Prediction of Chronic Kidney Disease Using C4.5 Algorithm

M. Praveena, N. Bhavana

Abstract— Chronic Kidney Disease (CKD) is a major medical problem and can be cured if treated it in the early stages. Usually, people are not aware that medical tests we take for different purposes could contain valuable information concerning kidney diseases. Consequently, attributes of various medical tests are investigated to distinguish which attributes may contain helpful information about the disease. The objective of this paper is to make use of such attributes. The information says that it helps us to measure the severity of the problem, the predicted survival of the patient after the illness, the pattern of the disease and work for curing the disease. Hence we considered a data-set with different attributes that can be found in general medical tests, machine learning is applied by developing a decision tree using the C4.5 algorithm and predicted whether the person is normal or suffering from kidney problem. This proposed model will be developed using Java language and is implemented in Net-Beans platform.

Keywords: Chronic Kidney Disease (CKD), Decision Tree, C4.5 Algorithm, Machine Learning.

1. INTRODUCTION

Chronic kidney disease is otherwise called as chronic renal disease. It may show symptoms while developing. However, if you are suffering from tiredness, feeling less energetic, suffering from embarrassment, weak bones, nerve damage, swollen legs and ankles during the night may give you a chance of suffering from a kidney disease. Age isn't a matter for a chronic kidney disease for cause. However, the risk of getting prone to the disease might be higher if a person is diabetic, has high blood pressure or if anyone in the family has a chronic kidney disease. The best way to check the kidney functioning is to use the GFR-glomerular filter. The earlier detection of a kidney disease helps in the early cure of that particular disease which is why we have chosen prediction techniques of machine learning i.e. Decision Trees.

2. LITERATURE SURVEY

- Nishanth Anandanadarajah et al. To detect chronic kidney disease they used (LDA) Linear Discriminant Analysis and (CSP) common spatial pattern filter.
- Radha et al. Different classification techniques are applied over the patient record available and proved that radial basis function works better
- Uma N Dulhare, Mohammad Ayesha Naïve Bayesian and k-nearest neighbour algorithms have been used to predict the disease. They ended up with k-nearest neighbour showing more accuracy than naïve Bayesian.
- Hajar Mousannif et al. Datasets are being used to store medical record. They used support vector machine and Bayesian network to predict kidney disease and select the efficient one among them.
- Parul S, Poonam S SVM and KNN classifiers are compared based on their accuracy and execution time for CKD prediction and proved KNN classifier is better.

3. PROPOSED-MODEL

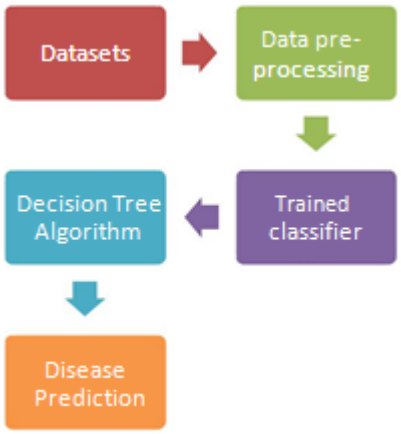

Fig 1

Fig 1 shows the flow of the work done. We considered data-set with patient data and pre-processed it which includes cleaning of the data i.e. noisy data removal, filling up the empty blocks. As we considered C4.5 algorithm all the cells need not be filled in the dataset as the algorithm works even for discrete values and this algorithm is advanced than the ID3 algorithm which works only for continuous values. After pre-processing is done we train the data. The trained data is used to construct a decision tree showing the result whether the patient is normal or affected by the disease. For each input give decision tree generated will be displayed as output.

4. WORK FLOW

The process begins with the gathering of raw data, which is processed and classified. This is followed by data pre-processing prior to sampling. The data for the sample is

Revised Version Manuscript Received on March 08, 2019.

M. Praveena, Assistant professor, Department of computer science and engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, Andhra Pradesh, India. (E-Mail: praveena.mandapati@kluniversity.in)

N. Bhavana, Student, Department of computer science and engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, Andhra Pradesh, India. (E-Mail: pnallamala1@gmail.com)

then generated and is split into training and test datasets. This training data-set is pre-processed again and is trained in the learning algorithm. Optimization of the data and its parameters followed by data post-processing. All of these sampling steps are grouped under validation and processing. Finally, the final model evaluation of the test data set is performed and new data are classified using the classification model.

Data entered will be processed and a decision tree will be generated by calculating entropy and information gain values as per the rules of c4.5 algorithm. From root node to the leaf node, prioritizing of the node to be placed depends on the homogeneity of the node and also the calculated values. In decision tree, the prediction usually occurs at the leaf node. Here performance analysis of CKD and NCKD will be displayed in terms of graph.

5. DATA DESCRIPTION & RESULT

The information used in this paper was obtained from UCI source and comprises of the data samples with 24 attributes, most of them are clinical and the rest are physiological. Some of the attributes are numerical and some are nominal.

Attribute	Representation of the attribute
Age	Age
Blood pressure	Bp
Specific gravity	Sg
Albumin	Al
Sugar	Su
Red blood cells	Rbc
Pus cell	Pc
Pus cell clumps	Pcc
Bacteria	Ba
Blood glucose random	Bgr
Blood urea	Bu
Serum creatinin	Sc
sodium	Sod
Potassium	Pot
Haemoglobin	Hemo
Packed cell volume	Pcv
White blood cell count	Wc
Red blood cell count	Rc
Hypertension	Htn
Diabetes mellitus	Dm
Coronary artery disease	Cad
Appetite	Appet
Pedal edema	Pe
Anemia	Ane

Information of the attributes:

Numerical	Age
	Blood pressure
	random Blood glucose
	Blood urea
	Serum creatinin
	Sodium
	Potassium
	Haemoglobin
	Packed cell volume
	White blood cell count
	Red blood cell count

Nominal	Specific gravity
	Albumin
	Sugar
	Red blood cells
	Pus cell
	Pus cell clumps
	Bacteria
	Hypertension
	Diabetes mellitus
	Coronary artery disease
	Appetite
	Pedal edema
	Anemia

6. RESULTS

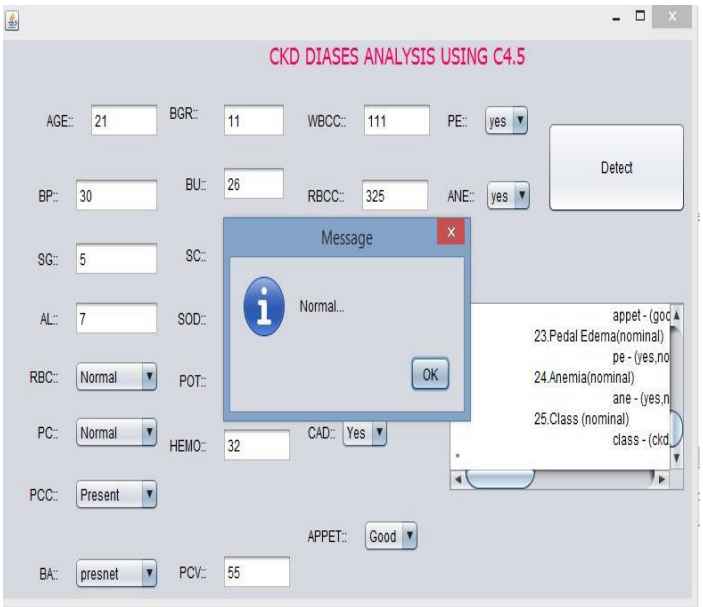

Fig 2: Disease Prediction

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

This output shows that the person is not affected with chronic disease.

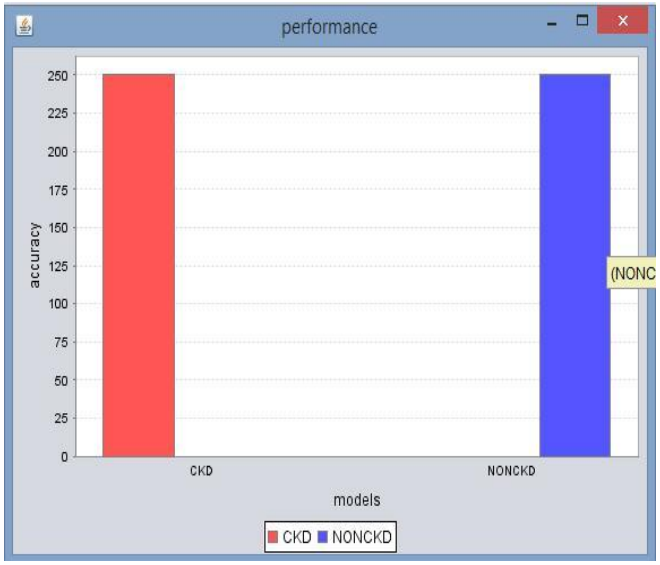

Fig3: Performance Analysis

7. CONCLUSION

The use of machine learning techniques for predictive health analysis works very well. Since it gives us the power to predict diseases in earlier stages, it saves the lives of people by anticipating cures. In this work, we used the c4.5 learning algorithm to predict patients with chronic kidney failure (ckd) disease and patients who do not (notckd) suffer from the disease. The anticipation of diseases remains a major medical challenge and urges us to increase our efforts to develop more machine learning algorithms to intelligently exploit information and extract the best knowledge.

REFERENCES

1. Di Sciascio, David N, Binetti Giulio et al. An end stage kidney disease predictor based on an artificial neural networks ensemble. Expert Syt Appl. Issue 11: 4438-4445 (2013)
2. Arora M, Sharma EA. Chronic kidney disease detection by analysing medical datasets in Weka. Int. Journal. Comput. Appl; vol-6: 20-26, Aug 2016.
3. Tharmarajah Thiruvaran et al., IEEE Identifying important attributes for early detection of chronic kidney disease.
4. Verhaar MC, knepper MA et al. Exosomes and the kidney: prospects for diagnosis and therapy of renal diseases. Kidney Int. 80: 1138-1145 Aug 2011.
5. Shuo Yang, Ran Wei et al. Semantic inference on clinical documents: Combining Machine learning algorithms with an inference engine for effective clinical diagnosis and treatment (2017)
6. Mirja k, Samoylenko A et al. Exosomes as renal inductive signals in health and disease, and their application as diagnostic markers and therapeutic agents. Front. Cell Dev. Biol. 2015 Oct.
7. Ling Yu, Duke Billie J. et al. Exosomal Gapdh from proximal tubule cells regulate ENaC activity Nov 2016
8. GuillermoGarcia-Garcia, kunitoshi Iseki et al. Chronic kidney disease: global dimension and perspectives. May 2013.
9. Meenambal S. et al. Velocity bounded Boolean particle swarm optimization for improved feature selection in liver and kidney disease diagnosis. Expert Syst. Appl. 28-47 (2016).
10. J Stankovic, Salekin A Detection of chronic kidney disease and selecting important predictive attributes in healthcare informatics (ICHI), IEEE, pp.262-270, oct 2016.
11. Parul S and Poonam S Comparative Study of Chronic kidney disease prediction using KNN and SVM. Int. Journ. IJERT Dec2015.
12. Mark hall. The WEKA data mining soft: an update
13. Lichman et al. UCI ML repository
14. Fatima M, P Maruf survey of ML algorithms for disease diagnostic. (2017).

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication