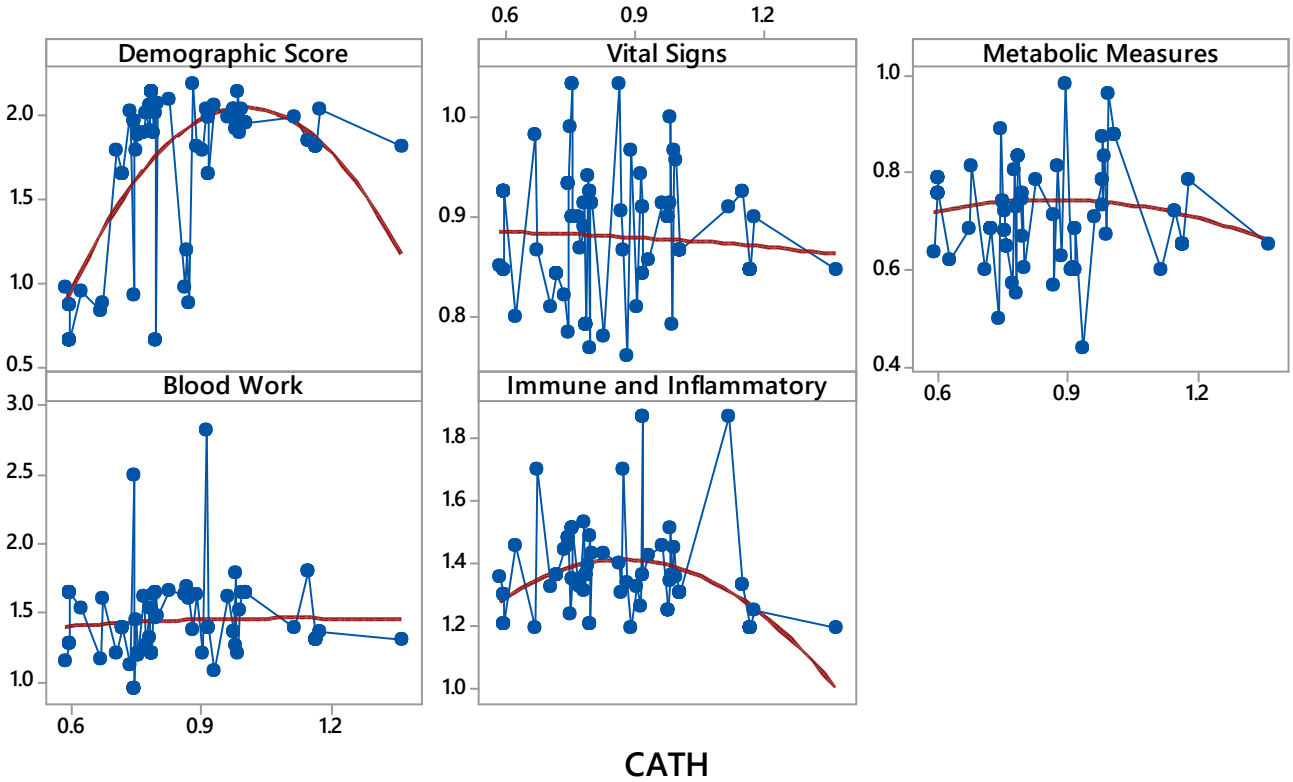

Figure 2. Scatter Plots and trend lines between CATH and various health variable scores.

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

CATH score was calculated. Table 1 consists of the descriptive statistics of all ten variables.

Table 1. Descriptive Statistics of patient data.

Variable	Mean	SE Mean	StDev	CoefVar	Median
Demographic Score	1.6754	0.0498	0.5074	30.28	1.8980
Vital Signs	0.87954	0.00579	0.05907	6.72	0.86667
Metabolic Measures	0.7327	0.0100	0.1022	13.95	0.7407
Blood Work	1.4316	0.0267	0.2719	18.99	1.3820
Immune and Inflammatory	1.3551	0.0148	0.1509	11.13	1.3333
CATH	0.8510	0.0177	0.1803	21.19	0.7919

Then, after providing the weight to each variable, a formula as shown in Eq.7 is used to calculate the CATH score.

$$Cath\ Score = 0.15 \times Demographic\ Score + 0.1 \times Vital\ Signs\ Score + 0.1 \times Respiratory\ Symptoms\ Score + 0.2 \times Cardiac\ Symptoms + 0.1 \times Metabolic\ Measures\ Score + 0.1 \times Blood\ Work\ Score + 0.1 \times Immune\ and\ Inflammatory\ Score + 0.05 \times Length/200 + 0.1 \times Edema + 0.15 \times Angioplasty \quad (7)$$

After, using this Eq. (7), the CATH score for all 104 patients is calculated. Figure 1 shows the variation in the CATH score value of each patient. Then, this calculated CATH score is used to select the most suitable and efficient regression model.

Then, based on a scatter plot as depicted in Figure 2, a relationship is predicted between the CATH score and the various health scores. We clearly observe that the Metabolic Measures, Blood Work, and Immune and Inflammatory scores show a broader dispersion of data points with obvious outliers. It indicates that they have a non-linear relationship with CATH. The outlier values of Blood Work and Immune and Inflammatory scores significantly affect the prediction accuracy. This analysis clearly indicates the need for a robust statistical model.

Mathematical Modelling

The analysis presented in this research paper evaluates the impact of various cardiovascular health indicators through a regression model where the dependent variable, CATH (catheterization necessity), is regressed against predictors such as Demographic Score (DS), Vital Signs (VS), Metabolic Measures (MM), Blood Work (BW), and Immune and Inflammatory (II) scores.

Result and Discussion

Mathematical Model

The ANOVA table 3 in our statistical model delineates the effects of various predictors on the catheterization necessity (CATH) through linear, square, and interaction terms. Linear terms like Vital Signs (VS) significantly directly impact CATH, evidenced by an F-value of 11.12, indicating their strong influence.

Table. 2. Model summary statistics and selection for predicting CATH variable.

Model Type	R-Squared	Adjusted R-Squared	PRESS	Suggested Model
Linear	0.407	0.377	1.985	No
2FI	0.703	0.652	0.996	Yes
Quadratic	0.736	0.682	0.998	Yes
Cubic	0.266	0.229	2.457	No

In the comparative analysis of statistical models as tabulated in Table 2 for predicting the variable CATH, the Linear, Two-Factor Interaction (2FI), Quadratic, and Cubic models present varying levels of performance based on R-Squared, Adjusted R-Squared, and PRESS values. The Linear model, with an R-Squared of 0.407 and an Adjusted R-Squared of 0.377, offers moderate explanatory power, explaining about 40.7% of the variance in CATH. Its PRESS score of 1.985 suggests reasonable predictive accuracy, although it is not recommended due to better-performing alternatives. The 2FI model significantly outperforms the Linear model, with an R-Squared of 0.703 and an Adjusted R-Squared of 0.652, reflecting strong explanatory capability and the best predictive accuracy among the models, as evidenced by the lowest PRESS score of 0.996. Consequently, it is highly recommended for use. The Quadratic model also shows excellent performance, with the highest R-Squared of 0.736 and an Adjusted R-Squared of 0.682, alongside a very competitive PRESS score of 0.998, making it another recommended model. In contrast, the Cubic model, despite the complexity it adds, performs poorly with an R-Squared of 0.266 and an Adjusted R-Squared of 0.229, combined with the highest PRESS score of 2.457, indicating its inadequacy in both explanatory and predictive capacities. Hence, it is not recommended. The selection between the 2FI and Quadratic models should consider specific analysis goals and the desired balance between model complexity and interpretability.

Additionally, square terms reveal non-linear dynamics; particularly, the Immune and Inflammatory (II) scores squared (II*II) present a quadratic relationship with CATH, having a high F-value of 53.45. This suggests

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

increasing impact at higher levels of II scores. Two-way interactions, such as between VS and II, with an F-value of 35.70, indicate that the effects of vital signs on CATH are considerably moderated by the immune-inflammatory state, underscoring the complexity of these relationships in predicting healthcare needs.

Table 3. Analysis of Variance (ANOVA).

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	20	2.56067	0.128034	22.56	0.000
Linear	5	0.12248	0.024495	4.32	0.002
DS	1	0.02639	0.026392	4.65	0.034
VS	1	0.06312	0.063124	11.12	0.001
MM	1	0.01074	0.010735	1.89	0.173
BW	1	0.00517	0.005175	0.91	0.342
II	1	0.00636	0.006355	1.12	0.293
Square	5	0.39286	0.078572	13.85	0.000
DS*DS	1	0.00001	0.000010	0.00	0.966
VS*VS	1	0.00962	0.009618	1.69	0.197
MM*MM	1	0.00414	0.004145	0.73	0.395
BW*BW	1	0.00346	0.003465	0.61	0.437
II*II	1	0.30333	0.303331	53.45	0.000
2-Way Interaction	10	0.47118	0.047118	8.30	0.000
Error	83	0.47102	0.005675		
Lack-of-Fit	18	0.22057	0.012254	3.18	0.000
Pure Error	65	0.25045	0.003853		
Total	103	3.03170			
Model Summary					
S	**R-sq**	**R-sq(adj)**	**R-sq(pred)**		
0.0753323	**84.46%**	**80.72%**	**55.95%**		

The high R-squared (84.46%) reflects the model's efficacy, indicating a strong explanatory power. Adjustments for predictors' degrees of freedom result in an adjusted R-squared of 80.72%, ensuring that multiple predictors do not overestimate the model accuracy. The regression equation provided synthesizes these interactions into a predictive formula, offering coefficients for each term that quantify their specific contributions to the prediction of CATH. This mathematical modeling facilitates a nuanced understanding of how multiple factors interplay to affect the likelihood of requiring catheterization, which is crucial for targeted interventions and patient management in cardiovascular care.

The second-order regression models' generalized equation can be written as Eq. 8:

$$y_j = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} b_{ii} x_i^{\,2} + \sum_i \sum_j b_{ij} x_i x_j \quad (8)$$

Where, i=1, 2,.., n and j=1, 2,..,n and n is the total number of input parameters, y_j is the output response and b_0, b_{ii}, b_{ij} are the regression coefficients and x_{ij} denoted the values of different control factors for i^{th} observation and j^{th} level. On the basis of this generalized equation, a second-order polynomial prediction equation is developed for CATH value as mentioned in Eq.(9).

*CATH =17.52+1.465 DS- 7.71 VS+7.06 MM+ 0.82 BW- 18.18 II- 0.0036 DS*DS- 2.77 VS*VS+ 0.831 MM*MM- 0.0539 BW*BW+ 2.254 II*II- 0.977 DS*VS+ 0.307 DS*MM+ 0.0836 DS*BW- 0.530 DS*II+ 0.95 VS*MM- 1.455 VS*BW+ 11.65VS*II+ 0.779 MM*BW+ 2.35 MM *II+ 0.095 BW*II* (9)

Parametric Analysis

Figure 3 is a normal probability plot with corresponding statistical data for five variables. The plot suggests deviations from normality, as evidenced by points straying from the diagonal reference line, particularly in the tails. This indicates potential skewness or kurtosis beyond what is expected for a normal distribution. The accompanying table confirms these findings, with all variables displaying small p-values (<0.005) from the Anderson-Darling normality test, leading to a rejection of the hypothesis that the data are normally distributed. This suggests the need for non-parametric statistical methods or data transformation before analysis, as the assumption of normality is a key prerequisite for many conventional parametric tests.

The Figures (4-12) present in 3D surface plot visualizing the relationship between various health

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

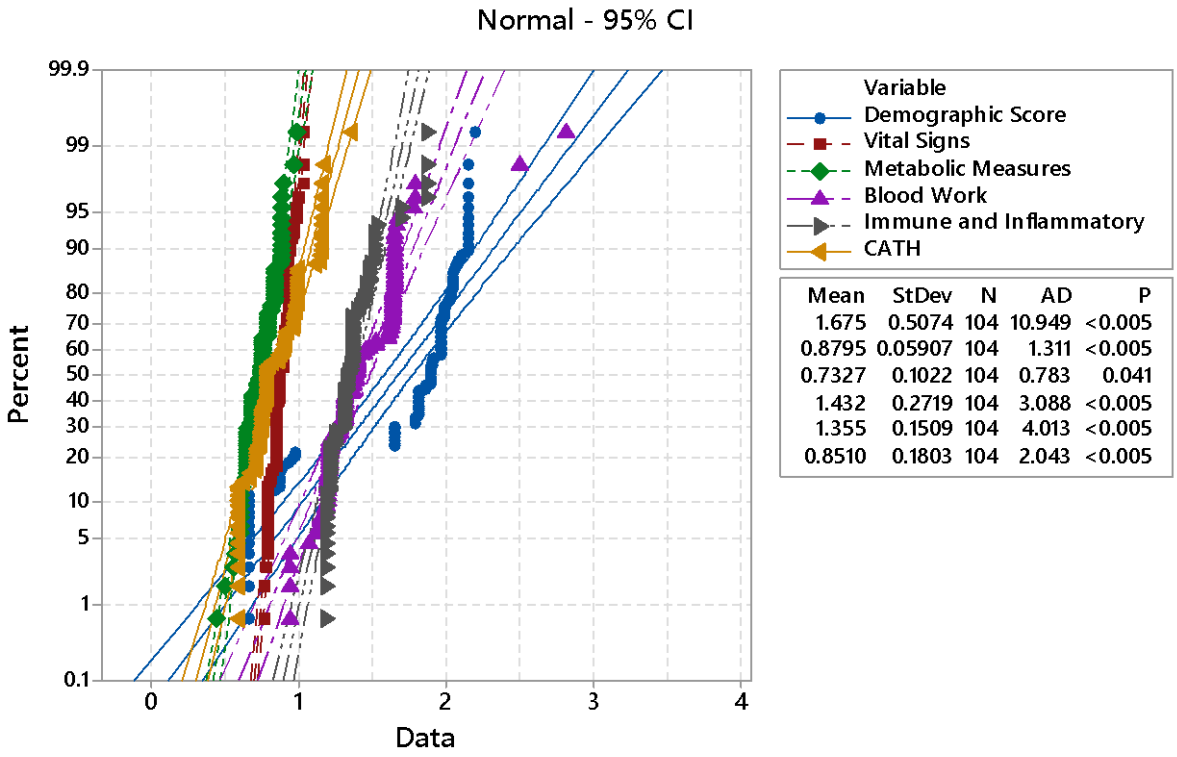

Figure 3. Normality assessment of clinical variables with CATH scores included.

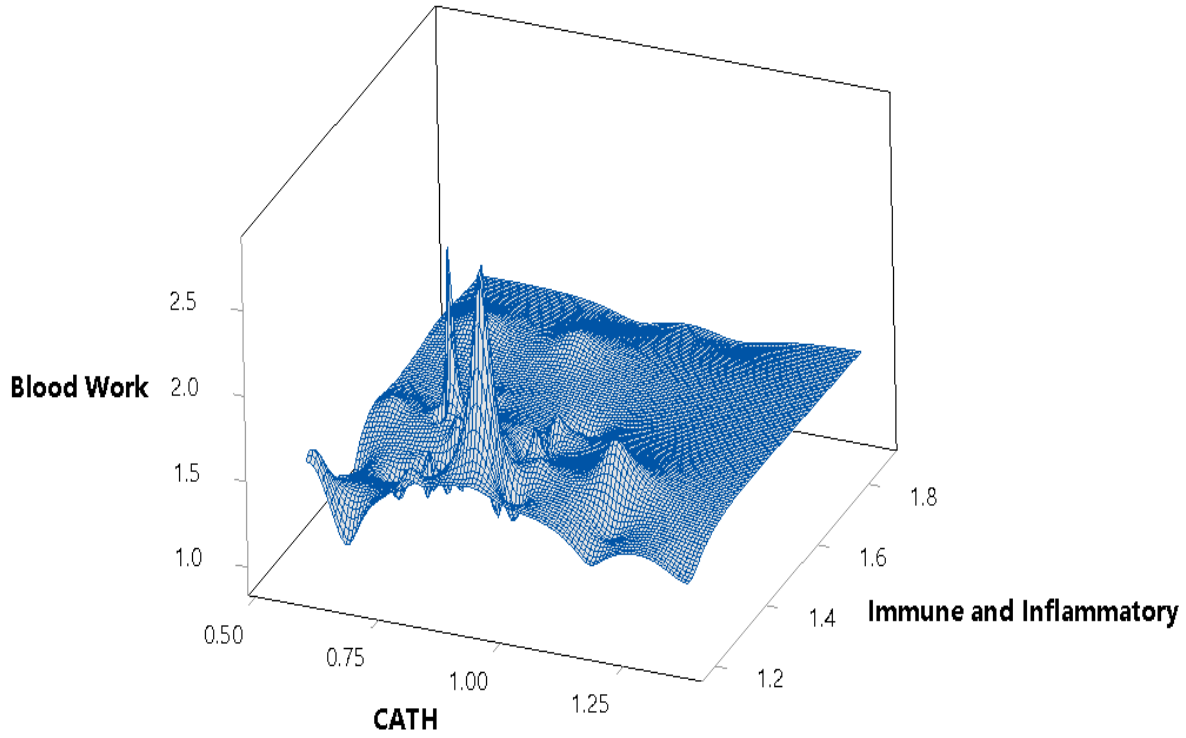

Figure 4. 3-D surface plot for the interactive effect of Blood work & Immune and Inflammatory parameters on CATH score.

parameters and "CATH." The surface peaks and troughs indicate the variability in CATH influencing the other two parameters. There are sharp peaks, suggesting possible outliers or anomalous values affecting the "CATH" variable. These could indicate data entry errors, measurement errors, or true variability needing further investigation. The nature of the data and these spikes warrant careful consideration in the context of Modeling, as they may unduly influence statistical results. It's also crucial to consider data pre-processing to address these potential anomalies for robust model performance.

In Fig. 4, 3-D surface plot shows the impact of Blood Work and Immune and Inflammatory parameters on CATH score. As shown in the plot, the strong interactive effect of both parameters results in a higher CATH score or greater possibility of CKD.

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

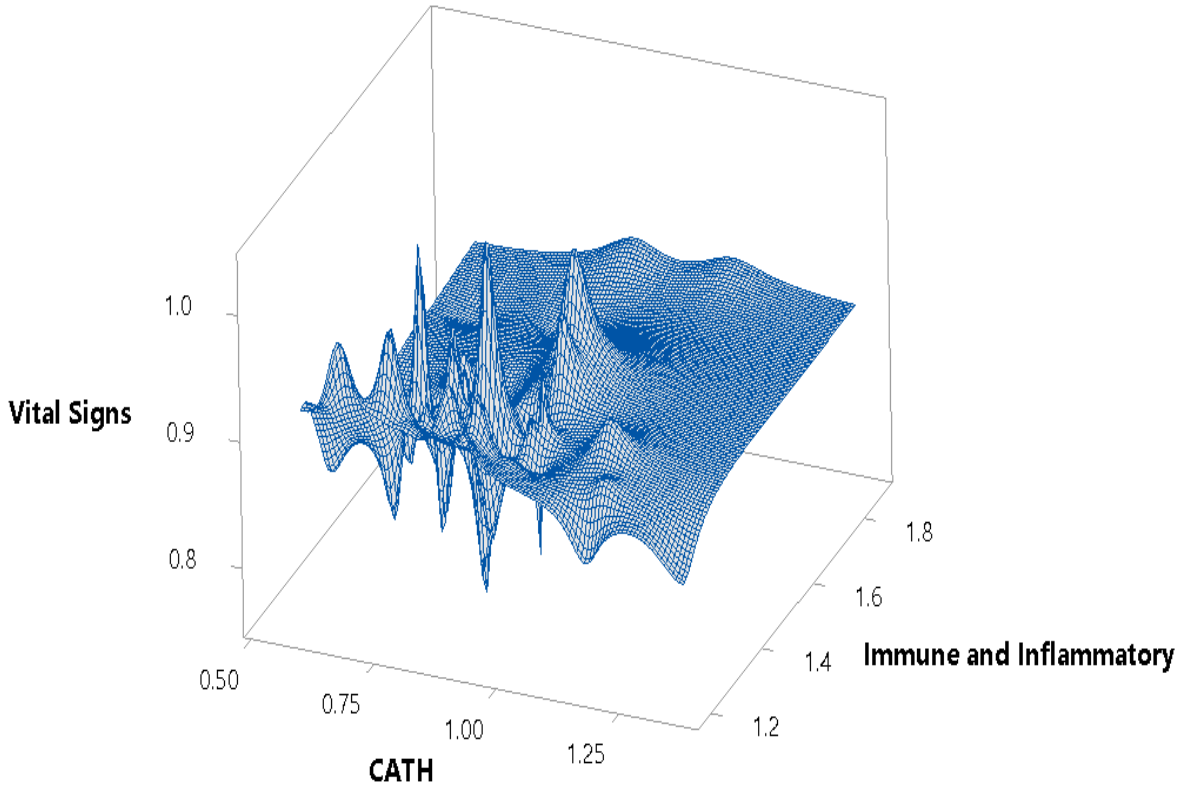

Figure 5. 3-D surface plot for interactive effect of Vital Signs and Immune and Inflammatory on CATH score.

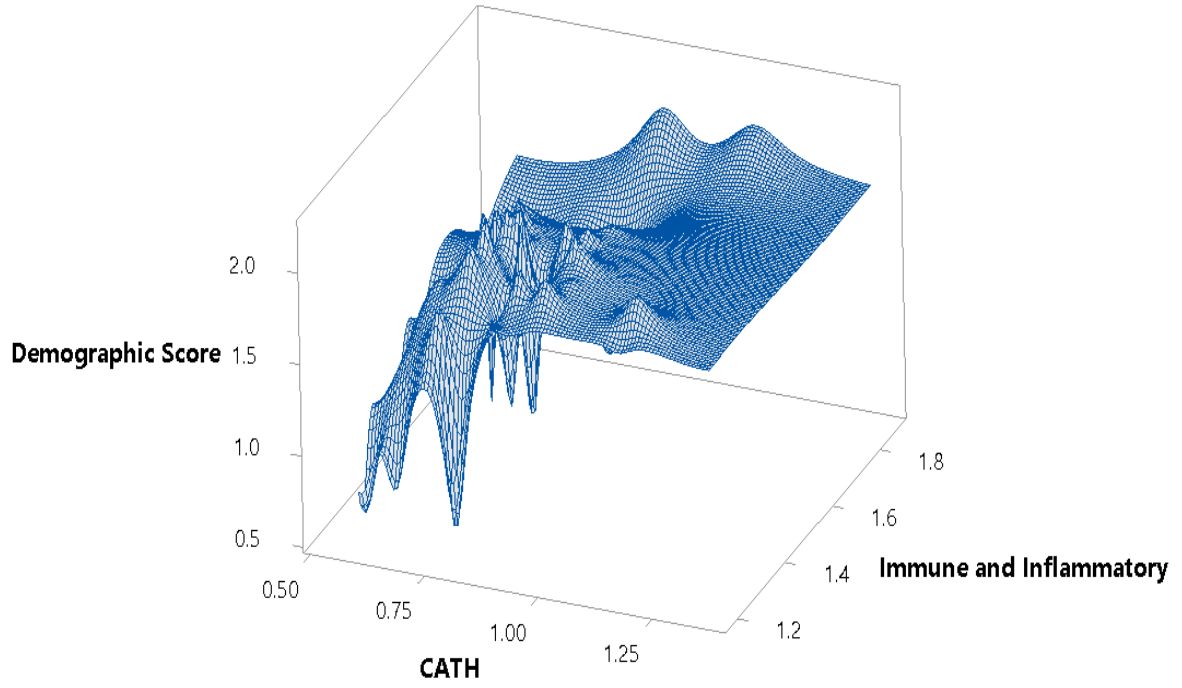

Figure 6. 3-D surface plot for the interactive effect of Demographic Score and Immune and Inflammatory on CATH score.

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

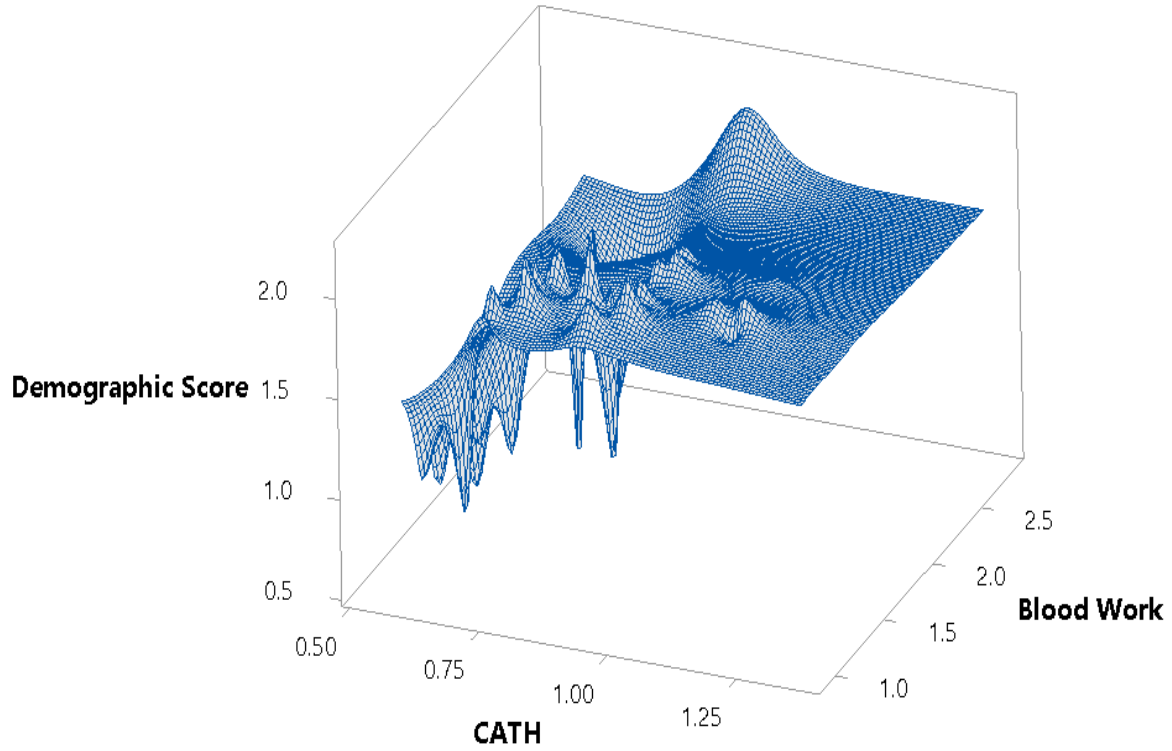

Figure 7. 3-D surface plot for interactive effect of Demographic Score and Blood work on CATH score.

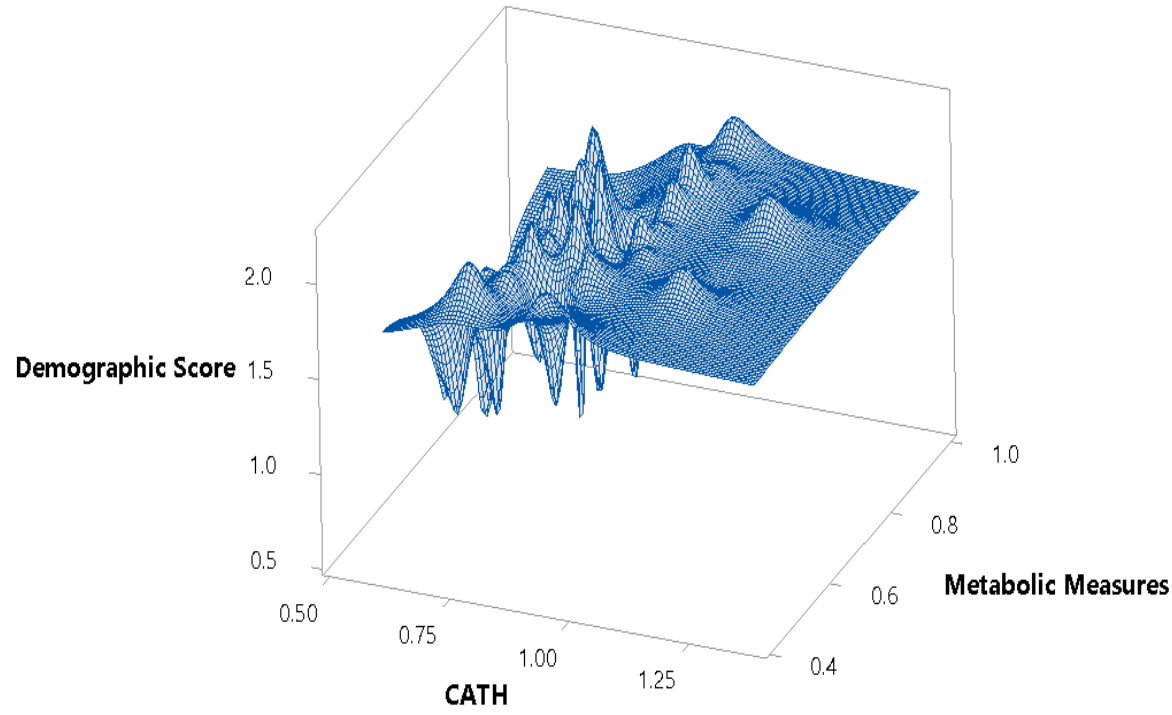

Figure 8. 3-D surface plot for interactive effect of Demographic Score and Metabolic Measures on CATH score.

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

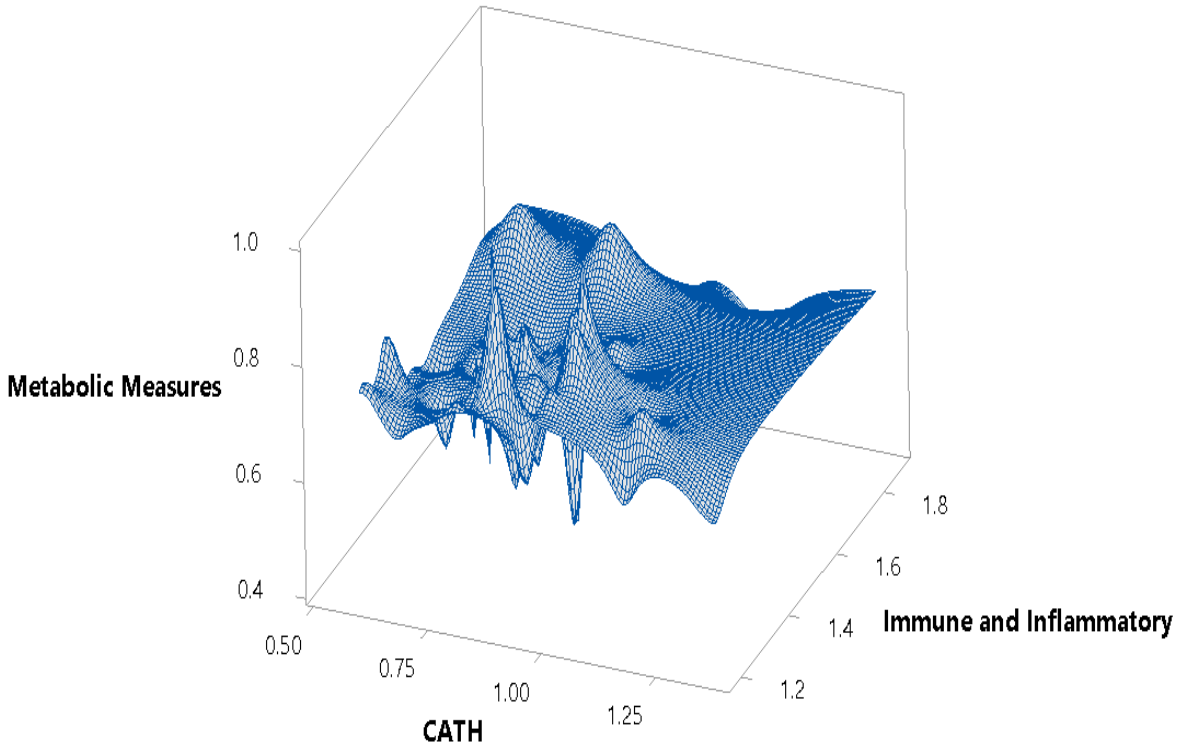

Figure 9. 3-D surface plot for interactive effect of Metabolic Measures and Immune and Inflammatory CATH score.

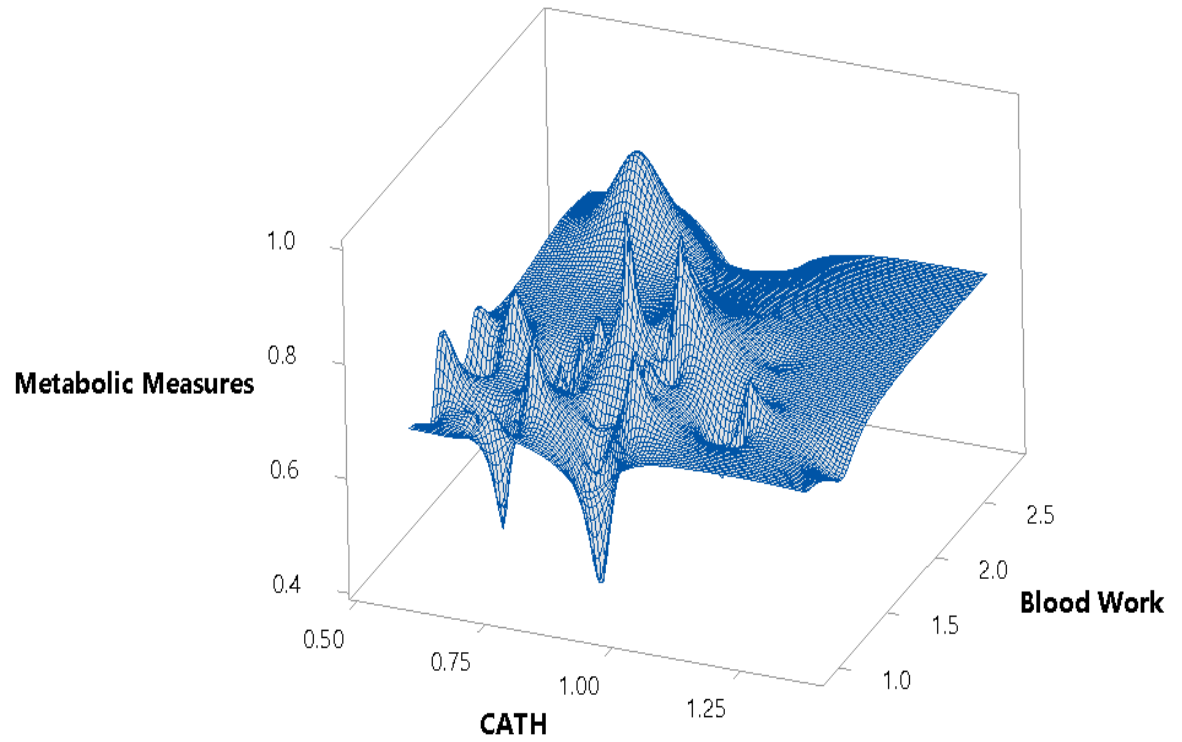

Figure 10. 3-D surface plot for the interactive effect of Metabolic Measures and Blood Work on CATH score.

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

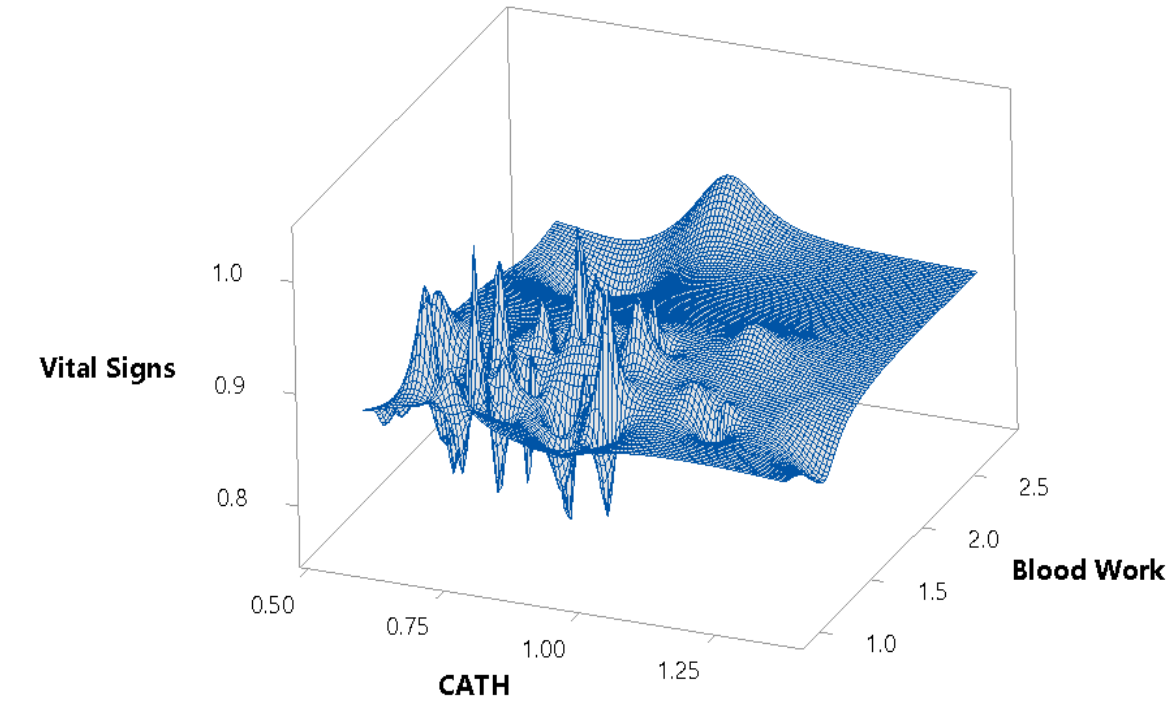

Figure 11. 3-D surface plot for interactive effect of Vital Signs and Blood Work on CATH score.

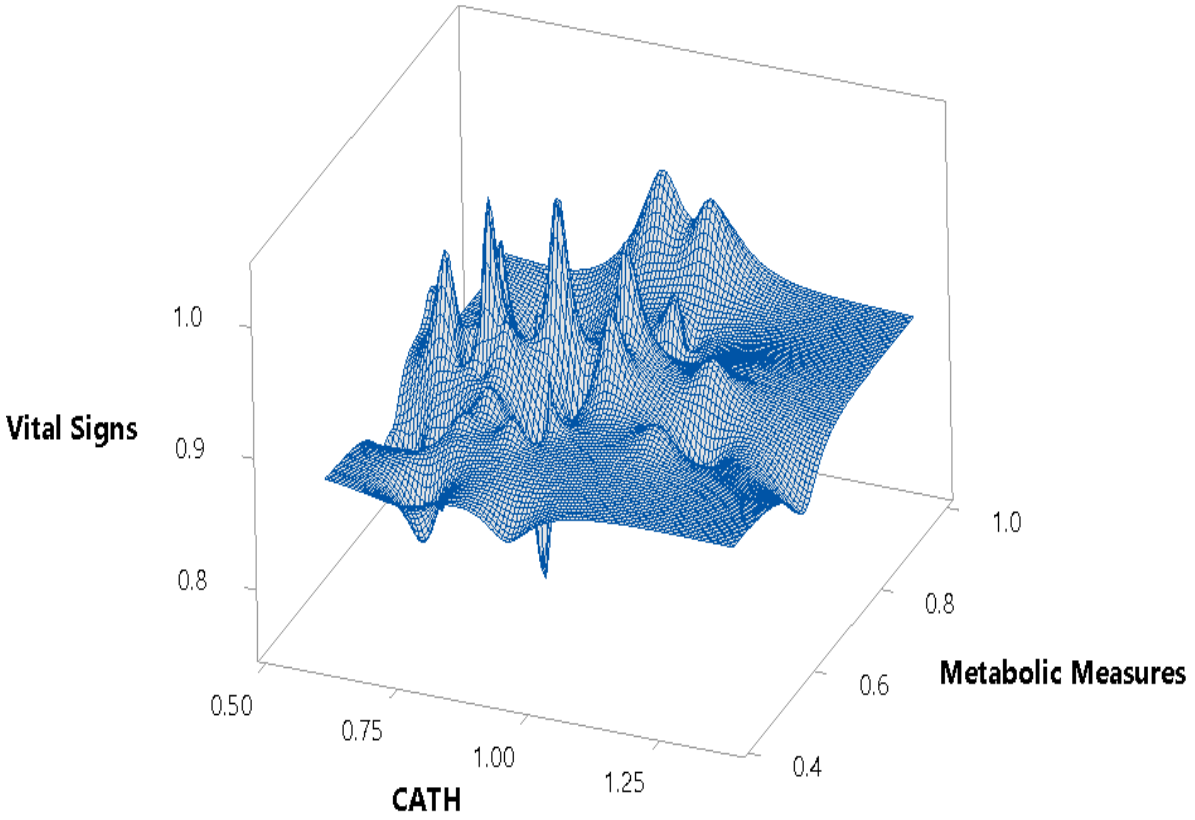

Figure 12. 3-D surface plot for the interactive effect of Vital Signs and Metabolic Measures on CATH score.

DOI: https://doi.org/10.52756/ijerr.2024.v45spl.023

On the other side, Figure 5 depicts the interactive effects of Vital Signs and Immune and Inflammatory parameters on CATH scores. This plot reveals that at certain levels of immune and inflammation, patient vital signs have an exacerbated impact on CATH score. However, Figure 6 clearly shows that demographic factors with the patient's body immune responses could have complex and combined impacts on Kidney's health. Moreover, it shows the multifaceted nature of CKD depends on patient's demographic condition.

Figure 7 shows a 3-D surface plot demonstrating how the Demographic Score and Blood Work interact to impact CATH scores, a proxy for cardiovascular intervention necessity. The peaks in the plot may indicate demographic and blood work profiles that are associated with an increased likelihood of requiring such interventions. Conversely, lower regions suggest less intervention might be needed.

Figure 8 depicts a 3-D surface plot demonstrating the interaction between Demographic Score, Metabolic Measures, and CATH scores. The plot may suggest how demographic variables combined with metabolic health indicators contribute to cardiovascular risk assessments. Peaks might point to specific demographics with metabolic profiles indicative of higher cardiovascular intervention needs, while the troughs could indicate lower-risk demographics.

Figure 9 illustrate a 3-D surface plot that details how Metabolic Measures and Immune and Inflammatory markers impact CATH scores. This plot can indicate the combined effect of a patient's metabolic status and immune system activity on their cardiovascular health assessment. Peaks on the surface could highlight areas where metabolic and immune-inflammation scores correlate with increased cardiovascular risk, whereas lower areas might suggest a reduced need for intervention.

Figure 10 depicts the combined impact of Metabolic Measures and blood Work on CATH score. As represented by the peaks of this graph, it is clearly indicated that the association of metabolic abnormalities with certain blood work may increase the possibility of CKD. Further, Figure 11 reveals that Vital Signs and Blood Work coincide with increased CATH scores, indicating an enhanced need for cardiovascular interventions. Lastly, Figure 12 suggests that the patient's physiological health and metabolic status could have a nuanced impact on their cardiovascular assessment.

Conclusion

In the present research article, the authors tried to develop a regression model to predict kidney disease accurately. First, the authors identified a suitable type of regression model. Then, the quadratic model is selected based on its higher R-Squared value of 0.736 over the linear, 2FI, and cubic models. Thereafter, ANOVA analysis has been performed. During ANOVA, it was observed that it provided satisfactory values of R-sq at 84.46% and R-sq(adj) at 80.72%. This shows that the developed model has higher fitness and adequacy. Then, a second-order polynomial quadratic model is developed. Later, the authors performed parametric analysis to identify the impact of kidney disease parameters on CATH value. The authors believe that the developed model could be used to predict CKD in the future.

Acknowledgement

The authors are highly thankful to the management of Galgotias University, Greater Noida, India for providing fruitful support to complete this work.

Conflict of Interest

The authors declare no conflict of interest.

References

Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific reports, 12*(1), 8377. https://doi.org/10.1038/s41598-022-12316-z

Chen, F., Kantagowit, P., Nopsopon, T., Chuklin, A., & Pongpirul, K. (2023). Prediction and diagnosis of chronic kidney disease development and progression using machine-learning: Protocol for a systematic review and meta-analysis of reporting standards and model performance. *Plos One*, *18*(2), e0278729. https://doi.org/10.1371/journal.pone.0278729

Chien, K.L., Lin, H.J., Lee, B.C., Hsu, H.C., Lee, Y.T., & Chen, M.F. (2010). A prediction model for the risk of incident chronic kidney disease. *The American Journal of Medicine*, *123*(9), 836-846. https://doi.org/10.1016/j.amjmed.2010.05.010

Dritsas, E., & Trigka, M. (2022). Machine learning techniques for chronic kidney disease risk prediction. *Big Data and Cognitive Computing, 6*(3), 98. https://doi.org/10.3390/bdcc6030098

Echouffo-Tcheugui, J.B., & Kengne, A.P. (2012). Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Medicine*, *9*(11), e1001344. https://doi.org/10.1371/journal.pmed.1001344

Hassan, M.N., Mahmud, M.S., Nipa, K.F., & Kamrujjaman, M. (2023). Mathematical Modeling and Covid-19 Forecast in Texas, USA: a prediction model analysis and the probability of disease outbreak. *Disaster Medicine and Public Health Preparedness*, *17*, e19. https://doi.org/10.1017/dmp.2021.151

Hosseini Sarkhosh, S.M., Hemmatabadi, M., & Esteghamati, A. (2023). Development and validation of a risk score for diabetic kidney disease prediction in type 2 diabetes patients: a machine learning approach. *Journal of Endocrinological Investigation*, *46*(2), 415-423. https://doi.org/10.1007/s40618-022-01919-y

Jiang, W., Wang, J., Shen, X., Lu, W., Wang, Y., Li, W., Gao, Z., Xu, J., Li, X., Liu, R., & Zheng, M. (2020). Establishment and validation of a risk prediction model for early diabetic kidney disease based on a systematic review and meta-analysis of 20 cohorts. *Diabetes Care*, *43*(4), 925-933. https://doi.org/10.2337/dc19-1897

Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A novel study on machine learning algorithm-based cardiovascular disease prediction. *Health & Social Care in the Community*, *2023*(1), 1406060. https://doi.org/10.1155/2023/1406060

Krishnamurthy, S., Ks, K., Dovgan, E., Luštrek, M., Gradišek Piletič, B., Srinivasan, K., Li, Y.C., Gradišek, A., & Syed-Abdul, S. (2021). Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan. *In Healthcare*, *9*(5), 546. https://doi.org/10.3390/healthcare9050546

Liu, P., Sawhney, S., Heide-Jørgensen, U., Quinn, R.R., Jensen, S.K., Mclean, A., Christiansen, C.F., Gerds, T.A., & Ravani, P. (2024). Predicting the risks of kidney failure and death in adults with moderate to severe chronic kidney disease: multinational, longitudinal, population based, cohort study. *BMJ*, *385*. https://doi.org/10.1136/bmj-2023-078063

Matsushita, K., Jassal, S.K., Sang, Y., Ballew, S.H., Grams, M.E., Surapaneni, A., Arnlov, J., Bansal, N., Bozic, M., Brenner, H., & Brunskill, N.J. (2020). Incorporating kidney disease measures into cardiovascular risk prediction: development and validation in 9 million adults from 72 datasets. *E-Clinical Medicine*, *27*. https://doi.org/10.1016/j.eclinm.2020.100552

Nelson, R.G., Grams, M.E., Ballew, S.H., Sang, Y., Azizi, F., Chadban, S.J., Chaker, L., Dunning, S.C., Fox, C., Hirakawa, Y., & Iseki, K. (2019). Development of risk prediction equations for incident chronic kidney disease. *JAMA*, *322*(21), pp.2104-2114. https://doi.org/10.1001/jama.2019.17379

Ningthoujam, S.S., Nath, R., Sarker, S.D., Nahar, L., Nath, D., & Talukdar, A.D. (2024). Prediction of medicinal properties using mathematical models and computation, and selection of plant materials. *In Computational Phytochemistry*, pp. 91-123. https://doi.org/10.1016/B978-0-443-16102-5.00011-0

Ozcan, M. and Peker, S., 2023. A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, *3*, p.100130. https://doi.org/10.1016/j.health.2022.100130

Schena, F.P., Anelli, V.W., Trotta, J., Di Noia, T., Manno, C., Tripepi, G., D'Arrigo, G., Chesnaye, N.C., Russo, M.L., Stangou, M., & Papagianni, A. (2021). Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy. *Kidney International*, *99*(5), 1179-1188. https://doi.org/10.1016/j.kint.2020.07.046

Sawhney, R., Malik, A., Sharma, S., & Narayan, V. (2023). A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. *Decision Analytics Journal*, *6*, 100169. https://doi.org/10.1016/j.dajour.2023.100169

Singh, V., Asari, V.K., & Rajasekaran, R. (2022). A deep neural network for early detection and prediction of chronic kidney disease. *Diagnostics*, *12*(1), 116. https://doi.org/10.3390/diagnostics12010116

Tangri, N., Stevens, L.A., Griffith, J., Tighiouart, H., Djurdjev, O., Naimark, D., Levin, A., & Levey, A.S. (2011). A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*, *305*(15), 1553-1559. https://doi.org/10.1001/jama.2011.451

Wang, S., Chen, Y., Cui, Z., Lin, L., & Zong, Y. (2024). Diabetes Risk Analysis Based on Machine Learning LASSO Regression Model. *Journal of Theory and Practice of Engineering Science*, *4*(01), 58-64. https://doi.org/10.53469/jtpes.2024.04(01).08

Xiao, J., Ding, R., Xu, X., Guan, H., Feng, X., Sun, T., Zhu, S., & Ye, Z. (2019). Comparison and development of machine learning tools in the prediction of chronic kidney disease progression.

Journal of Translational Medicine, *17*, 1-13. https://doi.org/10.1186/s12967-019-1860-0

Yang, H., Lin, X., Li, J., Zhai, Y., & Wu, J. (2023). A review of mathematical models of COVID-19 transmission. *Contemporary Mathematics*, pp.75-98. https://doi.org/10.37256/cm.4120232080

How to cite this Article:

Sonal Saini, Ajay Shanker Singh and Alok Katiyar (2024). Development of a Regression Model for Prediction of Chronic Kidney Disease Risk. *International Journal of Experimental Research and Review*, *44*, 288-300.

DOI : https://doi.org/10.52756/ijerr.2024.v45spl.023