Prediction of Chronic Kidney Disease using Machine Learning Techniques

Suresh Reddy. G[1*], Raja Vamsi Bandireddy [2], Likhitha Koppula[3], Pravallika Sree Vani Nidadavolu [4], Navya Pagadala[5]

[1] Professor at VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
[2,3,4,5] Student at VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
sureshreddy_g@vnrvjiet.in[1], rajavamsi_b19@vntvjiet.in[2], likhitha_k19@vnrvjiet.in[3], pravallikasreevani_n19@vnrvjiet.in[4], navya_p19@vnrvjiet.in[5]

Abstract. Utilizing machine learning calculations, this study plans to foresee ongoing kidney illness. A condition wherein the kidneys cannot dispose of poisons from the body is known as persistent kidney sickness. It is conceivably the deadliest disorder, and a wrong finding can achieve mortality. Chronic kidney disease (CKD) influences around 800 individuals for every million individuals (pmp), while end-stage renal disease (ESRD) influences 150-200 individuals for every million individuals (pmp). In India, 18,000 to 20,000 patients (or 10% of new ESRD cases) use dialysis. A hemodialysis treatment costs between $15 and $40, and erythropoietin adds $150 to $200 per month. The patient may not exhibit any symptoms during the disease's early stages that would prevent it from becoming chronic. Several data mining and machine learning algorithms can aid the diagnosis of CKD. This study's dataset was obtained from the UCI repository. Expectation strategies included the K-Nearest Neighbor method, Random Forest, Decision Tree, Support Vector Machine, Gradient Boosting, XG boost, Ada boost, and Ensemble. The characteristics of prediction and their correlation were examined from a medical perspective.

Keywords: CKD, ESRD, XG boost, Ada boost, K - Nearest Neighbor method, Support Vector Machine, Gradient Boosting, and, Ensemble.

1 Introduction

The steady decrease in renal capability that in the end prompts the deficiency of all kidney capability is a sign of ongoing kidney sickness. Before all else periods of diligent kidney disorder, secondary effects are honest and vague, which addresses a bit [12]. Sometimes, people learn about their condition when it is in its final stages when there is little chance of recovery. Constant renal sickness positioned seventeenth among worldwide reasons for death as indicated by the 2015 Global Burden of Disease (GBD)

© The Author(s) 2023
C. Kiran Mai et al. (eds.), *Proceedings of the Fourth International Conference on Advances in Computer Engineering and Communication Systems (ICACECS 2023)*, Atlantis Highlights in Computer Sciences 18,
https://doi.org/10.2991/978-94-6463-314-6_13

study (age-normalized yearly passing pace of 192 for every 100,000 population) [1]. This disease has many causes, like diabetes and high blood pressure, and numerous side effects, including headache, vertigo, and frequent urination. In India, chronic renal disease ranks eighth among the leading causes of death, according to GBD 2015 [2]. One survey found that CKD affects one in seven Americans. In India, roughly 87% are connected to hypertension, 37% to diabetes, 22% to CVD (cardiovascular sickness), 6.7% have a background marked by intense renal injury, and 23% have been involved in selective drugs in the past [2]. In its advanced stages, chronic kidney disease is costly to treat and cure. Prediction at an early stage is therefore crucial. Other than the kidneys, other internal organs in the body can be affected by chronic kidney disease, which can also cause other diseases. A proper diagnosis can lead to effective treatment and a healthy lifestyle. This application challenges minimal consciousness of ongoing kidney illness. The disease with the lowest awareness level is chronic kidney disease. Due to the absence of symptoms, people are more likely to be in the terminal stage than others. People's ignorance is exacerbated by the high cost of both diagnosis and treatment.

The goal is to develop an awareness so that people can become mindful of their medical issues and the seriousness of this illness. In this course of the investigation, a variety of calculations, including Logistic Regression, K-Nearest Neighbour, Random Forest, Decision Tree, Support Vector Machine, Gradient Boosting, XG Boost, Ada Boost, and Ensemble, were utilized. Preprocessing has separated the dataset into training and testing sets. By feeding data into our models and evaluating their predictions' accuracy, we have created models. Parameters like age, blood pressure, the number of red and white blood cells, and more are considered. Figure 1 portrays the work process.

2 Literature Review

2.1 Diagnosis of Patients with chronic kidney disease by using two fuzzy classifiers

To analyze CKD patients, the task was to create fluffy classifiers, fluffy rule-building master frameworks, and fluffy ideal-associated memories [3]. The dataset that was utilized was taken from the archive for AI at UCI [3]. They likewise incorporated a couple of composite datasets with different corresponding sounds to dissect the two sorts of fluffy techniques. The preparation and forecast sets were converged in pairs [3] after each mathematical trait was separately exposed to around 11 degrees of the corresponding commotion. A matrix with 121 information classifications was developed to look at order rates.

A while later, the presentation of two fleecy classifiers was evaluated using imitated datasets; For FuRes and Froth, the typical expectation rates were 98.1% + 0.5 percent and 97.2% + 1.2 percent, individually. These two fluffy classifiers are pivotal instruments for precisely anticipating CKD patients [3].

2.2 Diagnosis of chronic kidney disease by using random forest

The quantitative and subjective consequences of this venture were investigated, and the outcomes showed that the irregular woodland classifier played out the best [6]. Random Forest is the sole starting point for this undertaking. Accuracy is achieved when the outcomes are contrasted with the clinical determination. This would be an impediment. Moreover, the tested dataset does not matter in the present.

2.3 Prevalence of chronic kidney disease in China: A cross-sectional survey

To recognize the essential driver of constant kidney sickness, this drive has provided an overview [7]. The discernment that no previous move in China has consolidated an expected glomerular filtration rate (eGFR) and albuminuria is an improvement for this review. A representative trial of Chinese adults was compiled in this outline. An eGFR of less than 60 mL/min/1.73 within seeing albuminuria was described as CKD [4]. The etiology of persistent renal disease was the primary focus of the paper, rather than the development of an information-input-based ML model for its preparation and evaluation. Additionally, China must utilize it [9].

2.4 Incorporating temporal EHR data in Predictive Models for risk stratification of Renal Function Deterioration

Models inherent in this venture utilized transient information from electronic health records (EHRs) [8], which all together work on the administration of constant infections. The results demonstrated the way that recollecting transient information for a patient's clinical history can work on the assumption of renal capacity decline [5]. The way that this drive has depended on well-being records from earlier years and puts less accentuation on persistent kidney infection is a huge downside.

2.5 Prevalence of chronic kidney disease in an adult population

In a grown-up populace reconnaissance program, this drive decided the predominance of persistent kidney illness and the hazard factors for it. 73% of 51-year-elderly people ladies without a background marked by ongoing renal sickness were inspected in a cross-sectional investigation of 600 adults.

Questions were requested from the participants [4]. Over a portion of individuals who participated had diabetes mellitus (DM), hypertension, or weight in their families, and 30% had persistent kidney illness [4].

3 Existing System

Utilizing huge scope CKD information, a brain network-based classifier was developed, and its exactness on test information was 95% [10]. Moreover, the CKD informational index given by the UCI machine learning archive was used in by far most of the past examinations. A pre-owned picture enrollment to recognize changes in renal morphology.

CKD was researched utilizing K-nearest neighbor (KNN), support vector machine (SVM), and the delicate free appearance of class associations. Both KNN and SVM obtained a precision of 95.7% [11]. Moreover, CKD was analyzed utilizing a fluffy rule-building master framework, fluffy ideal cooperative memory, and halfway least squares discriminant investigation with a scope of 95.5% to 97.9% accuracy [12]. The diagnosis of CKD has been improved by their research.

4 Proposed System

Investigation made use of a wide range of machine learning techniques, including K-Nearest Neighbors, Support Vector Machine, Decision Tree, Logistic Regression, Random Forest, Gradient Boosting XG Boost, and AdaBoost [13].

The listed strategies are all used to make, train, and assess a model. There are 400 entries with 21 characteristics in the dataset that were obtained from the UCI repository [14]. Algorithms for classification received this data.

This comprises five modules:

- Data Collection
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on the test set

4.1 Data Collection

It is the first and most crucial stage in the model's construction. The information collection ought to be exact so the model's forecasts are more exact. The dataset has been taken from the repository for machine learning at UCI [14].

It contains 401 patient records. The prediction takes into consideration all 19 attributes. Table 1 contains data concerning patient attributes and the tests used to get them. There is a description of their classification, values, and description. It is possible that the prediction will not use every attribute in Table 1.

Table 1. Information on the Attributes

Attributes	Description	Type of Test	Attribute Type	Attribute Values
age	Age	-	Numerical	years
bp	Blood Pressure	BP_test	Numerical	Mm/Hg
sg	Specific Gravity	Tests_test	Nominal	1.006, 1.011
al	Albumin	Urine_test	Nominal	0, 1, 2, 3,4
su	Sugar	Urine_test	Nominal	0, 1, 2, 3,4
rbe	Red Blood Cells	Urine_test	Nominal	Normal, Abnormal
pe	Pus Cell	Urine_test	Nominal	Normal, Abnormal
pee	Pus Cell Clumps	Urine_test	Nominal	Present, Not Present
ba i	Bacteria	Urine_test	Nominal	Present, Not Present
bgr	Blood Glucose Random	Blood_test	Numerical	Mgs/dl
bu	Blood Urea	Blood_test	Numerical	Mgs/dl
se	Serum Creatinine	Blood_test	Numerical	Mgs/dl
sod	Sodium	Blood_test	Numerical	mEq/l
pot	Potassium	Blood_test	Numerical	mEq/l
hemo	Hemoglobin	Blood_test	Numerical	Gms
pev	Packed Cell Volume	Blood_test	Numerical	%
whe	White Blood Cell Count	Blood_test	Numerical	Cells/emm
re	Red Blood Cell Count	Blood_test	Numerical	millions/ emm
htn	Hypertension	BP_test	Nominal	Yes, no
dm	Diabetes Mellitus	Diabetes_test	Nominal	Yes, no
cad	Coronary Artery Disease	ECG, X-Rays, Blood_test	Nominal	Yes, no
appet	appetite	Doctor's A	Nominal	Good, Poor
pe	Pedal Edma	Oedema Assessment	Nominal	Yes, no
ane	Anemia	CBC(plete blood count)	By inal	Yes, no
class	Classification	-	Nominal	Ckd, No Ckd

In Figure 1, the "class" section contains either 0 or 1 qualities. The patient does not have chronic kidney disease (CKD) if the value is "0." A CKD patient is represented by the value "1."

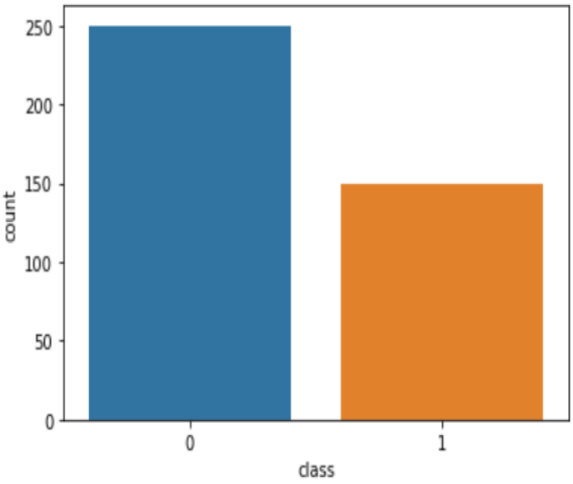

Fig.1. Information on the Attributes

4.2 Data Preprocessing

The arrangement of information contains both numeric and ostensible qualities. The numeric values "0" and "1" are created from the nominal values "Yes" and "No," "Present," and "Not Present." Figure 2's nominal values for pus cells, pus cell lumps, bacteria, and red blood cells are changed to the numerical values 0 and 1 in Figure 3 following preprocessing. To reduce data disturbance, some missing values in red blood cells were removed.

Utilizing capabilities from the Panda library, the clamor causing invalid qualities that impact the model's accuracy is dispensed with. To work on the general viability of the model, the information is separated into preparing and testing portions with an 80/20 proportion.

KNN attribution was utilized to adjust the qualities to the closest number. By computing the Euclidean distance for each example, invalid qualities were disposed of and values were adjusted.

	age	blood_pressure	specific_gravity	albumin	sugar	red_blood_cells	pus_cell	pus_cell_clumps	bacteria	blood_gluco
0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	
1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	
2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	
3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	
4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	
...	...	...	...	...	...	...	...	...	...	
395	55.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	
396	42.0	70.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	
397	12.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	
398	17.0	60.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	
399	58.0	80.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	

400 rows × 25 columns

Fig.2. Before Pre-Processing

age	blood_pressure	specific_gravity	albumin	sugar	red_blood_cells	pus_cell	pus_cell_clumps	bacteria	.
48.0	80.0	1.020	1.0	0.0	0	1	0	0	
7.0	50.0	1.020	4.0	0.0	1	1	0	0	
62.0	80.0	1.010	2.0	3.0	1	1	0	0	
48.0	70.0	1.005	4.0	0.0	1	0	1	0	
51.0	80.0	1.010	2.0	0.0	1	1	0	0	
...	...	...	...	...	...	...	...	...	
55.0	80.0	1.020	0.0	0.0	1	1	0	0	
42.0	70.0	1.025	0.0	0.0	1	1	0	0	
12.0	80.0	1.020	0.0	0.0	1	1	0	0	
17.0	60.0	1.025	0.0	0.0	1	1	0	0	
58.0	80.0	1.025	0.0	0.0	1	1	0	0	

umns

Fig.3. After Pre-Processing

4.3 Feature Selection

The connection between characteristics and class designations is shown by the heat maps' absolute values.

According to Figure 3, the highest correlation (>0.5) exists between hypertension (htn), packed cell volume (PCV), and hemoglobin (hemo). Albumin (al), red blood cells (RBC), pus cells (pc), blood glucose random (bar), red blood cell count (rc), diabetes mellitus (dm), appetite (applet), and pedal edema (pe) all have a relationship that is stronger than 0.30. The connection between the additional elements is closer to 0.30.

When evaluating the characteristics, the correlation and medical perspective were taken into consideration

4.4 Model Training

To improve the model's accuracy and efficiency, the data are divided into 80 and 20 ratios. whereas model training accounts for eighty percent. The model was trained using nine different algorithms was listed in Figure 4, and the one that performed the best during testing was used for prediction

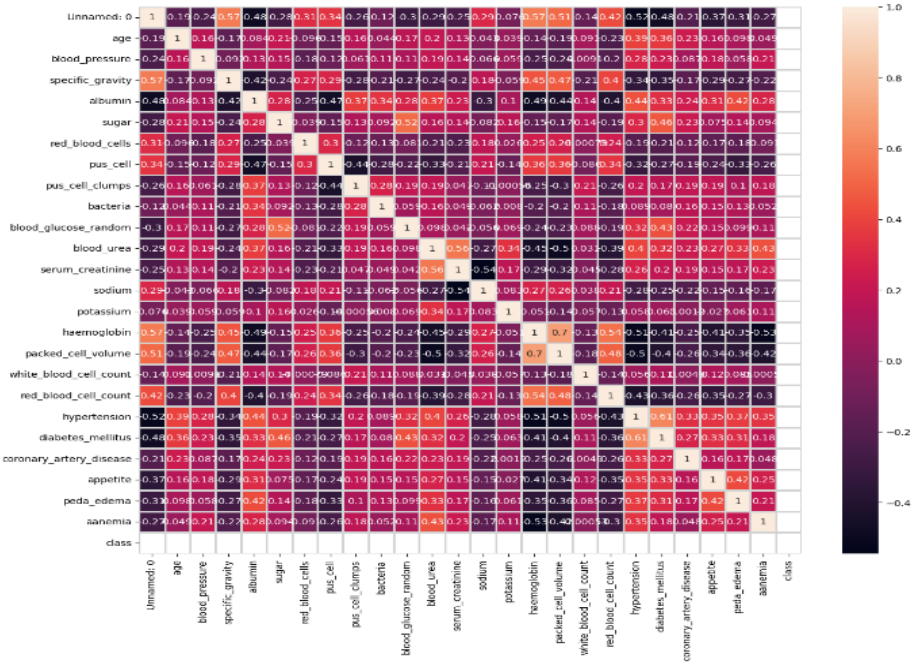

Fig.4. Heat Map (Correlation between attributes)

5 Evaluations

The experimental results showed that Ada boost and the RF classifier provides the greatest precision, with DT, XG Boost, and Gradient Boosting achieving respective levels of accuracy of 98.75%, 98.75%, and 97.50% [15]. We can ascertain which algorithm produced the most accurate disease predictions from Table 2 to 10 [16].

Nearest Neighbour

Training Results
Accuracy Score: 0.6500

Testing results
Accuracy Score: 0.6188

Table 2. Results of K-Nearest Neighbour

K- Nearest Neighbour					
	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	65%	0.0	65%	32.50	42.25
Recall	100%	0.0	65%	50.00	65.00
Fl-score	78.79%	0.0	65%	39.39	51.21
support	52	28	65	80.00	80.00

Logistic Regression

Training Results
Accuracy Score: 0.9250

Testing results
Accuracy Score: 0.9219

Table 3. Results of Logistic Regression

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	96%	86.67%	92.5%	91.33	92.73
Recall	92.30%	92.85%	92.5%	92.58	92.50
Fl-score	94.11%	89.65%	92.5%	91.88	92.55
support	52	28	92.5	80	80

Support Vector Machine

Training Results
Accuracy Score:0.6500

Testing results
Accuracy Score:0.6188

Table 4. Results of Support Vector Machine

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	65.00%	0.0	65%	32.50	42.25
Recall	100%	0.0	65%	50.00	65.00
Fl-score	78.78%	0.0	65%	39.39	51.21
support	52	28	65	80	80

Decision Tree

Training Results
Accuracy Score: 0.987

Testing results
Accuracy Score: 1.0000

Table 5. Results of Decision Tree

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	100%	96.55%	98.75%	98.28	98.79
Recall	98.07%	100%	98.75%	99.03	98.75
Fl-score	99.02%	98.24%	98.75%	98.63	98.75
support	52	28	98.75	80	80

Random Forest

Training Results
Accuracy Score: 1.0000

Testingresults
Accuracy Score:0.9844

Table 6. Results of Random Forest

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	100%	100%	100%	100	100
Recall	100%	100%	100%	100	100
F1-score	100%	100%	100%	100	100
support	52	28	100	80	80

Gradient Boost

Training Results
Accuracy Score: 0.9875

Testing results
Accuracy Score: 1.0000

Table 7. Results of Gradient Boost

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	100%	96.55%	98.75%	98.27	98.79
Recall	98.07%	100%	98.75%	99.03	98.75
Fl-score	99.02%	98.24%	98.75%	98.63	98.75
support	52	28	98.75	80	80

XG Boost

Training Results
Accuracy Score: 0.9875

Testing results
Accuracy Score: 0.9688

Table 8. Results of XG Boost

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	100%	96.55%	98.75%	98.27	98.79
Recall	98.07%	100%	98.75%	99.03	98.75
Fl-score	99.02%	98.24%	98.75%	98.63	98.75
support	52	28	98.75	80	80

Ada Boost

Training Results
Accuracy Score: 0.9875

Testing results
Accuracy Score: 1.0000

Table 9. Results of AdaBoost

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	100%	96.55	98.75%	98.27	98.79
Recall	98.07%	100%	98.75%	99.03	98.75
Fl-score	99.02%	98.24%	98.75%	98.63	98.75
support	52	28	98.75	80	80

Ensemble

Training Results
Accuracy Score: 1.0000

Testing results
Accuracy Score:0.9906

Table 10. Results of Ensemble

	0	1	Accuracy	Macro Avg	Weighted Avg
Precision	100%	100%	100%	100	100
Recall	100%	100%	100%	100	100
Fl-score	100%	100%	100%	100	100
support	52	28	1	80	80

6 Results

Ada boost and the RF classifier provide the greatest precision, according to the experimental results, with DT, XG Boost, and Gradient Boosting achieving respective levels

of accuracy of 98.75%, 98.75%, and 97.50%. We can ascertain which algorithm produced the most accurate disease predictions from Table 11 [17-25].

Table 11. Comparison of Algorithms

S.No	Algorithm	Training Accuracy	Testing Accuracy
1	Support vector Machine	0.6188	0.6500
2	Decision Tree	1.0000	0.9870
3	KNN	0.6188	0.6500
4	Logistic Regression	0.9219	0.9250
5	Random Forest	0.9844	1.0000
6	XG Boost	0.9688	0.9875
7	Ensemble	0.9906	1.0000
8	Gradient Boost	1.0000	0.9875
9	Ada Boost	1.0000	0.9875

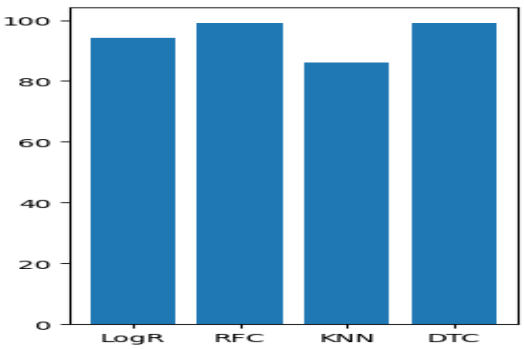

Fig.5. Comparison of Accuracy

The above figure 5 tells which algorithm has the highest accuracy. Additionally, a web application is developed with MySQL as the backend and the Python framework Flask as the front end. Based on basic blood and urine test results, this application helps determine whether a person has CKD.

7 Conclusion

Chronic kidney disease currently is the most widespread disease across the world. Diagnosis in its early stages is crucial but minor symptoms make it a difficult task. In our proposed system we have successfully achieved the highest accuracy with a random classifier. This can help people to detect CKD easily. This project can further be extended as a more viable app. It can also be developed to detect the stages by taking X-rays as input.

References

1. Lalithambika devi B.N, Sanjay Maitra. Assessment of sequential changes occurring in the body fluid composition by bio-impedance analyzer in pre and early post-renal transplant patients, Nephrology Dialysis Transplantation, Volume 35, Issue Supplement_3, June 2020, gfaa142.P1616, https://doi.org/10.1093/ndt/gfaa142.P1616
2. Kumar, Vivek, Ashok Kumar Yadav, Jasmine Sethi, Arpita Ghosh, Manisha Sahay, Narayan Prasad, Santosh Varughese, et al. "The Indian chronic kidney disease (ICKD) study: baseline characteristics." Clinical Kidney Journal 15, no. 1 (2022): 60-69.
3. Magrini, Anna, Stefano Lazzari, Ludovica Marenco, and Giulia Guazzi. "A procedure to evaluate the most suitable integrated solutions for increasing energy performance of the building's envelope, avoiding moisture problems." International Journal of Heat and Technology 35, no. 4 (2017): 689-699.
4. Divya, C. D., H. L. Gururaj, R. Rohan, V. Bhagyalakshmi, H. A. Rashmi, A. Domnick, and Francesco Flammini. "An efficient machine learning approach to nephrology through iris recognition." Discover Artificial Intelligence 1 (2021): 1-15.
5. Liu, Na, Chiyue Ma, Man Xu, Yun Ge, and Dan Gan. "Physiological Status Prediction Based on a Novel Hybrid Intelligent Scheme." Computational Intelligence and Neuroscience 2022 (2022).
6. Subasi, Abdulhamit, Emina Alickovic, and Jasmin Kevric. "Diagnosis of chronic kidney disease by using random forest." In CMBEBIH 2017: Proceedings of the International Conference on Medical and Biological Engineering 2017, pp. 589-594. Springer Singapore, 2017.
7. Zhang, Luxia, Fang Wang, Li Wang, Wenke Wang, Bicheng Liu, Jian Liu, Menghua Chen, et al. "Prevalence of chronic kidney disease in China: a cross-sectional survey." The lancet 379, no. 9818 (2012): 815-822.
8. Singh, Anima, Girish Nadkarni, Omri Gottesman, Stephen B. Ellis, Erwin P. Bottinger, and John V. Guttag. "Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration." Journal of Biomedical Informatics 53 (2015): 220-228.
9. Cueto-Manzano, Alfonso M., Laura Cortés-Sanabria, Héctor R. Martínez-Ramírez, Enrique Rojas-Campos, Benjamin Gómez-Navarro, and Marcelo Castillero-Manzano. "Prevalence of chronic kidney disease in an adult population." Archives of Medical Research 45, no. 6 (2014): 507-513.
10. H. Polat, H. D. Mehr, and A. Cetin. "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods", Journal of Medical Systems., vol. 41, no. 4, p. 55, Apr. 2017.
11. Levey, A. S., R. Atkins, J. Coresh, E. P. Cohen, A. J. Collins, K-U. Eckardt, M. E. Nahas, et al. "Chronic kidney disease as a global public health problem: approaches and initiatives–a position statement from Kidney Disease Improving Global Outcomes." Kidney International 72, no. 3 (2007): 247-259.
12. Jha, Vivekanand, Guillermo Garcia-Garcia, Kunitoshi Iseki, Zuo Li, Saraladevi Naicker, Brett Plattner, Rajiv Saran, Angela Yee-Moon Wang, and Chih-Wei Yang. "Chronic kidney disease: global dimension and perspectives." The Lancet 382, no. 9888 (2013): 260-272.
13. N. R. Hill, S. T. Fatoba, J. L. Oke, et al., "Global prevalence of chronic kidney disease – a systematic review and meta-analysis," PLoS One, vol. 11, no. 7, article e0158765, 201

14. H. Nasri, “World Kidney Day 2014; chronic kidney disease and Aging: a global health alert,” Iranian Journal of Public Health, vol. 43, no. 1, pp. 126-127, 2014.
15. G. Abraham, S. Varughese, T. Thandavan, et al., “Chronic kidney disease hotspots in developing countries in South Asia,” Clinical Kidney Journal, vol. 9, no. 1, pp. 135–141, 2016.
16. K. T. Mills, T. Xu, W. Zhang, et al., “A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010,” Kidney International, vol. 88, no. 5, pp. 950–957, 2015
17. B. Ene-Iordache, N. Perico, B. Bikbov, et al., “Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study,” The Lancet Global Health, vol. 4, no. 5, pp. e307–e319, 2016
18. S. Anand, M. A. Khanam, J. Saquib, et al., “High prevalence of chronic kidney disease in a community survey of urban Bangladeshis: a cross-sectional study,” Glob Health, vol. 10, no. 1, p. 9, 2014
19. L. Ali, K. Fatema, Z. Abedin, et al., “Screening for chronic kidney diseases among an adult population,” Saudi Journal of Kidney Diseases and Transplantation, vol. 24, no. 3, p. 534, 2013.
20. M. J. Hasan, M. A. Kashem, M. H. Rahman, et al., “Prevalence of chronic kidney disease (CKD) and identification of associated risk factors among rural population by mass screening,” Community-Based Medical Journal, vol. 1, pp. 20–26, 2013.
21. M. J. Lysaght, “Maintenance dialysis population dynamics: current trends and long-term implications,” Journal American Society Nephrology, vol. 13, suppl 1, pp. S37–S40, 2002.
22. M. Bakhshayeshkaram, J. Roozbeh, S. T. Heydari, et al., “A population-based study on the prevalence and risk factors of chronic kidney disease in an adult population of Shiraz, southern Iran,” Galen Medical Journal, vol. 8, no. 935, p. 935, 2019.
23. K. U. Eckardt, J. Coresh, O. Devuyst, et al., “Evolving importance of kidney disease: from subspecialty to global health burden,” The Lancet, vol. 382, no. 9887, pp. 158–169, 2013.
24. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities,” IEEE Access, vol. 5, pp. 8869–8879, 2017.
25. V. Mohan, “Decision trees: a comparison of various algorithms for building decision trees,” 2013, http://cs.jhu.edu/ vmohan3/document/ai_dt.pdf.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.