Journal of Positive Sciences (JPS) eISSN: 2582-9351, https://positive-sciences.com JPS. 2023 Sep; 3(5): 1-5

Article info
Received on: 29.08.2023
Accepted on: 29.09.2023
Published on: 30.09.2023
doi: https://doi.org/10.52688/ASP75091

Research Article

Identification and prediction of kidney diseases by using machine learning techniques

Shaymaa Adnan Abdulrahman [1, *], Mohamed Roushdy [2], Abdel-Badeeh M. Salem [3]

[1, 2] University of Information Technology & Communications, Business Information Technology Department, Baghdad, Iraq

[2] Faculty of Computers & Information Technology, Future University in Egypt, New Cairo, Egypt
[3] Department of Computer Science, Ain Shams University, Cairo, Egypt

* dr.shaymaa.adnan@Uoitc.edu.iq

ABSTRACT

Machine learning (ML) applications in health informatics are receiving more and more attention. One scenario that highlights the important function of ML diagnostic algorithms is the timely diagnosis of renal disease and the following prompt reaction to it. The goal of ML in Kidney Disease Diagnosis (MLKDD), an active research area, is to help doctors diagnose kidney diseases using computer-aided systems. Numerous studies have attempted to evaluate the viability, applicability, and superiority of various ML techniques over one another. This study uses Convolutional Neural Networks (CNN) and other machine learning techniques to determine healthy or patient individuals from medical photographs. We used CNN to evaluate a number of machine learning techniques, including Naive Bayes, K- Nearest Neighbour, Decision Trees, Support Vector Machines, and deep neural networks. The experiments show that the CNN Classifier produces the best classification outcomes.

Keywords: Kidney Disease, Machine, Learning (ML), Support Vector Machine (SVM), K.-Nearest Neighbour (KNN)

INTRODUCTION

Kidney malfunction has far-reaching effects because the kidney is responsible for so many bodily functions, including controlling the concentration of water. and minerals in. the blood and maintaining. internal equilibrium. A patient's prognosis for recovery from Chronic Kidney condition(CKD), also “known as Chronic Renal Failure (CRF)”, is greatly improved if the condition is detected and treated early on. From 1990 to 2013, the global annual death toll attributable to CKD grew by 90%, making it the thirteenth greatest cause of mortality worldwide [1]. The prevalence of renal disease is estimated to be 850 million worldwide [2][3]. At least 2.4 million people die annually from kidney-related causes, as reported by World Kidney Day 2019. The World Health Organisation ranks it as the (6-sixth) leading cause of death today. The rising global prevalence, of chronic kidney disease poses significant public. health challenges. The problem is exacerbated in countries with low incomes due to decreased funds for The Glomerular Filtration Rate (GFR) gauges how well the kidneys are functioning, and the National Kidney Foundation has classified CKD into five phases based on aberrant kidney function and lower GFR. Stages 1 and 2 are the mildest, with little noticeable symptoms, while stage 5 is the final, life-threatening stage, also known as renal failure. Total renal failure treatment, known as Renal Replacement Therapy (RRT), can rack up hefty medical bills. Most underdeveloped countries, including Ethiopia, do not have access to the treatment. Therefore, in developing nations, it is challenging to manage renal failure and its complications due to a lack of facilities, specialists, and the high. cost of treatment [4.][5].As a result, detecting CKD early is crucial for reducing costs and improving treatment outcomes [6]. Early diagnosis of CKD through predictive analysis utilising machine learning approaches can aid in effective and timely therapies [7]. Renal failure is the end result of this disorder, which is characterised by a steady decline in renal function. In its early stages, CKD has no outward signs. The kidney's function may decline by 25% before the disease is recognised [8]. In addition, CKD has worldwide effects on human health, including a high morbidity and mortality rate [9]. The risk of developing cardiovascular disease may be reduced [10]. A fatal and degenerative condition, chronic kidney disease (CKD). Being able to predict and diagnose CKD early on may enable patients to receive therapy on time, delaying the disease's progression. The process of teaching a machine to execute a certain task by studying instances of that task and extrapolating its properties from those examples is known as machine learning. [11]. This method has the potential to offer accurate and affordable illness diagnosis.; hence, it may serve as a useful tool in the detection of (CKD.) With the advent of “IT” it has evolved. into a novel sort of medical. instrument with wide ranging potential applications thanks to the proliferation of EHRs. Machine learning has already been put to use in the medical industry, where it may be used to identify human body status,

***Corresponding author**
Shaymaa Adnan Abdulrahman,
University of Information Technology & Communications, Business Information Technology Department, Baghdad, Iraq
e-mail: dr.shaymaa.adnan@Uoitc.edu.iq

analyse relevant disease characteristics, and diagnose a wide range of illnesses. Machine learning(ML) algorithm have been applied to the diagnosis of numerous disorders, including cardiovascular disease [14], diabetes and retinopathy [15], acute renal injury [16], cancer [17], and others [18] [19]. All methods based on regression, trees, probabilities, decision surfaces, and neural networks performed admirably in these models. The National renal Foundation has divided CKD into five phases based on aberrant renal function and reduced GFR, which is a measurement of the degree of kidney function. The earliest stages are 1 and 2, which have barely perceptible symptoms, whereas stage 5, often known as renal failure, is the most severe and potentially fatal stage. Total renal failure treatment, known as Renal Replacement Therapy (RRT), can rack up hefty medical bills. Most underdeveloped countries, including Ethiopia, do not have access to the treatment. Consequently, in developing countries, the care of renal failure and associated consequences is challenging because of a lack of facilities, specialists, and the high cost of therapy [18]. Therefore, detecting CKD early is crucial to reduce costs and improve treatment outcomes. Predictive analysis based on machine learning(ML) approaches can aid in the early detection of (CKD) allowing for more effective and timely therapies to be implemented. In this research, CKD detection was attempted utilising a many machine learning method, such as Decision Trees (DT), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Naive Bayes; and deep neural. network with CNN. Most studies have only looked at CKD in two categories, which makes it hard to establish therapy recommendations based on the severity. of the disease. The outline of the paper looks like this: The following sections comprise the paper: sections: 2. Introduction; 3. Overview of the Methodology; 4. Classifier; 5. Experimental Results; 6. Conclusion.

LITERATURE REVIEW

Given the vast volume of everyday medical data produced, determining the most efficient methods for disease diagnosis is a crucial requirement for this literature. Researchers in this field have focused on two main areas: (1) creating novel ML techniques, and (2) comparing the various existing methods in order to identify the most effective computer aided systems for improved the detection of kidney disorders. Several machine-learning algorithms may be useful for early detection of chronic kidney disease, as explored by Aljaaf et al.[19]. Despite the fact that this has been studied at length, we are bolstering our technique with predictive analytics. which examines the association between data factors and target group or class characteristics [19]. Deep learning combined with machine learning is used to by the author [20] when implemented their solution. In particular, the SVM model outperformed all other machine learning (ML) algorithm in terms of accuracy. Highly precise CKD prediction models were put into use by the authors of [21] by combining such as (SVM / AdaBoost / linear discriminant analysis/ (LDA) and gradient boosting (GBoost.) method. The effectiveness of these models was measured using data mined from the repository for machine learning at UCI. The best accuracy (99.80) percent, was achieved by the gradient boosting classifier. The (LR), Decision Tree (DT,) and(k- NN) algorithms were employed to training three alternative models for predicting CKD by the authors of [22]. When compared to DT (96.21%) and kNN (71.24%) , LR (97) % accuracy is much superior. Gorzaáczany, M.B.; RudziEski, F.(2017) [23] explains how the Internet of. Medical Things (IoMT) can be used to analyse camera-captured photos for signs of disease in humans, with a 96.78% verification rate. for histopathology images. Fourth on the list of technological revolutions is deep learning (DL) and machine learning (ML), with DL relying on artificial neural networks [24]. K nearest neighbour (KNN) and support vector machine (SVM) classification algorithms, as well as logistic regression. (LR), decision tree classifiers. were all identified by G.-S. Fu et.al. (2019) [25] as being useful for predicting the likelihood of CKD. The results showed that the SVM classifier has the greatest accuracy (97.6%) in the experiment. After evaluating the suggested strategy, SVM has the highest sensitivity. From this comparison, it is feasible to infer that an SVM classifier is used to predict chronic renal illness. [26]. A machine learning (ML) prediction models for CKD early detection was created G.-S. Fu et.al. (2019) [27]. Models have been evaluated and validated by using the provided input features, which were acquired from the CKD dataset. The CKD diagnosis was made using three different machine learning (ML) classifier: the decision tree (DT) classifier and the random forest classifier, and the support vector (SVM) classifier. (RF) Random forest, K-nearest neighbours, and neural network were all tested by A. Salekin and J. Stankovic [28] for their ability to identify CKD. They used a dataset of24 attributes based on information from 400 UCI patients. The characteristics that accurately diagnose this illness have been identified by a feature reduction study utilising the wrapper technique. “Albumin, specific gravity, diabetes mellitus, haemoglobin, and hypertension” are all considered to make a.98 F1 and a.11 RMSE prediction of CKD. In another report [28], researchers looked into how morphological surgeries influence the categorization and evaluation of kidney stones. GAC segmentation, in addition to extraction and morphological techniques, has been tried to assess the position and the kidney stone's size. Several kidney pictures have been processed using the proposed techniques, with promising results [28] as shown Table 1

*Corresponding author
Shaymaa Adnan Abdulrahman,
University of Information Technology & Communications, Business Information Technology Department, Baghdad, Iraq
e-mail: dr.shaymaa.adnan@Uoitc.edu.iq

Table 1: Summary of, some related works

NO	Author	Techniques applied	Result	Years
1	Aliaaf etal [19]	FOURE Machine learning	OUTCOME OF AUC 0.995,	2018
2	Debabrate swain etal [20]	Machine learning SVM	HIGHEST ACCURACY	2023
3	Ghosh ,p etal [21]	SVM, Machine learning AdaBoost, linear discriminant (LD) analysis (LDA),	highest accuracy of 99.80%	2022
4	fraz, G.M[22]	LR, Decision Tree (DT), and k- NN	accuracy (97%)	2022
5	Gorzaáczany, M.B etal [23]	Deep learning (DL)	96.88% verification rate for histopathological images.	
6	K. R. A. Padmanaban et al[24]	Machine learning (KNN), (SVM), (LR),	highest accuracy, 98.3%.	2016
7	G.-S. Fu, Y. Levin et al [25]	Machine learning by using DT, RF SVM methods	-	2019
8	A. Salekin and J. Stankovic [26]	Machin learning	predict the CKD with .98 F1 and 0.11 RMSE.	2016
9	M. Shahina and H. S. Mahesh [27]	Machine learning GAC segmentation	high efficiency	2019

OVERVIEW OF THE METHODOLOGY

The next part provides a detailed, step-by-step explanation of the approach taken in this study.

PROPOSED SYSTEM

A strong model is proposed in this study that may accurately diagnose chronic renal disease. An efficient and precise prediction model was created using an ML-based strategy. A simplified diagram is shown in Figure 1. It shows how the proposed system develops over time

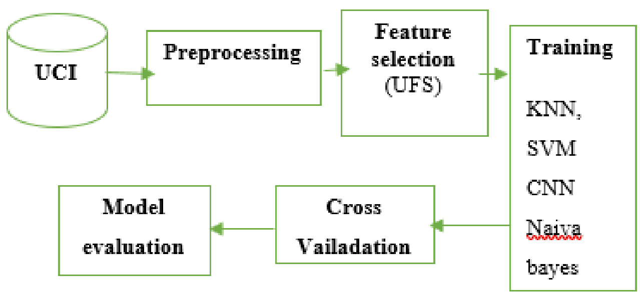

Figure 1: Block diagrams of the .proposed system.

DATASET COLLECTION

CKD patients at Apollo Hospital in India provided the data for this study over a two-month period in 2015. Chronic_Kidney Disease DataSet [18] contains the information and can be found in the UCI data repository. There are some missing and noisy values in these 400- observation data set. There are 250 CKD patient records and

150 healthy control records in the database. Therefore, 62.5%

***Corresponding author**
Shaymaa Adnan Abdulrahman,
University of Information Technology & Communications, Business Information Technology Department, Baghdad, Iraq
e-mail: dr.shaymaa.adnan@Uoitc.edu.iq

of the population has CKD while 37.5% of the population does not. These reports come from people as young as two and as old as ninety. According to Table 2, there are a total of

25 features in the CKD dataset; this includes the categorization or condition of CKD as well as 11 numerical features and 13 nominal features.

TABLE 2: Description of [CKD] Dataset

Name	Description	Type: unit/ values
[Age.]	(Patient's age)	(Numeric: years)
[Blood pressure, (bp.)]	(Blood pressure of the patient.)	(Numeric:mm/Hg)
[Specific gravity (sg)]	(The ratio of the density of urine).	(Nominal: 1.005, 1.010, 1.015,
[Albumin (al)]	("Albumin level in the blood")	(Nominal:[0,1,2,3,4,5)]
[Sugar (su)	(Sugar level of the patient:	Nominal::[0,1,2,3,4,5]
[Red. blood cells (rbc)	9Patients' red blood cells count)	Nominal:: normal, abnormal
[Pus. cell (pc)	(pus cell count of patient"	Nomina:l: normal, abnormal
[Pus cell .clumps (pcc)	(Presence of pus cell clumps in the blood)	Nominal:: present, not , present
[Bacteria ,(ba)	(Presence of bacteria in the blood)	Nominal:: present, not present
[Blood glucose .(bgr)	(blood glucose random count)	Numeric: :[mgs/dl].
[Blood urea (bu)	(blood urea level of the patient")	Numeric::[mgs/dl].
[Serum creatinine	(serum creatinine level in the blood)	Numeric::[mgs/dl]
[Sodium (sod)	(sodium level in the blood)	Numeric:[;mEq/L]
[Potassium (pot)	"potassium. level in the blood;	Numeric::[mEq/L]
[.Hemoglobin (hemo)	hemoglobin level in the blood	Numeric: ;(gms).
[.Packed cell volume	:packed cell volume in the bloo	[Numeric.]
.[White blood cell count (wc)	"white blood cell count of the patient)	.[Numeric. millions/[cmm]
.[Red blood cell count (rc)	(red blood cell count of the patient:	Numeric millions/[cmm]
.[Hypertension (htn)	"Does the patient has hypertension on not)	Nominal: :yes, no
.[Diabetes mellitus].	9Does the patient has diabetes or not)	Nominal:: yes, no
["Coronary artery disease (cad)	(Does the patient has coronary artery disease or not)	Nominal: yes, no
["Appetite (appet];)	(Patient's appetite0)	[.Nominal: yes, no
"Pedal Edema (pe)].	Does patient has pedal edema or not)	[.Nominal: yes, no
["Anemia (ane];)	Does patient has anemia or no)	[.Nominal: yes, no
{"Class	(Does the patient has kidney disease or not)	[.Nominal: CKD, not CKD

DATA PREPROCESSING

The immediately method was used to encode all categorical information. These included insurance coverage, level of education, and primary diagnosis. If more than half of a variable's values were missing during model building, that

REFERENCES

[1] Janez, Potočnik Renewable energy sources and the realities of setting an energy agenda Science, 315 (2007), pp. 810-811
[2] Dincer Ibrahim, Renewable energy and sustainable development: a crucial review, Renewable & Sustainable Energy Reviews, 4 (2) (2000), pp. 157-175
[3] Hü seyin Benli, Durmus Aydın, Evaluation of ground-source heat pump combined latent heat storage system performance in greenhouse heating, Energy and Buildings, 41 (2009), pp. 220-228
[4] EPA. A short primer and environmental guidance for geothermal heat pumps.US Environmental Protection Agency;1997.
[5] Q.C. Yang, J. Liang, L.C. Liu, Numerical Model for the Capacity Evaluation of Shallow Groundwater Heat Pumps in Beijing Plain, China. Procedia Environmental Sciences, 10(A) (2011), pp. 881-889

***Corresponding author**
Shaymaa Adnan Abdulrahman,
University of Information Technology & Communications, Business Information Technology Department, Baghdad, Iraq
e-mail: dr.shaymaa.adnan@Uoitc.edu.iq

[6] Wang Meizhi, Liujijun, Tianjianhui, et al. Feasibility discussion on ground-source heat pump technique application in pig farms in Beijing. Proceedings of the symposium of rural renewable energies and low- carbon technologies in China. Chinese Society of Agricultural Engineering2011:414-417.

[7] Lilong Chai, Chengwei Ma, Ji-Qin Ni.Performance evaluation of ground source heat pump system for greenhouse heating in northern China, Biosystems engineering, 111 (1) (2012), pp. 107-117

[8] Wang Meizhi, Liu Jijun, Wu Zhonghong, et al. Effect of ground-source heat pump applied to piggery on energy saving and emission reduction, Transactions of the Chinese Society of Agricultural Engineering, 27 (4) (2011), pp. 251-254

[9] J. Zhao, C.S. Dai, Design and application of ground source heat pump system, China Building Industry Press, Beijing (2009)

[10] People's Bank of China. (2012). The exchange rate between Chinese Yuan and U.S. dollar. http://www.pbc.gov.cn/publish/main/537/index.html.(in Chinese).

[11] M. Barbari, L. Conti, Use of different cooling systems by pregnant sows in experimental pen, Biosystems engineering, 103 (2) (2009), pp. 239-244.

***Corresponding author**
Shaymaa Adnan Abdulrahman,
University of Information Technology & Communications, Business Information Technology Department, Baghdad, Iraq
e-mail: dr.shaymaa.adnan@Uoitc.edu.iq