ISSN (ONLINE) : 2045-8711
ISSN (PRINT) : 2045-869X

INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY & CREATIVE ENGINEERING

December 2023
Vol - 13 No - 12
@IJITCE Publication

UK: Managing Editor
International Journal of Innovative Technology and Creative Engineering
1a park lane,
Cranford
London
TW59WA
UK

USA: Editor
International Journal of Innovative Technology and Creative Engineering
Dr. Arumugam
Department of Chemistry
University of Georgia
GA-30602, USA.

India: Editor
International Journal of Innovative Technology & Creative Engineering
36/4 12th Avenue,
1st cross St,
Vaigai Colony
Ashok Nagar
Chennai, India 600083

Email: editor@ijitce.co.uk

www.ijitce.co.uk

IJITCE PUBLICATION

International Journal of Innovative Technology & Creative Engineering

Vol.13 No.12

December 2023

IJITCE
INNOVATE TO CREATE

www.ijitce.co.uk

Dear Researcher,

Greetings!

Articles in this issue discusses about study endeavours to recent trends in E-Banking.

We look forward many more new technologies in the next month.

Thanks,
Editorial Team
IJITCE

Editorial Members

Review Board Members

Contents

RISK PREDICTION OF RENAL FAILURE FOR CHRONIC DISEASE POPULATION BASED ON ELECTRONIC HEALTH RECORD BIG DATA

K.SARANYADEVI
Research scholar, Navarasam Arts &Science college for women, Arachalur, Erode
Dr. P. RATHIGA
Assistant professor, PG and Research Department of Computer Science, Erode arts and Science college,

Abstract

Renal failure is a deadly condition that is causing worldwide concern. Previous risk models for renal failure relied heavily on the identification of chronic kidney disease, which lacks evident clinical signs and so goes misdiagnosed, resulting in a considerable underreporting of high-risk individuals. In this research, we developed a system for predicting the probability of renal failure directly from a chronic illness population's big data repository, without the need for a prior diagnosis of chronic kidney disease. During a 3-year follow-up, the electronic health data of 42,256 patients with hypertension or diabetes in Shenzhen Health Information Big Data Platform were obtained, with 398 suffering from renal failure. Five cutting-edge machine learning algorithms are used to develop risk prediction models for renal failure in a chronic illness population. Extensive experimental findings indicate that the suggested framework performs admirably.

The XGBoost, in particular, achieves the greatest performance, with an area under the receiving-operating-characteristics curve (AUC) of 0.9139. We discovered that serum creative, age, urine acid, systolic blood pressure, and blood urea nitrogen are the top five risk factors for renal failure by assessing the influence of risk factors. In comparison to previous models, our approach may be included into normal chronic disease management processes, allowing for more proactive, widely-covered screening of kidney hazards, which would lessen disease damage by prompt intervention. Diagnosis of CKD is still inadequate at the clinical level and it is not possible to detect the CKD in early stage. Recently machine learning based approaches provides the efficient result in disease diagnosis. The present study retrospect's the recent researches related to the chronic kidney disease diagnostic using machine learning approaches. This research assists to analyze the drawbacks of the prior study and provides a path for most applicable detection system. In comparison to previous models, our approach may be included into normal chronic disease management processes, allowing for more proactive, widely-covered screening of kidney hazards, which would lessen disease damage by prompt intervention.

Keywords:
Renal failure Risk prediction, Electronic Health record, Big Data, Machine Learning.

1. Introduction

Renal failure, also known as end-stage kidney disease (ESKD), is a pathological state of partial or total loss of renal function caused by the development of chronic kidney diseases (CKD) to the later stage. Patients with renal failure would soon suffer from uremia or even deadly consequence, and the treatment can only be dialysis or renal transplantation. The prevalence and total mortality of renal failure continue to increase [1]. In 2016, there were 720,000 patients with renal failure in the United States, and the hospital mortality rate of all dialysis patients was 0.5% [2]. In China, the number of renal failure patients was about 2.9 million and the mortality rate among dialysis patients was 28.42 per thou- sand years[3]. The difficulty of reversing renal damage increases steadily with the disease progression, thus early detection of high- risk groups for renal failure is particularly important to enable

early interventions. .

Extracting and analyzing retrospective population data from electronical health record (EHR) big data platforms would largely ex- tend the feasibility of many clinical studies in the scope of data availability, and we will demonstrate this in our renal failure study as well.

Several prospective cohort studies and cross-sectional studies have been conducted to develop CKD risk prediction models [4], such as SCORED score [5], ARIC/CHS score [6], Framingham score [7], QKidney score [8], Taiwan score [9], Japan/HIV score [10], and ADVANCE model [11]. The investigated risk factors mostly include age, gender, body mass index, blood pressure, diabetes status, serum creatinine, proteinuria, serum albumin, and total protein. In addition, some studies added novel biomarkers such as smoking, kidney stones, and family history of kidney disease, or genetic fac- tors [12] to improve model performance. Subsequently, risk models for predicting progression to ESKD have been developed by meta- analysis, the most famous of which is the 4-variable Kidney Failure Risk Equation (KFRE), using gender, age, estimated glomerular filtration rate (eGFR), and urine albumin-to-creatine ratio (ACR) [13]. There are also two ESKD prediction equations based on 6 variable. which specifically boost the efficiency of observational studies. On the other hand, machine learning techniques are being used more and more widely for clinical analysis due to its strong potential to use com- plex mathematics operations to compute large assertion and prevention of renal failure are mainly focused on CKD patients. However, the awareness rate of early CKD is low, which is less than 10% in developing and developed countries, and only 12.5% in China [1,3]. Most patients with CKD have no obvious symptoms in the early stage of onset, resulting in a very high rate of missing diagnosis among general population. A low awareness rate for doctors also exists, and nearly half of the country's attending and deputy doctors have a lower av- erage understanding of CKD guidelines [1]. The high undiagnosed rate of CKD poses a severe challenge to renal failure prevention, as a large portion of high-risk patients were not monitored for dis- ease risk in the early. In this paper, we strove to extend the feasibility of renal risk prediction from CKD patients to general chronic disease populations. A total of 42,256 registered patients with hypertension or diabetes were selected from Shenzhen Health Information Big Data Platform. After rigorous population screening, only 5,974 patients were retained, of whom 398 had renal failure during a three-year follow-up. Five machine learning algorithms were used to establish the three-year risk models of renal failure, among which the integrated algorithm XGBoost achieved the optimal performance on the test set. Furthermore, we analyzed the univariate effect of renal failure and showed nine continuous variables that were non- linearly correlated with renal failure risk. The contribution of our work can be summarized into three scopes. Firstly, for the first time we extended risk modelling for renal failure to non-CKD patients by conducting a large-scale retrospective study, which was achieved by more efficient curation of target data through the aid of big data technologies. Secondly, with sophisticated machine learning methods, we were able to study a relatively large number of features simultaneously. The high undiagnosed rate of CKD poses a severe challenge to renal failure prevention, as a large portion of high-risk patients were not monitored for dis- ease risk in the early. In this paper, we strove to extend the feasibility of renal risk prediction from CKD patients to general chronic disease popula- tions. A total of 42,256 registered patients with hypertension or diabetes were selected from Shenzhen Health Information Big Data Platform. After rigorous population screening, only 5,974 patients were retained, of whom 398 had renal failure during a three-year follow-up. Five machine learning algorithms were used to establish the three-year risk models of renal failure, among which the integrated algorithm XGBoost achieved the optimal performance on the test set. Furthermore, we analyzed the univariate effect of renal failure and showed nine continuous variables that were non- linearly correlated with renal failure risk. The contribution of our work can be summarized into three scopes. Firstly, for the first time we extended risk modelling for renal failure to non-CKD patients by conducting a large-scale retrospective study, which was achieved by more efficient curation of target data through the aid of big data technologies. Secondly, with sophisticated machine learning methods, we

were able to study a relatively large number of features simultaneously. As a result, we discovered some novel biomarkers of renal failure, including uric acid (UA), aspartate aminotransferase (AST), alanine transaminase (ALT), and total bilirubin (TBIL), which were not included in previous models, and identified their nonlinear role in renal function disorder. Thirdly, the proposed model was based on daily monitoring and physical examination data that are easy to acquire for both CKD and non-CKD chronic disease patients. Therefore, it can be deployed into chronic disease management systems to aid physicians to early identify high-risk population for timely intervention. Materials and methods.

Data resource

The data used in this paper are from Shenzhen Health Information Big Data Platform, which has access to more than 4,000 health institutions including 85 hospitals and more than 650 community health service centers. The platform covered medical service records including outpatient, inpatient, biochemical test, imaging examination, physical examination, and regular follow-up records of registered patients with hypertension, diabetes, cancer, and other diseases. At present, the platform has more than 5 billion medical service records and 598 million electronic medical records, covering a time span from 2010 to 2020. Medical records among different institutions of the same patient can be associated with a unique personal identification number. Due to the case that all medical records were collected in routine clinical activities and the anonymous nature of the obtained data, following the Guidelines of the WMA Declaration of Helsinki term 32, a waive-of-consent protocol was adopted and was approved by the SIAT IRB with No. SIAT-IRB-151115-H0084. The causes of renal failure are complex, diabetic nephropathy (43.2%) and hypertension (23%) form the main causes of renal failure worldwide [2]. Moreover, a large portion of patients with diabetes and hypertension tend to receive periodic physical examinations, thus a large number of laboratory test result data needed for renal risk prediction have been accumulated, as in the case of the Shenzhen Health Information Big Data Platform. Therefore, this study mainly focused on predicting renal failure risk for these two types of chronic disease patients with high incidence and standardized management. The main goal of this work is to establish a high-precision three-year short-term risk prediction model for the two major chronic disease population of hypertension and diabetes, based on the real-world population.

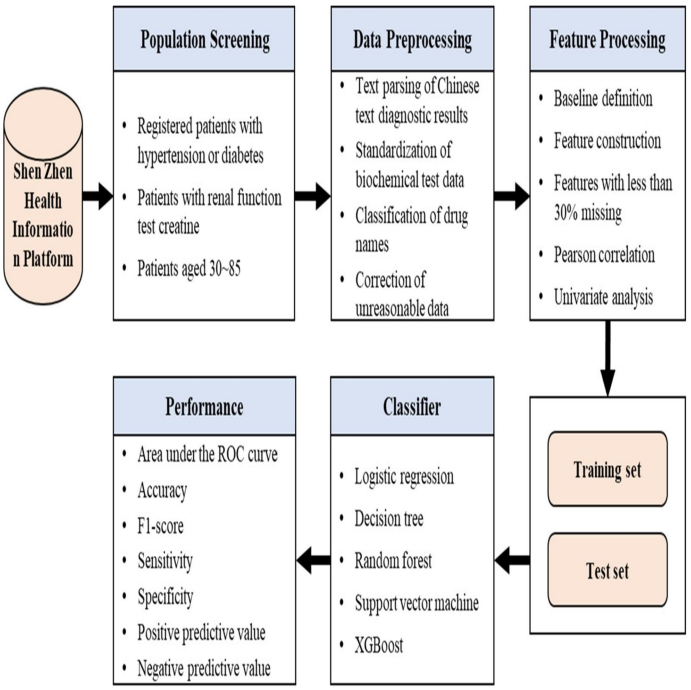

Fig. 1. The pipeline of the study. ROC: receiver operating characteristic

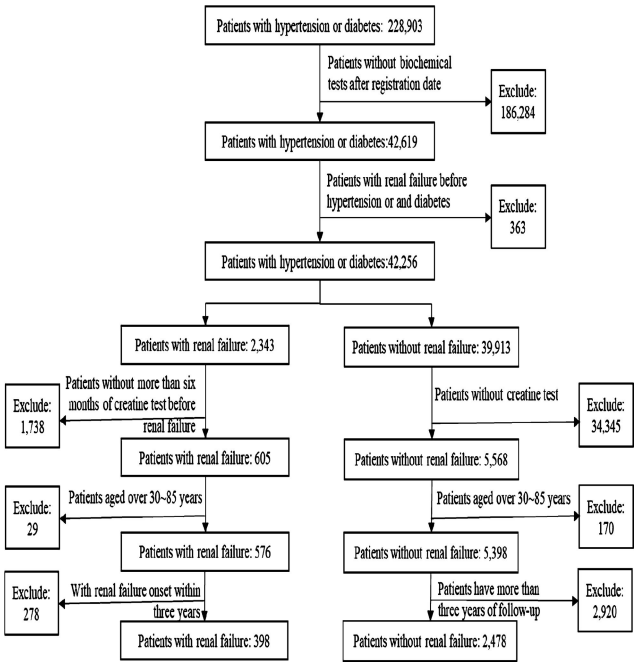

Fig. 2. The process of screening study population

2. Related Works

N.Bhaskar and M.Suchetha [2] used the combination of convolutional neural network with support vector machine to develop an automated diagnosis system. The traditional CNN is enhanced with SVM in order to overcome the drawbacks of CNN. This model utilizes the salivary urea as potential biomarker for CKD diagnosis. Indian patient data is collected for this research experiments. The detection system is evaluated with all possible performance metrics and achieved 98.67% accuracy.

Qin et al [3] adopted the machine learning algorithm to discover the CKD using the samples collected from UCI repository. The data in UCI contains large missing value and processed with KNN. The authors used best performing six ML approaches and evaluated this model and integrate the logistic regression and random forest with perceptron for disease diagnosis. The model with KNN and combine LR with RF achieved 99.83% accuracy.

Ahmed Abdelaziz et al [4] introduced an approach for diagnosing and predicting the disease anywhere through cloud computing. The LR and NN algorithm is utilized for critical factor analysis and prediction respectively. Windows azure is accessed for cloud environment and the present model achieved 97.8% accuracy. Based on the model a case study is performed with three different patient data.

Jain, Divya and Vijendra Singh [5] in this examination, a quick, novel versatile classification system is displayed for the conclusion of chronic diseases. For this reason, the proposed methodology utilizes a hybrid methodology involving PCA and Relief technique with enhanced Support Vector Machine classifier.

Besra, et.al [6] in this original copy, we have proposed a system that will generate a prediction of CKD with higher precision esteem, trailed by the estimation of kidney harm rate. The fundamental goal of this analysis is to robotize a prediction system that will analyze the various stages in CKD. It begins with the prehandling steps, closes with the classification, recognizes the effectively characterized occasions, and then figures its GFR esteem.

El-Houssainy et al [7] build up the classifier model with PNN, MLP, SVM and RBF machine learning algorithms for CKD stages prediction process. The dataset is collected from UCI repository which has 361 instances. The system achieved great performance of 96.7% accuracy when using PNN. This model attempts to predict the different stages of CKD.

Jivan parab et al [8] the principle goal of this paper is to give an account of research where we exploited those accessible technological progressions to create prediction models for CKD prediction on diabetes patients, and furthermore the primary objective of restorative information mining systems is to get best calculations that depict given information from various perspectives. In this examination, three information mining methods (BP- ANN and PLSR) are utilized to inspire learning about the connection between these factors and patient endurance. The model is developed based on two factors like blood urea and glucose. The principle component analysis is further used to improve the accuracy of the model.

Wang et al [9] introduced the multi-task deep and wide neural network classifier to discover the Renal failure prediction from heart patients. The investigation depends on the EHR information containing right around too many years of clinical perceptions gathered at PLA General Hospital, a huge medical clinic in Beijing with one of the most established electronic health records in the china. The dataset is collected from Chinese hospital and start the prediction process with missing value elimination and normalization. Multi-task deep and wide neural network classifier (MT-DWNN) is then applied to diagnosis the renal failure which is illustrated in figure 1. The input layer and all the hidden layers are shared layers, while the output layer is a specific layer for different tasks. The Roc and Auc is computed to evaluate the performance of this study.

3. Discussion

Here developed a high-precision risk prediction model of renal failure for chronic disease patients with hypertension or diabetes based on electronic heath records from the Shenzhen Health Information Big Data Platform. Unlike existing studies, our model does not require patients to be diagnosed with CKD, which avoid the severe defect of low coverage for previous models led by the high undiagnosed rate of CKD patients in clinical practice. Collecting blood samples from large-scale non-CKD population and per- forming long-term follow-up have been difficult and costly. How- ever, in our work, we manage to curate the data with the aid of big data technologies through extracting useful information from routine clinical records in the large-scale regional medical information

platform, making it feasible to perform massive observational cohort studies more efficient Our findings partially overlap with some other early studies on patients with CKD. For example, ARIC/CHS score and Framingham score include age, gender, hypertension, diabetes, BMI, and HDL-C. Taiwan score and ADVANCE model include ACR, UA, glucose, and proteinuria. Also, the prediction model of CKD progression KFER includes CREA, ALB, and history of CKD, stroke, heart failure, and arrhythmia. More importantly, we further identified several new prediction biomarkers such as AST, ALT and TBIL with the power of sophisticated machine learning methods, and discovered their non- linear role in renal dysfunction. The effect of nonlinear correlation justifies the necessity of adopting sophisticated nonlinear machine learning models over traditional linear regressions. Furthermore, with non-linear ensemble algorithms such as XG Boost used in our work, there is no need to select variables in advance even when the number of potential variables is large, which is different from most traditional clinical studies and enables identification of novel biomarkers with both linear and non-linear effects during modeling process through mining large-scale population data. This is another advantage brought by big data technologies.

4. Conclusion

In conclusion, have developed and validated a highly accurate risk model for predicting renal failure of chronic disease patients with hypertension or diabetes, without necessarily early diagnosis of kidney diseases, which advance the state-of-the-arts for renal failure prediction. The model uses routinely available physical and laboratory examination data and could predict the short-term risk of renal failure with high accuracy. Due to the ease of access to data, it could be easily implemented in laboratory information systems or EHR systems to help with a more pervasive, preemptive screening of renal failure risk, enabling higher efficiency of early disease prevention and intervention. Our works also justify the advantages of adopting big data technologies in public health as well.

References

[1] GBD Chronic Kidney Disease Collaboration, Global, regional, and national bur- den of chronic kidney disease, 1990–2017: s systematic analysis for the Global Burden of Disease Study 2017, Lancet 10225 (395) (2020 February 29) 709–733, https://doi.org/10.1016/S0140-6736(20)30045-3, Epub 2020 Feb 13.

[2] Somnath Pal, Primary causes of end-stage renal disease, US Pharm. 41 (8)(2016) 6.

[3] L. Zhang, M.H. Zhao, L. Zuo, Y. Wang, F. Yu, H. Zhang, H. Wang, CK-NET Work Group, China Kidney Disease network (CK-NET) 2015 annual data report, Kid-ney Inter., Suppl. 9 (1) (2019 Mar) e1–e81, https://doi.org/10.1016/j.kisu.2018. 11.001.

[4] J.B. Echouffo-Tcheugui, A.P. Kengne, Risk models to predict chronic kidney disease and its progression: a systematic review, PLoS Med. 9 (11) (2012) e1001344,

[5] H. Bang, S. Vupputuri, D.A. Shoham, P.J. Klemmer, R.J. Falk, M. Mazumdar,

December 2023

Vol - 13 No - 12

@IJITCE Publication