Innovations in Public Health Services: Leveraging Data Science for Kidney Health Assessment

Aakansha Soy[1], Sutar Manisha Balkrishna[2]

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India*

KEYWORDS

Kidney Health Assessment, Disease, Data Science, Public Health Services, Recursive feature Elimination (RFE), Differential Evolution optimization with Upgraded logistic regression (DEO-ULR).

ABSTRACT

Improvements in Public Health Services use advanced analytical methods to address kidney-related illnesses early through extensive data research and to enhance public health outcomes. The difficulty would be an extreme dependence on data, which could distort physical condition assessment by ignore kidney patient aspect like existence and socioeconomic conditions. To overcome this problem, use the machine learning (ML) approach in the proposed method. Differential Evolution optimization with Upgraded logistic regression (DEO-ULR) utilizedfor the kidney health evaluation. A diagnosis of CKD has comprehensive health information dataset available in open source Kaggle website. The gathered data is preprocessed using Min-Max normalization and employed feature selectedutilizing Recursive Feature Elimination (RFE). The suggested method is also compared the other traditional algorithms, and this study is experiment with in the Python platform. The findings show the suggested technique achieve enhanced performance in accuracy, precision, F1-Score and recall. The study demonstrates the probable of data science in improvingpublic health for kidney health assessment through advanced analytics, resulting in more precise diagnoses, efficient treatment strategies, and earlier detection for patients.

1. Introduction

Public health developments are always altering the way that handle healthcare problems, and data science is becoming a major player in altering how understand and handle different public health problems [1]. The assessment of kidney health is one such critical area of concentration, where the utilization of advanced analytical techniques offers significant improvements. Chronic kidney disease (CKD) affects a lot of people globally and has a considerableunenthusiasticcollision on both health and the economy [2]. As such, kidney health continues to be a main public health distress. The established CKD organization has undergone a revolution driven by data science, which has completely proactive, modified healthcare strategies [4]. The creativetechniques for kidney disease research and treatment use data science, utilizing huge data like genetic data, electronic health records (EHR), and real-time patient monitoring, replacing immediate procedures [5]. Accomplished discerning early indications for kidney malfunction through the examination of complicated patterns in patient data, customizing treatment programs, and enabling immediate intervention. Establishment can enhance the experiences of public health and enhance the kidneyhealthcare system's resilience in treating chronic illness by utilizing data-driven insights to build focused public health initiatives and allocate resources methods [7]. Data science is revolutionizing kidney health evaluation, enabling quick disease recognition, efficient treatment, and population-wide avoidance of kidney diseases [8]. The objective of the research is to enhance diagnostic precision and patient outcomes, data-driven models utilizing ML techniques to created for the early recognition and public health evaluation of kidney health in patients [6]. The rest of the paper was arranged into sections. The objective based related works shown in Section 2. Section 3 included a thorough methodology. Section 4 presents an outcome of research, and Section 5 provides the conclusion [13].

Related works

The triple-marker screen, educational initiatives, clinical champions, and telemedicine technologies are all part of the cascade of care (C3) a Veterans Affairs (VA) effort, to improve CKD identification in dangerous primary care patients [10]. Using pre-post observation design and secondary results, the first phase seeks to develop a universal infrastructure for the diagnosis of CKD. In [11] investigated CKD affects a large portion of the population worldwide and is a major health concern. Accurate predictions of CKD have been made by investigators using ensemble ML techniques, underscoring the possibility of rapid identification and individualized treatment. In the identification of CKD, a Public health cause of mortality globally, the research [12] developed an ML-based approach. The random forest (RF) method outperforms the other three classification algorithms in the system, which employs a public health of kidney disease patients. The article [14] described how ML techniques and information from the UCI repository can be used to predict CKD. Classifier methods were used, and the greatest accuracy in the synthetic minority over-sampling strategy was obtained by linear support vector machine (LSVM) with penalty L2. The maximum accuracy was attained by deep neural networks (DNN) [3]. There was disagreement on the need for early detection and intervention for CKD, a worldwide health problem [16] Immediate evaluation, risk assessment, and public health medical care for high-risk persons are critical, according to a symposium organized by kidney disease: Improving Global Outcomes (KDIGO).

2. Methodology

In this section, to improve the model's performance public health for kidney health assessment in data science, the preprocess used for Min-Max normalization and the feature selection used for RFE, the DEO-ULR hybrid model is suggested. Figure 1 shows the flow of the suggested approach.

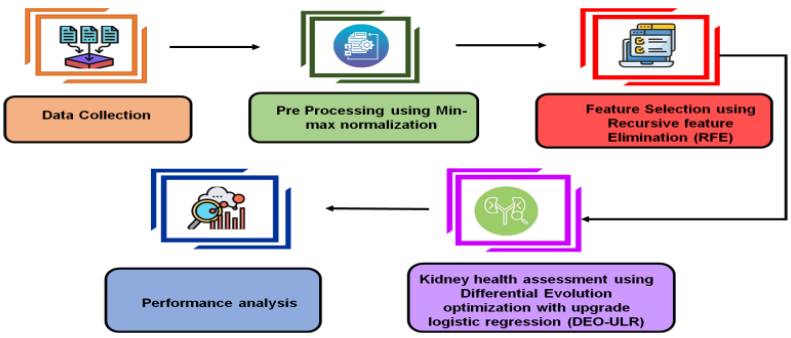

Figure 1 Flow of the suggested method

Dataset

The dataset was collected from open-source Kaggle website https://www.kaggle.com/datasets/rabieelkharoua/chronic-kidney-disease-dataset-analysis. The 1,659 people with a diagnosis of CKD have comprehensive health information available in this dataset. The dataset includes information on demographics, lifestyle characteristics, clinical measures, medical history, prescription drug use, symptoms, quality of life ratings, exposure to the environment, and health-related behaviors. Every patient has a unique Patient ID, and the information has a private column with the name of the attending physician.

Preprocessing using Min-Max normalization

A scaling technique called min-max normalization is utilized to convert data to an established range, usually across 0 and 1, to enable fair comparisons. It is used in public health evaluation to harmonies many health indicators to provide precise assessmentand diagnosis. Min-max normalization is a method of normalization that involves linearly transforming initial data to achieve a balance between pre- and post-process data comparisons using equation (1).

$$W_{new} = \frac{W - \min(W)}{\max(W) - \min(W)} \tag{1}$$

W_{new} = The updated value derived from the findings that were normalized

W = Previous value

$\max(W)$ = Maximum value inside the collection of data min (W) = Minimum value inside the collection of data

Recursive feature Elimination (RFE) using feature selection

The RFE selection procedure in relation to the evaluation of renal health, is essentially a recursive process that ranks features based on some metric of their value. To properly assess highly significant public health in kidney disease characteristics, a progressive elimination method that analyses feature importance and eliminates less relevant ones needs recursion. Robust feature selection customized for particular applications such as kidney health monitoring is ensured by public health iterative technique. While employing kernels to manage kidney disease, the experience has shown that utilizing linear models produces superior results in cases with low ratios. An RFE creates a linear decision function for a problem with classes, $yi \in \{-1, +1\}$.

$$C(w) = sign(w \cdot x) \quad (2)$$

The linear decision function is supplied by the vector x, which is perpendicular to the hyperplane, and w, which represents a vector containing the components of a specified spectrum, the kidney disease risk minimization public health principle using equation (1). The kidney feature's selection significance to the decision function is indicated by the components of x (x_i). When a component $x_i$has a large value, it suggests that feature i has a relatively high difference between the classes, as x points in the direction of the maximum separation. An issue can be broken down in kidney disease problems, $r = d(d-1)/2$, that each distinguishes between two d classes, to overcome the problems in the feature selection technique RFE.

$$C_j(w) = sign\left(w.x_j\right)j = 1 \ldots r. \quad (3)$$

The weight vectors $x_j$that represent every issue related to renal illness are then averaged.

$$x = \frac{1}{q}\sum_{j=1}^{r} x_j \quad (4)$$

The position of the characteristics is using equations (3-4) of the apparatus of X. To choose features for the RFE, the kidney health evaluation utilizes the components of X as metrics of feature implication.

Prediction of kidney health assessment using Differential Evolution optimization with Upgraded logistic regression (DEO-ULR)

The DEO-ULR technique is a way to enhance model concert by combining an enhanced version of ULR with the DEO algorithm. It is particularly helpful in conditions connecting intricate difficulty-solving, public health such as kidney health assessment, since it aids in the precise predictionof kidneyphysical condition parameters. By efficiently optimizing the model parameter, this technique guarantees precise and dependable health evaluation [15].

Upgraded logistic regression (ULR)

The $logit$ of several independent factors can explain the outcome variable, once additional manner a strong resemblance to ULR research following equation (5):

$$logit\, \gamma = \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3 + \cdots + \beta_m w_m \quad (5)$$

To expound on the previous instance, it considers whether age plays a role in the relationship between endothelial health and CKD. Both the risk of CKD and the likelihood of endothelial dysfunction, the predictive variable, are impacted, making it a potential confounding factor once more. Additionally, age cannot be regarded as an access effect because endothelial health has public health effect, the incorporated to research of ULR. Adjusting the maximal response to acetylcholine (ACh) did not alter the relationship between endothelial function and the risk of CKD. declared differently, there was minimal impact on the connection between CKD incidence and endothelial function.

Differential Evolution Optimization (DEO)

A heuristic approach called DEO is used to optimize complicated functions. When used in the assessment of kidney health in public health, DEO can assist in the research of medical data and enhance diagnostic precision by iteratively optimizing parameters [17]. A population of Mo, C-dimensional vectors is first initialized in classical DE, with parameter values distributed with kidney health assessment between lower and upper initial parameter boundaries that have been specifiedpublic health, $w_{i,high}$ and $w_{i,low}$, using equation (6).

$$w_{i,j,h} = w_{i.low} + rand(0,1) \cdot (w_{i,high} - w_{i,low}), \quad i = (1,2, \dots, C), \quad j = (1,2 \dots, Mo), \ h = 0 \tag{6}$$

The related public health evaluation indices are indicated, and the generation index is represented by the subscript h,i and j. The i^{th}dimension of the j^{th}increase of population in generations h is therefore $w_{j,h}$. An equally spaced value within the range $(0, 1)$ is produced using the random number generator rand $(0,1)$.To modify parameter values in DE, which appear as conventional floating-point numbers, one can employ standard floating-point operators. DEO modifies a vector from renal disease patients to develop a trial solution by adding the scaled variation among two more vectors from the current public health for kidney disease.

$$u_{j,h} = w_{q1,h} + E \cdot (w_{q2,h} - w_{q3,g}) q1, q2, q3 \in \{1,2, \dots, Mo\} \tag{7}$$

The vector indices $q1, q2, and\ q3$are chosen at random, but they are all unique and distinct from the population index. The j^{th} population vector, $w_{j,h}$, also known as the target vector, and any number of the values for this mutant vector's parameters, $u_{j,h}$, are consistently crossed $v_{j,h}$, the trial vector is the result followed by equations (7-8).

$$v_{i,j,h} = \begin{cases} u_{i,j,h} \ if \ rand(0,1) \le Dq \ \ or \ \ i = i_{rand}; \\ \qquad w_{i,j,h} \quad otherwise \end{cases} \tag{8}$$

The crossover constant controls how much of the traits from the mutant vector are transferred to the trial vector, which ranges from $0.0 < Cr < 1.0$. To guarantee that at least one parameter separates the trial vector from the vector that was compared to, the randomly chosen mutant vector parameter is always passed down to the trial vector index i_{rand}. Trial If the function parameters for vectors are equal to or below those of the target vector, then vectors replace the target vectors in subsequent generations.

$$w_{j,h+1} = \begin{cases} v_{j,h} \ if \ e(v_{j,h}) \le e(w_{j,h}) \\ \quad w_{j,h} \ otherwise \end{cases} \tag{9}$$

When issues include boundary restrictions, Excessive trial parameters appear back from the restriction by the degree of infraction to preserve the viability of solutions using equations (9-10).

$$v_{i,j,h} = \begin{cases} 2 \cdot x_{i,low} - v_{i,j,h} \ if \ \ v_{i,j,h} < x_{i,low} \\ 2 \cdot x_{i,high} - v_{i,j,h} \ if \ \ v_{i,j,h} > x_{i,high} \end{cases} \tag{10}$$

By combining the strength of DEO with the improved predictive capability of ULR, DEO-ULR performs better and more effectively in challenging data modeling tasks in public health for a kidney health assessment.

3. Results and discussion

Python 3.11 was significantly used in this investigation. Itsupplies Intel Core i9 laptops running Windows 10 on 33GB. Here the effectiveness of the proposed system is assessed, accuracy, precision, recall, and F1-score are among the evaluation metrics. Comparing Extreme Gradient Boosting (XGBoost) [12], Random Forest (RF) [12], Logistic Regression (LR) [18], and Adaptive Boosting (AdaBoost) [12], the proposed method DEO-ULR. Table 1shows the numerical results of the parameters.

Table 1 Numerical results of parameters

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
XG Boost [12]	94	92.56	93.42	95.2
RF [12]	93	90.4	93	91.3
LR [12]	90	92	92.7	92
AdaBoost [12]	91	91	90.5	94.5
DEO-ULR [Proposed]	**98.9**	**98.2**	**96.7**	**96.2**

Accuracy and Precision: To accurately diagnose and track kidney function, a kidney health assessmentfor public health accuracy is essential. Appropriate diagnosis and timely treatment of kidney-related disorders is ensured by reliable diagnostics. XGBoost scored 94%, RF scored 93%, LR scored 90% and AdaBoost scored 91%. The proposed DEO-ULR algorithm has a better accuracy of 98.9%.A precise and thorough analytical technique planned to observe the health and purpose of the public health to kidneys with a high degree of specificity and reliability is recognized as a precision kidney health assessment. XGBoost scored 92.56%, RF scored 90.4%, LR scored 92% and AdaBoost scored 91%. The proposed DEO-ULR algorithm has the betterprecision of 98.2%. Figure 2 shows the result of (a) Accuracy and (b) Precision.

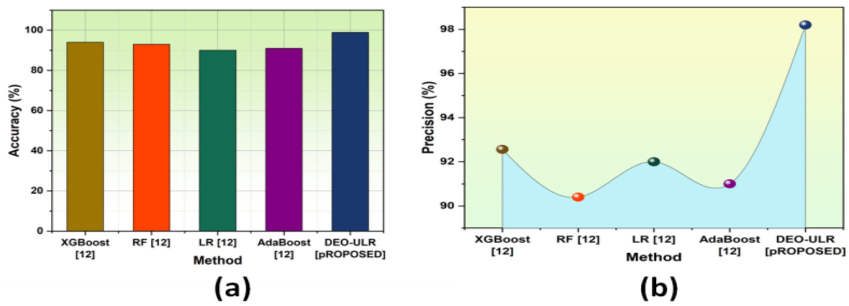

Figure 2 Outcome of (a) Accuracy (b) Precision

Recall and F1-Score: A medicinalevaluation or assessmentprocess targeted at examining or locating data applicablepublic health to kidney health is the majorityprobabledestined by recall kidney health consideration. It proposes an organized way of recall and evaluating significant information on kidney function or diseases [9]. XGBoost achieved 93.42%, RF achieved 93%, LR achieved 92.7% and AdaBoost achieved 90.5%. The suggested DEO-ULR algorithm has the highest recall of 96.7%. The F1-score is a statistic that is utilized to assess the dependability of kidney health evaluation by public health corresponding recall recording all positive instances and precision correctlyidentifying positive patients in a single determination. XGBoost achieved 95.2%, RF achieved 91.3%, LR achieved 92% and AdaBoost achieved 94.5%. The proposed DEO-ULR algorithm has a better F1-Score of 96.2%. Figure 3 shows the outcome of (a) Recall and (b) F1-Score.

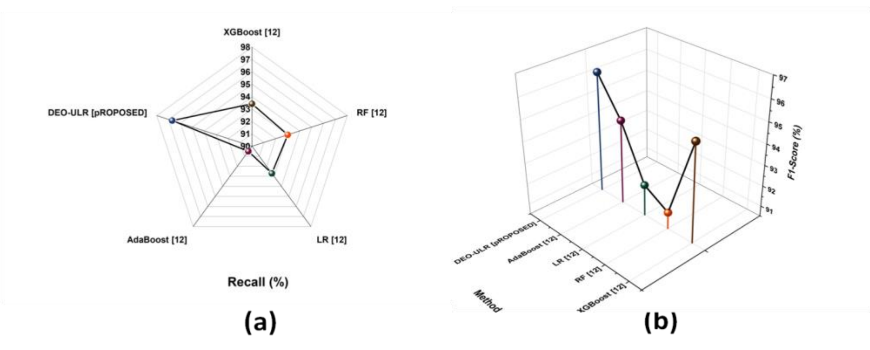

Figure 3 Outcome of (a) Recall (b) F1-Score

4. Conclusion and future scope

Data science for kidney public health assessment, monitoring, and forecasts kidney health with ML and data visualization. This makes kidney diseases treatable on an individual basis and enables early detection, enhancing patient outcomes and medical decision-making. The preprocessing uses Min-Max normalization and feature selection using RFE. When the DEO-ULR model was compared to other approaches, the researchfinds that present were important inenhancing the performance metrics for the proposed technique such as accuracy (98.9%), precision (98.2%), F1-Score (96.2%), and recall (96.7%). This study emphasizes data science for kidney health kidney assessment, with a focus on DEO-ULR models. Data science for kidney health assessment utilizes numerical examination for public health, ML, and data to evaluate, monitor, and forecast kidney health, enabling early identification and personalized treatment of kidney disorders, and improving patient outcomes and clinical executive.Future data science for kidney health evaluation might contain wearable approaches, immediate monitoring, and Artificial Intelligence (AI)improvements for early diagnosis, modified treatment regimens, and precise prediction models for the course of kidney illness.

Reference

[1] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy," *IEEE Access*, 8, pp.55012-55022, 2020. DOI: https://doi.org/10.1109/ACCESS.2020.2981689

[2] C. Kaur, M.S. Kumar, A. Anjum, M.B. Binda, M.R. Mallu, and M.S. Al Ansari, "Chronic kidney disease prediction using machine learning," *Journal of Advances in Information Technology*, 14(2), pp.384-391, 2023. DOI: https://doi.org/10.1109/ACCESS.2020.2981689

[3] Kutlu, Y., & Camgözlü, Y. (2021). Detection of coronavirus disease (COVID-19) from X-ray images using deep convolutional neural networks. Natural and Engineering Sciences, 6(1), 60-74.

[4] M.U. Emon, R. Islam, M.S. Keya, and R. Zannat, "Performance research of chronic kidney disease through machine learning approaches. *In 2021 6th International Conference on Inventive Computation Technologies (ICICT)* (pp. 713-719), 2021. DOI: https://doi.org/10.1109/ICICT50816.2021.9358491

[5] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B.Chen, "A machine learning methodology for diagnosing chronic kidney disease," IEEE Access, 8, pp.20991-21002, 2019. DOI: https://doi.org/10.1109/ACCESS.2019.2963053

[6] Sathyanarayanan, S., & Srikanta, M.K. (2024). Heart Sound Analysis Using SAINet Incorporating CNN and Transfer Learning for Detecting Heart Diseases. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 15(2), 152-169. https://doi.org/10.58346/JOWUA.2024.I2.011

[7] P. Sinha, and P. Sinha, "Comparative study of chronic kidney disease prediction using KNN and SVM," *International Journal of Engineering Research and Technology*, 4(12), pp.608-12, 2015.

[8] S. Vashisth, I. Dhall, and S. Saraswat, "Chronic kidney disease (CKD) diagnosis using multi-layer perceptron classifier," *In 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 346-350). IEEE, 2020. DOI: https://doi.org/10.1109/Confluence47617.2020.9058178

[9] Amiruzzaman, M., Islam, M. R., Islam, M. R., & Nor, R. M. (2022). Analysis of COVID-19: An infectious disease spread. Journal of Internet Services and Information Security, 12(3), 1-15.

[10] J.A. Lamprea-Montealegre, P. Joshi, A.S. Shapiro, E. Madden, K.Navarra, O.A. Potok, L.P. Gregg, T. Podchiyska, A. Robinson, M.K. Goldstein, and C.A. Peralta, "Improving chronic kidney disease detection and treatment in the United States: the chronic kidney disease cascade of care (C3) study protocol," *BMC nephrology*, 23(1), p.331, 2022. DOI: https://doi.org/10.1186/s12882-022-02943-z

[11] V. Srikanth, "CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS", 2023.

[12] E.M. Senan, M.H. Al-Adhaileh, F.W. Alsaade, T.H. Aldhyani, A.A. Alqarni, N. Alsharif, M.I. Uddin, A.H. Alahmadi, M.E. Jadhav, and M.Y. Alzahrani, "Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques," *Journal of healthcare engineering*, 2021(1), p.1004767, 2021. DIO: https://doi.org/10.1155/2021/1004767

[13] Kavitha, S. V., & Balasubramanian, P. (2022). Utilization of E-Resources among Women Faculty Members in Higher Educational Institutions in South Tamil Nadu. Indian Journal of Information Sources and Services, 12(1), 28–33.

[14] P. Chittora, S. Chaurasia, P. Chakrabarti, G. Kumawat, T. Chakrabarti, Z. Leonowicz, M. Jasiński, Ł. Jasiński,R. Gono, E.

Jasińska, and V. Bolshev, "Prediction of chronic kidney disease-a machine learning perspective," *IEEE access*, 9,pp.17312-17334, 2021. DOI: https://doi.org/10.1109/ACCESS.2021.3053763

[15] Salkić, Z., Lugović, B., & Babajić, E. (2021). Geochemistry and Petrogenesis of Oligocene Dacites from the Central Bosnia and Herzegovina with Insight in the Post collisional Tectonic Evolution of Central Dinaridic Ophiolite Belt. Archives for Technical Sciences, 1(24), 17–30.

[16] M.G.Shlipak, S.L. Tummalapalli, L.E. Boulware, M.E. Grams, J.H. Ix, V. Jha, A.P. Kengne, M. Madero, B. Mihaylova, N. Tangri, and M. Cheung, "The case for early identification and intervention of chronic kidney disease: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference," *Kidney international*, 99(1), pp.34-47, 2021. DOI: https://doi.org/10.1016/j.kint.2020.10.012

[17] Kabasa, B., Chikuni, E., Bates, M.P., & Zengeni, T.G. (2023). Data Conversion: Realization of Code converter using Shift Register Modules. Journal of VLSI Circuits and Systems, 5(1), 8-19

[18] M.A.R. Rahat, M.T. Islam, D.M. Cao, M. Tayaba, B.P. Ghosh, E.H. Ayon, N. Nobe, T. Akter, M. Rahman, and M.S. Bhuiyan, "Comparing Machine Learning Techniques for Detecting Chronic Kidney Disease in Early Stage," *Journal of Computer Science and Technology Studies*, 6(1), pp.20-32, 2024. DOI: https://doi.org/10.32996/jcsts.2024.6.1.3