



(RESEARCH ARTICLE)



Machine learning approach using hierarchical decision- tree projection algorithm to predict the presence of chronic kidney disease

J Sarada * and NV Muthu Lakshmi

Department of Computer Science, Sri Padmavathi Mahila Vishvavidyalayam (SPMVV), Tirupati, Andhra Pradesh – India.

Global Journal of Engineering and Technology Advances, 2021, 08(02), 088–095

Publication history: Received on 20 July 2021; revised on 26 August 2021; accepted on 28 August 2021

Article DOI: <https://doi.org/10.30574/gjeta.2021.8.2.0142>

Abstract

Chronic Kidney disease is one of the eminent diseases which is commonly seen the patients which the various ailments which results in the step by step failure of kidneys which may result in the fatality of the human. The Chronic Kidney disease which is precisely called as CKD in medical terms is predicted with various symptoms that evolves in the human body. Strategically predicting the CKD using the machine learning algorithm is the challenging proportion. This paper solves the issues of predicting the CKD using the Hierarchical Decision-Tree Projection Algorithm which takes the various clinical study results of the human body into the dataset format and is algorithmically evaluated. Through this method the whole study in completely evaluated with the various parameters which is obtained from the human study. The variations in the whole study with respect to the parametrical correlation are taken into consideration. The results are obtained from each parametrical evaluation and with the results the prediction and presence of the Chronic Kidney disease is evaluated. The Experimental results show the algorithmic evaluations are showing the comparatively high accuracy and performance.

Keywords: Chronic Kidney Disease; Clinical Evaluation; Hierarchical Decision-Tree Projection Algorithm; Machine Learning

1. Introduction

Chronic kidney ailment incorporates conditions that harm your kidneys and decline their capacity to keep you solid by carrying out the responsibilities recorded. In the event that kidney illness deteriorates, squanders can work to elevated levels in your blood and cause you to feel debilitated. You may create confusions like hypertension, frailty (low blood tally), powerless bones, poor dietary wellbeing and nerve harm. Likewise, kidney malady expands your danger of having heart and vein illness. These issues may happen gradually over an extensive stretch of time. Interminable kidney sickness might be brought about by diabetes, hypertension and different issue. Early recognition and treatment can frequently shield ceaseless kidney infection from deteriorating. At the point when kidney infection advances, it might in the end lead to kidney disappointment, which requires dialysis or a kidney transplant to look after life.

Chronic Kidney Disease is a very dangerous health problem that has been spreading globally due to alterations in life style such as food habits, changes in the atmosphere, etc. So, it is essential to decide on any remedy to avoid and to predict the disease in early stage which helps to avoid wastage of life. We show that feature selection approach is well suited for chronic kidney disease prediction. Principal Component Analysis is one of the feature selection techniques that filters out less important attributes; it also picks attributes of importance from the dataset. We also compare different data classification approaches in terms of how accurately they predict chronic kidney disease.

*Corresponding author: J Sarada

Research Scholar, Department of Computer Science, Sri Padmavathi Mahila Vishvavidyalayam (SPMVV), Tirupati, Andhra Pradesh – India.

We use Hierarchical Algorithm for the grouping and ordering of the similar datasets for the information. Hierarchical clustering, also known as hierarchical cluster analysis, is an algorithm that groups similar objects into groups called clusters. The endpoint is a set of clusters, where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other, Hierarchical clustering starts by treating each observation as a separate cluster. Then, it repeatedly executes the following two steps: (1) identify the two clusters that are closest together, and (2) merge the two most similar clusters. This continues until all the clusters are merged together.

We use Futuristic Enhance able Decision- Tree Projection algorithm to Implement the Machine Learning Concept into the proposes system. Decision-Tree Projection is a model for machine learning. It consists of a Decision-Tree tree at a single level. This tree is a Decision-Tree tree directly linked to each leaf with a root node. Decision-Tree Projection creates a production determined by the value assumed by exactly one input feature. They are also called 1 rule from time to time. As the input function changes, there may be several variations. Decision-Tree Projection are frequently used as components in ensemble techniques for machine learning.

2. Related study

Chronic Kidney Disease is a tranquil condition. Signs and signs of CKD, if present, are ordinarily not unequivocal in nature and not in the least like a couple of other steady ailments, (for instance, congestive cardiovascular breakdown and unending obstructive lung disease), they don't reveal a snippet of data for assurance or earnestness of the condition. Early acknowledgment and treatment can often keep steady kidney disease from deteriorating. CKD is a powerful condition that results in important grimness and mortality. Because of the huge employment the kidneys play in caring for homeostasis, CKD can impact basically every body structure. Early affirmation and intervention of CKD is essential to slow disease development, to keep up close to home fulfilment and improve results. CKD can more over be described as damage to kidney or Glomerular Filtration Rate (GFR) < 60 mL/min/1.73 m² for a very long time or continuously, free of the explanation.

Kidney hurt in kidney related afflictions can be achieved by the closeness of albuminuria, described as egg whites to creatinine extent >30 mg/g in two of three spot pee models. GFR can be surveyed from adjusted serum creatinine and assessing conditions. A bit of the available conditions for GFR include are Modification of Diet in Renal Disease (MDRD) study condition, the Cockcroft- Gault condition, etc., which are giving definite evaluations of GFR in various clinical settings. GFR is generally assessed as renal room of an ideal filtration marker i.e., inulin from plasma. This intentional GFR is considered as the most excellent level yet isn't judicious for step by step clinical use on account of multifaceted nature of the assessment system. Surveying GFR subject to a filtration marker (for the most part serum creatinine) is as of now by and large recognized as a basic test. Five phases and their individual GFR levels are communicated.

Prosperity Informatics is conveying gigantic proportion of data and planning of this delivered tremendous proportion of data makes more open doors for data to be gotten. This got information can improve the organization idea of social protection to patients. The amount of issues develops while dealing with this enormous proportion of data, one among them is the way to analyze this data in a trust worthy way. The principal target of Health Informatics is to use authentic clinical data to improve our appreciation of prescription and clinical practice. As of now a-days, Data mining methodology is gotten together with AI to eliminate disguised models similarly concerning assessment purposes.

Data mining is portrayed as "a strategy of nontrivial extraction of saw, effectively dark and potentially accommodating information from the data set aside in a data set. Clinical data mining has mind boggling potential like researching the covered plans which can be utilized for clinical finding of any disease dataset. There are two strategies to perform data mining, explicitly managed and independent learning. In managed learning, a readiness set is used to learn model boundaries while in independent adjusting no planning set is used. Portrayal is a controlled learning used to discover disguised models from existing clinical data. Portrayal is fundamental for treatment of patients. Game plan is a huge data mining task and the rule inspiration driving gathering is to propose a portrayal work or then again gathering model (called classifier).

The portrayal models can change the data in the information base or dataset to an express class. Game plan advancement strategies include: Guileless Bayes, SMO, K-NN, Support Vector Machine, Radial Premise Function Network, Multi-Layer Perceptron, Logistic Relapse, Decision-Tree Tree, Back Propagation Neural Network, Arbitrary Forest, Ultrasonography (USG) and so forth Feature Decision-Tree is a method to reduce the dimensionality. The standard use of this method is to eliminate little subsets of significant features from the first dataset reliant on evaluation model. Generally, it is used to improve precision.

3. Proposed research work

The Chronic Kidney Disease which mainly affects the Kidney dues to various human body ailments. The Predication of the CKD remained challenging until the machine learning algorithm was introduced into the whole technological study. The Proposed research work is categorized into four main Phases namely 1) Parametrical Evaluation 2) Data Segregation 3) Algorithmic Implementation and 4) Result Generation.

The above mention four phases play a vital role in the prediction and identification of the presence of Chronic Kidney Disease.

3.1. Parametrical Evaluations

This phase constitutes of the importing of the various parametrical value from the human study with clinical test results that are evaluated and obtained from the various tests parameters that are taken from the serum and other body samples. The various parameters that are obtained from the serum and other body samples in which the parameters needed for the evaluation of the chronic kidney disease is listed below with its reference range.

Table 1 Parameters taken for the Study from Human Clinical Evaluations

ParameterName	ReferenceValue
BloodPressure	Yes/No
Hemoglobin	13-16gm/dl
Diabetic	Yes/No
PackedCellVolume	0.40-0.52PCV
PusCells	0-4PVF
PusCellClumps	25-30HPF
Potassium	3.5-5.3mEq/L
WhiteBloodCells	4-11(10^9 /L)
RedBloodCells	4.5-6.5 (10^{12} /L)
Sodium	135-153mEq/L)
Appetite	Ranges
PedalEdema	Yes/No
BloodGlucose Random	90-140GRBS
SerumCreatinine	0.7-1.3mg/dl
BloodUrea	10-50mg/dl
CoronaryArteryDiseae	Yes/No
Albumin	3.4-5.4g/DL
SpecificGravity	1003-1030
Fibrinogen	1.5-4.5g/L
ErythrocyteSedimenta tionRate(ESR)	1-10mm/hr

The above-mentioned parameter values areobtained from the clinical evaluations of the patient. Each and every parameter is evaluated with the ranges from Normal-Mild-High with respect to the results that are obtained with the

clinical tests that are performed. Through this the whole results are taken into consideration with respect to the various obtains.

3.2. Dataset Segregation

This phase obtains all the clinical evaluation from the parametrical constitutes and makes them converted into the data sets as per the need of it into the machine learning concept. Each and every parametrical value is taken into consideration and parsed into the next phase. Before parsing into the algorithmic evaluation, it should be converted in the bind datasets which the machine learning algorithm accepts. For the evaluation few samples are taken into consideration and construction of the dataset is produced in this phase.

3.2.1. Trail-1 of Patient 1

Table 2 Clinical Evaluations of the Patient 1 with respect to various parameters in Human Study

ParameterName	ReferenceValue	DatasetConversionwithY/N
BloodPressure	140/90	Y
Hemoglobin	12.5 gm/dl	Y
Diabetic	Type-II	Y
Packed CellVolume	0.49PCV	N
PusCells	2PVF	N
PusCellClumps	27HPF	N
Potassium	3.2 mEq/L	Y
WhiteBloodCells	5 (10 ⁹ /L)	N
RedBloodCells	4.9(10 ¹² /L)	N
Sodium	140mEq/L)	N
Appetite	Hungry	Y
PedalEdema	Yes	Y
Anemia	Yes	Y
BloodGlucoseRandom	190GRBS	Y
SerumCreatinine	2.0 mg/dl	Y
BloodUrea	56 mg/dl	Y
CoronaryArteryDisease	No	N
Albumin	3.6 g/DL	N
Specific Gravity	1090	Y
Fibrinogen	1.6 g/L	N
ErythrocyteSedimentationRate(ESR)	2 mm/hr	N

The above table constitutes the various test results obtained from the human study of the patient 1. All the result considerations are taken into a single array for the conversion of the algorithmic matrix for the prediction and presence of the chronic kidney disease.

4. Results

$R = \{y, y, y, n, n, n, y, n, n, n, y, y, y, y, n, n, y, n, n\}$

All the above-mentioned results are converted into the 1 and 0 respectively for the Yes and No which is later converted in the resultant matrix for the evaluation

$R = \{1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0\}$

The evaluation matrix is constituted using the results obtained from the above result array R, all the obtained results are segregated into 4x4 matrix for the better understanding of the results

$$R_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} R_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} R_3 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$R_4 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} R_5 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

All the above resultant matrix $R \rightarrow R_1, R_2, R_3, R_4, R_5$ are sent into the algorithmic evaluation for the result generation.

4.1. Algorithmic Implementation

This phase constitutes of the algorithmic implementation of the resultant matrix obtained from the earlier phase.

This phase uses the Hierarchical Decision-Tree Projection Algorithm which is one of the machine learning algorithms which takes the resultant matrix as the input and gets the result for the evaluation.

4.1.1. Algorithm – Hierarchical Decision-Tree Projection Algorithm

Input = R (R₁, R₂ ... R_n)

Output = Value → V

Cycle C = (C₁, C₂, ... C_n)

Result R_R = Abnormality A → 1, Normal B → 0

Value V → Positive (P) / Negative (N)

Initialize

Initialize Cycle C

Input

$R \rightarrow R_1, R_2, \dots, R_n$

$R = (R_1(\text{Value}), R_2(\text{Value}) \dots R_n(\text{Value}))$

$R = (R_1(A/B), R_2(A/B) \dots R_n(A/B))$

$R = (R_1(A/B)(\text{Total}), R_2(A/B)$

$(\text{Total}) \dots R_n(A/B)(\text{Total})) R = V \rightarrow \text{Positive/Negative}$

Result R_R = Abnormality / Normal

End

The above mentioned evaluates the given dataset into various cycles and makes the decision whether the presence of the Chronic Kidney Disease is evaluable. The Resultant Vector R which is considered to be evaluation of the systematic approach on the machine learning principle for the result generation. The Abnormality in the resultant vector shows the result to be considered as the Significant or non-significant in the case of the Chronic Kidney Diseases, through which the accurate predication can be done in the machine learning concept.

4.2. Result Generation

The results are generated using the above-mentioned algorithm to evaluate the presence of the chronic kidney disease

Input= $R(R \rightarrow R_1, R_2, R_3, R_4, R_5)$

Output=Value $\rightarrow V$

CycleC= C_1, C_2, C_3, C_4, C_5

Rector=AbnormalityA $\rightarrow 1$, NormalB $\rightarrow 0$

ValueV \rightarrow Positive(P)/Negative(N)

Initialize

Initialize Cycle C1

Input R₁,

$R \rightarrow R_1$

$R_1 = (1, 1, 1, 0) \rightarrow (A, A, A, B)$

$R_1 = (3(P), 1(N))$

$R_1 = V \rightarrow \text{Positive}$

Input R₂,

$R \rightarrow R_2$

$R_2 = (0, 0, 1, 0) \rightarrow (B, B, A, B)$

$R_2 = (1(P), 3(N))$

$R_2 = V \rightarrow \text{Positive}$

Input R₃,

$R \rightarrow R_3$

$R_3 = (0, 0, 1, 1) \rightarrow (B, B, A, A)$

$R_3 = (2(P), 2(N))$

$R_3 = V \rightarrow \text{Positive}$

Input R₄,

$$R \rightarrow R_4$$

$$R_4 = (1, 1, 1, 0) \rightarrow (A, A, A, B)$$

$$R_4 = (3(P), 1(N))$$

$$R_4 = V \rightarrow \text{Positive}$$

Input R_5 ,

$$R \rightarrow R_5$$

$$R_5 = (0, 1, 0, 0) \rightarrow (B, A, B, B)$$

$$R_5 = (1(P), 3(N))$$

$$R_5 = V \rightarrow \text{Positive}$$

Resultant Vector $R_R = \text{Abnormality}$

The Abnormality detected in the given input of the Human Study regarding the Clinical Trials.

$$R = R_1 \rightarrow V(\text{Positive}), R_2 \rightarrow V(\text{Positive}), R_3 \rightarrow V(\text{Positive}), R_4 \rightarrow V(\text{Positive}), R_5 \rightarrow V(\text{Positive})$$

While Calculating the Patient record with the proposed Hierarchical Decision-Tree Projection Algorithm the above-mentioned results are obtained which shows the presence of Chronic Kidney Disease.

5. Conclusion

This paper evaluates the Prediction and Presence of the Chronic Kidney Disease (CKD) using the Hierarchical Decision-Tree Projection Algorithm through which the exact presence and accurate prediction of the presence of Chronic Kidney Disease is evaluated. The Hierarchical Decision-Tree Projection algorithm is the Enhanced version of the Hierarchical Decision Stump Algorithm which resembles better and more accuracy when compared with the existing algorithm. The theoretical evaluation clears the accuracy of the proposed algorithm with giving the better accuracy when compared with other proposed algorithms on the chronic kidney disease.

Compliance with ethical standards

Acknowledgments

I would like to thank Dr. NV Muthu Lakshmi, Assistant Professor, who is research supervisor for my Doctoral Research and Co-author of this article, Department of Computer Science, Sri Padmavathi Mahila Vishvavidyalayam (SPMVV), Tirupati, Andhra Pradesh – India for her moral support, and encouragement to carry out this research study.

Disclosure of conflict of interest

Authors would like to declare that there is no conflict of interest relevant to this article.

References

- [1] Noble A, Amerasinghe, P, Manthrilake H, Arasalingam S. Review of literature on chronic kidney disease of unknown etiology (CKDu) in Sri Lanka. International Water Management Institute (IWMI), 2014; 158: 41.
- [2] Abbas TF, Raheem OA, Abbas AN. Surface Fitting and Representation by Using 2D Least Squares Method in CAD Applications. Engineering & Technology Journal, 2011; 1325-1437.

- [3] Weaver VM., Fadrowski J.J. @ Jaar BG. Global dimensions of chronic kidney disease of unknown etiology (CKDu): a modern era environmental and/or occupational nephropathy? *BMC nephrology*, 2015; 16(01): 01-08.
- [4] Boor P., Ostendorf T, Floege, J., 2010 Renal fibrosis: novel insights into mechanisms and therapeutic targets. *Nature Reviews Nephrology*, 2010; 06(11): 643-656.
- [5] Kalyan K, et al. Artificial Neural Network Application in the Diagnosis of Disease Conditions with Liver Ultrasound Images. Hindawi Publishing Corporation - *Advances in Bioinformatics*, 2014; 708279(01): 01-14.
- [6] Abe C, KAHN JR CE, Doi K, Katsuragawa S. Computer Aided detection of diffuse liver disease. *Investigative radiology*, 1992; 27(01): 71-77.
- [7] Solomon C, Breckon T. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. 01 ed. West Sussex: John Wiley & Sons Ltd., 2011.
- [8] Albrechtsen F. Statistical texture measures computed from gray level concurrence matrices. *Image processing laboratory, department of informatics, university of Oslo*, 2008; 05(01): 01-14.
- [9] Vicas C, et al. Liver Fibrosis detection by the means of texture analysis. Limitations and further development directions. *Automat. Comput. Appl. Math.*, 2010; 19(01): 397-402.
- [10] NagannaChetty, Kunwar Singh Vaisla and Sithu D, Sudarsan. Role of attributes selection in classification of Chronic Kidney Disease patients”, *International Conference on Computing, Communication and Security (ICCCS)*, 2015; 4-5: 1-6.
- [11] L. JerlinRubini, Dr P Eswaran. Generating comparative analysis of early stage prediction of Chronic Kidney Disease, *International Journal of Modern Engineering Research (IJMER)*, July 2015; 5(7): 49-55.
- [12] Parul Sharma, Poonam Sinha. Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM”, *International Journal of Engineering Research & Technology (IJERT)*. December 2015; 4(12): 608-612.
- [13] S Vijayarani, S Dhayanand. Data Mining Classification Algorithms for Kidney Disease Prediction”, *International Journal on Cybernetics & Informatics (IJCI)*, August 2015; 4(4): 13- 25.
- [14] RenukaMarutirao Pujari and Vikas D Hajare. Analysis of Ultrasound Images for Identification of Chronic Kidney Disease Stages, *First International Conference on Networks & Soft Computing*. 19-20 Aug 2014; 380-383.
- [15] Abeer Y Al-Hyari. *Chronic Kidney Disease Prediction System Using Classifying Data Mining Techniques*, library of university of Jordan, 2012.
- [16] S Ramya, N Radha. Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms, *International Journal of Innovative Research in Computer and Communication Engineering*, January 2016; 4(1): 813-820.
- [17] Mohammed Siyad B, Manoj M. Fused Features Classification for the Effective Prediction of Chronic Kidney Disease, *International Journal for Innovative Research in Science & Technology*, March 2016; 2(10): 44-48.
- [18] L Ladha, T Deepa. Feature Selection Methods and Algorithms, *International Journal on Computer Science and Engineering*, May 2011; 3(5): 1787-1797.
- [19] Jiliang Tang, Salem Alelyani and Huan Liu, *Feature Selection for Classification: A Review*.
- [20] Yvan Saeys, Inaki Inza and Pedro Larranaga, A review of feature selection techniques in bioinformatics, *Bioinformatics*, August 2007; 23(19): 2507-2517.