Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766241

CHRONIC KIDNEY DISEASE PREDICTION AND DIAGNOSIS USING DIFFERENT MACHINE LEARNING ALGORITHMS

Mr. Anandkumar A. Sutariya[1] and Dr. Dushyantsinh B. Rathod[2]
[1] PhD Scholar, Computer Engineering, Sankalchand Patel College of Engineering,
Sankalchand Patel University, Visnagar, Gujarat, India
asutariya82@gmail.com
[2] Professor & HoD, Computer Engineering, Ahmedabad Institute of Technology,
Ahmedabad, Gujarat, India
Dushyantsinh.rathod@gmail.com

Abstract. Chronic Kidney Failure is the medical term for chronic kidney dis-ease. It portrays the moderate disintegration of renal disappointment and how, assuming that constant kidney infection has advanced to a high level stage, a high volume of fluid and undesirable electrolytes may develop in the body. We may see less evidence of chronic renal disease in the early phases. The treatment for chronic kidney disease focuses on slowing down the pro-cess of kidney damage. Without a trace of dialysis or kidney migration, per-sistent renal sickness can advance to the last periods of kidney annihilation, which is inoperable. The focal point of this examination is on early discovery of constant obstructive pneumonia illness utilizing different AI techniques, which are K-Nearest Neighbor, Decision Tree and Bayesian Classifier.
Keywords: Chronic Kidney Disease, K-Nearest Neighbor, Decision Tree, and Bayesian Classifier, AI techniques.

1 Introduction

Kidney disease is not a kidding medical problem that makes a ton of terrible things end up peopling, particularly in low-and center pay countries, where a great many individuals bite the dust consistently inferable from an absence of therapy. The le-thality of each persistent sickness is corresponding to the stage it has reached with-out being restored. Diabetic patients are turning out to be more normal, as are hyper-tension, coronary illness, diabetes, and a family background of renal disappointment. On the off chance that disease goes unrecognized and untreated, it can prompt hyper-tension and, in the most dire outcome imaginable, renal disappointment. We focus on CKD, which can help patients in an assortment of ways whenever identified early and suitably. It works on the possibilities of a fruitful treatment while additionally extending the patient's life. The objective of this paper is to utilize a few AI calcula-tions to early analyze constant obstructive aspiratory infection.

Classification is a type of data analysis in which models defining relevant data classes are extracted.

- K-Nearest Neighbor (K-NN): It is a simple strategy that employs classification and regression algorithms. Whether k-NN is used for classification or regression, the output of this method varies.
- Decision Tree (DT): An information base choice tree is a tree-like design of in-formation. It is utilized in tasks exploration and AI to pursue choices that will prompt

a significant end, as well as in information mining to characterize infor-mation and recover information.

- Bayesian Classifier: It is based on the probability theorem and can be used to perform medical diagnosis in a rational manner, especially in automated medical diagnosis decision support systems. It can deal with an unlimited number of inde-pendent variables, both continuous and categorical.

The paper is divided into several sections. The research works relevant to this study are discussed in Section II. The research objectives were outlined in Section III. The findings of all approaches are shown in Section IV. Section V brings the research to a close, while Section VI discusses the research's future directions.

2 Literature Survey

AI calculations have for quite some time been utilized to settle difficulties in the clinical field. This has been endeavored by various scientists. For the order and ex-pectation of the patient's disease status, different procedures and techniques have been utilized. They fostered a choice emotionally supportive network that utilizes characterization calculations to analyze and foresee ongoing renal disappointment.

J. Snegha et al. [1] tried two information mining strategies: irregular backwoods (RF) and back spread calculation. Various models are fabricated utilizing a CKD dataset acquired from www.kaggle.com, and the exhibition of these techniques is contrasted all together with figure out which is the best in foreseeing constant kid-ney illness. As indicated by the aftereffects of their investigations, the back engen-dering calculation has a conviction of 98.40 percent, contrasted with RF's 88.7 per-cent.

R. Gupta et al. [2] utilized three AI techniques: choice tree (DT), irregular woods (RF), and strategic relapse (LR). These models are assembled utilizing a CKD da-taset acquired from the UCI AI store, and their exhibition is assessed to decide the best classifier for anticipating persistent kidney sickness. As per the preliminary information, the LR classifier has the best exactness of 99.24 percent and the most noteworthy review of 100%. Moreover, subsequent to preparing the dataset, DT has the greatest accuracy of 100%.

In order to diagnose chronic renal illness, S. Vashisth et al. [3], used multi-layer perceptron, support vector machine (SVM), and nave bayes classifiers. He used a dataset taken from Apollo hospitals across India, and the testing results suggest that the multi-layer perceptron had the best accuracy of all, at 92.5 percent.

J. R. Lambert et al. [4] proposed a Correlation-based highlight choice - consecu-tive least streamlining (CFS-SMO) and Ranker-successive least advancement (Ranker-SMO) for relative examination of numerical and ostensible traits of con-stant kidney infection with Special clinical dataset for kidney disease, between 2 years to 83 years old, was acknowledged and distributed at the UCI AI archive. The CFS-SMO beats other Ranker-SMOs concerning results. With CFS-SMO, ostensible traits are utilized to further develop order.

P. Arulanthu and E. Perumal [5] has gathered information from Apollo Hospitals and given it to the UCI information vault, where it will be utilized as a preparation dataset for driving order calculations. JRip, Naïve Bayes, Sequential Minimal Opti-mization, and Instance Based Learner are the four calculations. While contrasting the presentation of the JRip and the exhibition of the other calculation, the presenta-tion outline plainly shows that the JRip performs better (for example precision of 98.8%).

3 Proposed Architecture

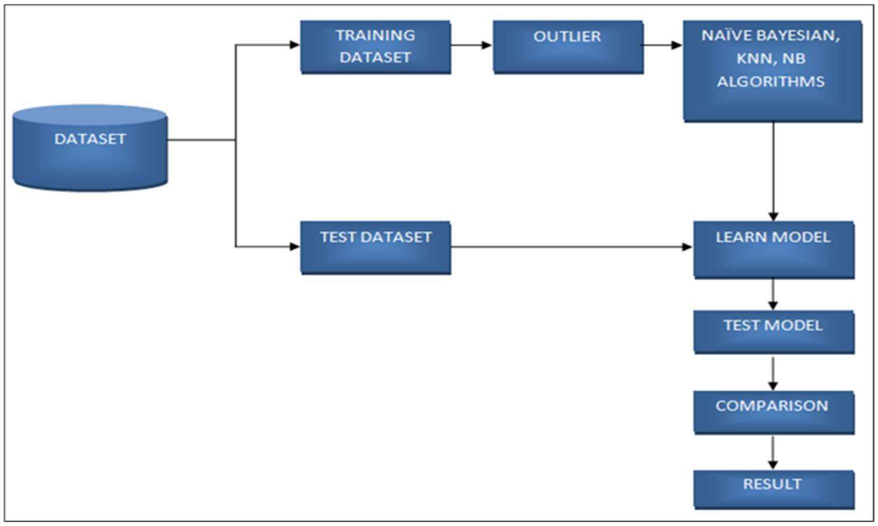

Fig. 1. Proposed Architecture

The essential objective of this study is to exhibit the worth of information mining in surveying way of life related messes. It is tried to audit the writing piece whose ex-amination action is centered on the two specialists and patients. The fundamental focuses (disease, technique, discoveries, precision) of the different review studies will be featured, as well as the use of instruments or strategies. At last, the objective is to recognize what locales request additional consideration from information min-ing and AI devices.

The examination objectives are to sort material as per conduct informatics and group information to dissect information designs. By surveying demonstrative data with regulated and solo AI calculations, we desire to work on the symptomatic exe-cution of present indicative methodologies for infection forecast. To survey the pro-posed approach's presentation utilizing models, for example, accuracy, review, F-measure, and precision with characterization rate. On various datasets, analyze the exhibition of various classifiers and bunching calculations.

4 Results

Fig. 2 shows the precision, recall, F1, and support scores for the proposed model in the classification report for the algorithm decision tree.

	precision	recall	f1-score	support
0	1.00	0.96	0.98	84
1	0.94	1.00	0.97	48
accuracy			0.98	132
macro avg	0.97	0.98	0.98	132
weighted avg	0.98	0.98	0.98	132

Fig. 2. Classification Report of Decision Tree (Precision, Recall, F1-score)

Accuracy suggests a classifier's ability to do whatever it takes not to name a neg-ative event as certain. Not entirely settled as the extent of authentic up-sides of how many certifiable up-sides and deceiving up-sides for each class. It's generally called positive assumption precision.

The constraint of a classifier to observe every one of the specific cases is known as review. Still up in the air as the extent of certifiable empowering focuses on how many veritable up-sides and misdirecting negatives for each class. It decides the degree of really perceived upsides.

The F1 score is a weighted consonant mean of precision and survey, with 1.0 be-ing the most imperative and 0.0 being the least. They are less careful than accuracy assessments since exactness and survey are considered along with the calculation. To dissect classifier models, utilize the weighted ordinary of F1 rather than overall accuracy as a rule.

The amount of genuine occasions of the class in the given dataset is known as help. Support doesn't shift between models; rather, it dissects the evaluation connec-tion.

As addressed in the confusing organization of the decision tree in Fig. 3, there are four methods for managing to choose if the gauges are correct or not.

- True Negative (TN): the case unendingly was projected to be negative.
- True Positive (TP): the case was positive and should be positive.
- FN (False Negative): the case was positive, but the outcome was projected to be negative.
- False Positive (FP): the case was negative, at this point, it was expected to be positive.

The confusion matrix reveals not just a predictive model's performance, but also which classes are successfully predicted, which are incorrectly forecasted, and what types of errors are being made.

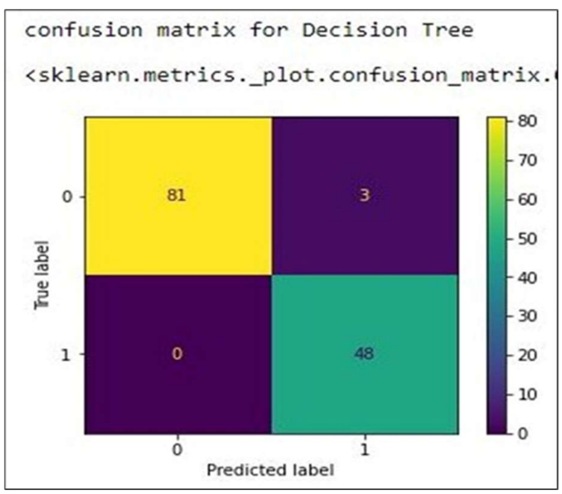

Fig. 3. Decision Tree Confusion Matrix

At the point when we utilize the term precision, we typically suggest exactness. The quantity of right expectations partitioned by the all-out number of information tests is the proportion. We considered the precision of each of the three calculations in this estimation (for example k-NN, Naïve Bayes and DT) are displayed in Fig. 4.

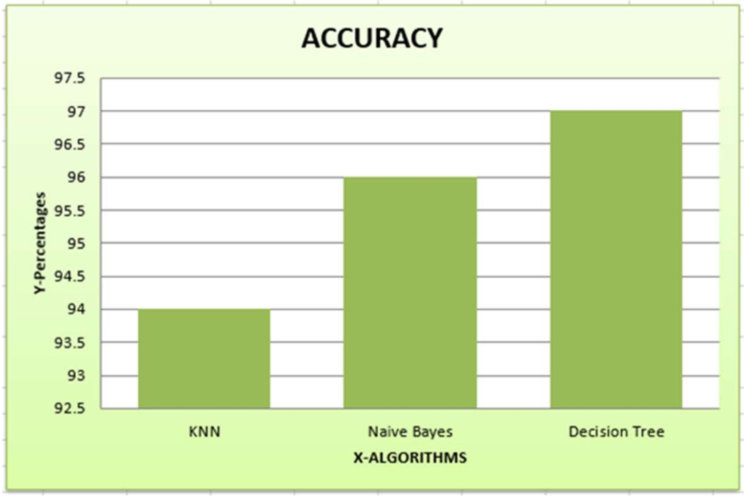

Fig. 4. Accuracy Chart of Machine learning algorithms

In the wake of ascertaining the presentation of proposed models and looking at them all, the best classifier to anticipate Chronic Obstructive Pulmonary Disease was picked. As per the exploratory information, the Decision Tree strategy has the greatest exactness of 97%, contrasted with 96% and 94 percent for the Nave Bayes and k-NN calculations, individually. The outcomes are displayed in Fig. 5.

Algorithms	Accuracy %
KNN	94
Naive Bayes	96
Decision Tree	97

Fig. 5. Accuracy Table of Machine learning algorithms

5 Conclusion

The reason for this review is to foster an original structure in light of bunching and order information digging methods for foreseeing and diagnosing these problems in the medical services region utilizing genomic data sets.

6 Future Direction

The recommended approach may also be applicable to other diseases classification challenges, including datasets of the same sort as those utilized in this study, accord-ing to a review of numerous literature papers. However, there is still much work to be done in terms of doing research on clustering, noise removal, and fuzzy rule-based illness diagnosis algorithms in order to fully harness their potential and utili-ty. In the future, the datasets for disease categorization and prediction utilizing in-cremental machine learning algorithms will demand greater attention. As a result, it is necessary to test this method on additional datasets, particularly large datasets, in order to demonstrate its efficacy in terms of large data computing time. In addition, the study looks into how the proposed technology may be adapted to work with vari-ous types of medical datasets.

References

1. J. Snegha, V. Tharani, S. D. Preetha, R. Charanya and S. Bhavani, "Chronic Kid-ney Disease Prediction Using Data Mining," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vel-lore, India, 2020, pp. 1-5, doi: 10.1109/ic-ETITE47903.2020.482.
2. R. Gupta, N. Koli, N. Mahor and N. Tejashri, "Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154147.
3. S. Vashisth, I. Dhall and S. Saraswat, "Chronic Kidney Disease (CKD) Diagnosis using Multi-Layer Perceptron Classifier," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 346-350, doi: 10.1109/Confluence47617.2020.9058178.
4. J. R. Lambert, P. Arulanthu and E. Perumal, "Identification of Nominal Attrib-utes for Intelligent Classification of Chronic Kidney Disease using Optimization Algorithm," 2020 International Conference on Communication and Signal Pro-cessing (ICCSP), Chennai, India, 2020, pp. 0119-0125, doi: 10.1109/ICCSP48568.2020.9182206.
5. P. Arulanthu and E. Perumal, "Predicting the Chronic Kidney Disease using Var-ious Classifiers," 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysuru, India, 2019, pp. 70-75, doi: 10.1109/ICEECCOT46775.2019.9114653.