International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

CHRONIC KIDNEY DIESEASE PREDICTION USING MACHINE LEARNING

Mr. Suresh R[1], Sudha R[2], Shanthini R[3], Madhumetha T[4]

[1]*Assistant Professor, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry-605107*
[2,3,4]*UG Students, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry-605107*

Abstract - *The Machine Learning are regularly used in medical services industry to "mined"clinical information to get covered up data for intelligent and affective deciding.Revelation of hidden patterns and relationships frequently goes unblemished, yet progressed Machine Learning methods are regularly useful as solution for the current situation. The fundamental objective of this project is to predict the Chronic Renal Disease (IPCRD). Information incorporate blood, urine test, and external symptoms are applied to foresee chronic renal disease. Information from the database is at first changed to Python (3.8) and Chi-Square technique is utilized for features selection. After regularizing data three classifiers were applied and efficiency of yield is. Mainly, three classifiers are analyzed: Yolo Classification, Yolo, Yolo Classification algorithm. Each method has its extraordinary strength in achieving the reason for characterized mining objectives. Machine Learning are regularly used in medical services industry to "mined"clinical information to get covered up data for intelligent and affective deciding.Revelation of hidden patterns and relationships frequently goes unblemished, yet progressed Machine Learning methods are regularly useful as solution for the current situation. The fundamental objective of this project is to predict the Chronic Renal Disease (IPCRD). Information incorporate blood, urine test, and external symptoms are applied to foresee chronic renal disease. Information from the database is at first changed to Python (3.8) and Chi-Square technique is utilized for features selection. After regularizing data three classifiers were applied and efficiency of yield is. Mainly, three classifiers are analyzed: Yolo Classification, Yolo, Yolo Classification algorithm. Each method has its extraordinary strength in achieving the reason for characterized mining objectives.*

yolo classification's efficiency has demonstrated a near edge over others. Further sensitivity and specificity tests are utilized as statistical measures to look at the presentation of a parallel characterization. Sensitivity estimates the portion of actual positives which are accurately recognized while Specificity measures the portion of negatives which are effectively recognized. Machine Learning approach is applied to construct the mining models. It comprises of six significant models: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

Key Words*: *Chronic Kidney Disease, Yolo classification, Chi square method.

1. INTRODUCTION

Chronic renal disorder is turning into a developing issue in both developed and developing world. As indicated by CNN report, in developing countries, because of expanding urbanization, people are adopting undesirable way of life style that advance diabetes and high blood pressure, the main clarification for Chronic Renal disease, while 10-20% of people with diabetes die of renal disorder. On the opposite side, in developed countries, similar to USA, 26 million American adults (1 in 9) have CKD and numerous others are at increased danger. US specialist have come up with an eight point risk factor agenda to foresee chronic renal disorder which incorporates more older age, anemia, female sex, hypertension, diabetes, peripheral vascular disease and any set of experiences of congestive coronary failure or disorder. within the beginning, renal failure could likewise be asymptomatic (not creating any side effects). As kidney function diminishes, the symptoms are identified with the inability to regulate water and electrolyte balances, to clear waste items from the body, and to showcase red blood corpuscle production. Laziness, weakness, shortness of breath, and generalized swelling may occur. Unseen or untreated, perilous conditions can develop. Generalized weakness are often due to anemia, a decreased red blood cell count, because lower levels of erythropoietin don't adequately stimulate the bone marrow. A decrease in red cells equals a decrease in oxygen-carrying capacity of the blood, leading to decreased oxygen delivery to cells for them to try to work; therefore, the body tires quickly. As well, with less oxygen, cells more readily use anaerobic metabolism (an=without+ aerobic=oxygen) resulting in increased amounts of acid production that can't be addressed by the already failing kidneys.

2. RELATED WORK

In this idea [1], the huge amount of knowledge collected by healthcare sector are often effective for analysis, diagnosis and deciding if it's mined properly. the precise techniques must be executed to accomplish better consequence. during this manuscript the potential of the classification of Support Vector Machine, Decision tree, Naïve Bayes and K-Nearest Neighbor algorithm, in analyzing the Chronic renal Disorder dataset collected from UCI repository. Data set has been inspected in terms of accuracy, Root Mean Squared Error, Mean Absolute Error and Receiver Operating distinctive function. Ranking algorithm provides basic improvement in classifications with proper number of attributes. 15 proves to be the atomic number for choosing attributes forthe given dataset resulting highest percent of improvement in accuracy.

In the idea focused in [2], the main motive ofthis work is to determine the existence of chronic kidney disease by imposing various classification algorithms on the patient medical history. This research work is primarily targeting on finding the simple suitable classification algorithm which may be used for the diagnosis of CKD supported the classification report and performance factors. A comparative analysis iscompleted on different algorithms like support vector machine (SVM), random forest (RF), XG Boost (XGB), logistic regression (LR), naive Bayes (NB) classifier to see the classification accuracy for diagnosing CKD. The results show that Random Forest and XG Boost both provides better results while compared to other classification algorithms and generates 99.29% accuracy.

The work done in [3], The studyput forward an adaptive neuro fuzzy inference system (ANFIS) for forecasting the kidney failure time frame of CKD supported real clinical data. This study used 10-year clinical record patients who are newly diagnosed by chronic kidney disease. The edge value of 15cc/kg/min/1.73m2 of glomerular filtration rate (GFR) was used because the marker of kidney failure. A Takagi Sugeno-Type ANFIS model was utilized to predict GFR values. The comparisons of the anticipated values with the important data showed that the ANFIS model could accurately estimate GFR variations altogether in a sequential period. Despite the high uncertainties of physical body and dynamic nature of CKD progression, our model can accurately predict the GFR variations at long future periods.

The aim is [4], To investigate the feasibility of salivary examination for Chronic Kidney Disease(CKD) discovery and there deliberately an computerized system to identify CKD through investigation of human salivation samples. We've executed an improved deep learning that blends both a one-dimensional Correlational Neural Network (1-DCorrNN) and bidirectional Long Short-Term Memory (LSTM) network for making accurate forecasts. The LSTM network isincorporated with the neural model to use the capacities of both these organizations to explore the time‐arrangement information. The proposed method accomplished a mean accuracy rate of 98.08% for the testing dataset. The outcomes show that the proposed recognition module and classification algorithm considerably advance the current systems, and gives more exact predictions compared to standardmethods.

This decision presents [5], a way towards early detection of chronic renal disease by mining sleep-wake measurements using deep learning. Exactly, a Long-Short- Term Memory network is applied on dataenhanced with analytic history of patients. Experiments and analysis are performed addressing detection at early stage and advanced disease stage supported different clinical data features. The results show for disease detection an averaged certainty of 0.62, 0.73, 0.77 for hypertension, diabetes and chronic kidney disease, respectively. Yet, compared to existing work, our approach shows an improvement in performance and establishes that predicting chronic diseases.

The thought carried out [6], Machine learning strategy has gotten reliable for clinical treatment. With the guide of a machine learning classifier algorithms, the specialist could be ready to identify the sickness on time. For this disposition, CKD dataset has been taken from the UCI repository. During analysis seven classifier algorithms have been applied like artificial neural network, C5.0, Chi-square Automatic interaction detector, logistic regression, linear support vector machine with punishmentL1 & with punishment L2 and random tree. The significant method known as feature selection was likewise applied to the dataset. for each classifier, the outcomes are figured upheld full features, connection based feature selection, Wrapper method feature selection, minimum absolute shrinkage and selection operator regression, synthetic minority over-sampling technique with full features.

This paper presents [7], a neural network‐based classifier to forecast whether an individual is in danger of developing chronic kidney disorder (CKD). The model is trained with the demographic data and medical aid information of two population groups: on the one hand, people diagnosed with CKD in Colombia during 2018, and on the opposite, a sample of individuals without a diagnosis of this disease. After the model is trained and evaluation metrics for classification algorithms are applied, the model accomplishes 95% accuracy within the test data set, making its application for disease prediction feasible. However, despite the determined efficiency of the neuralnetworks to predict CKD, this machine- learning paradigm is impenetrable to the expert regarding the reason of the result.

In this paper [8], to take a power at the force of deep learning strategies an Adaptive hybridized Deep Convolutional Neural Network (AHDCNN) has been proposed for the principal recognition of renal disorder productively and viably. Classification technology depends on the data set. to reinforce the accuracy of the system by reducing feature dimension an algorithm model has been developed using CNN.AHDCNN has been trained to realize the simplest object recognition performance within the large scale image of renal disorder. The smoothing and prior knowledge are often used to achieve an accurate segmentation of the integrated system through AHDCNN. The proposed approach is predicated on

the method for deeper learning and ROIs given by radiologists has shown leads to the classification of renal disorder.

In the model implemented in [9], An ensemble feature selection technique is applied to Chronic renal disorder (CKD) dataset. Density based Feature Selection (DFS) method is employed as a filter approach to rank the features of CKD.The results of DFS method is to a wrapper based optimization technique named Improved Teacher Learner Based Optimization (ITLBO) algorithm to seek out the optimal feature set that contains the foremost important features for prediction of CKD with high accuracy. The ensemble feature selection method are evaluated using SVM, Gradient Boosting, and CNN classification algorithms. Experimental results reveal that the proposed method is in a position to realize high classification accuracy of 93% for SVM, 97% for GradientBoosting and 97.75% for CNN.

In [10], The most focus during this paper is on the classification techniques on tree‐based decision tree, random forest, and logistic regression has been used for analyzing the renal disorder. They used techniques like K-nearest neighbor (KNN)and Naïve Bayes for predicting the presence of Chronic kidney failure (CRF). The dataset used then the filter method of feature selection that's univariate selection and matrix to seek out best features from the dataset.This selection compares each feature to the target variable to ascertain statistically significant relationship between the feature and therefore the target variable. Then, the random forest is that the number of trees perspective, converting this knowledge into a Machine Learning problem definition, and designing a preliminary plan to achieve the objectives.

3. PROPOSED METHODOLOGY

Machine Learning using PYTHON as a tool is helpful to reduce for feature selections, highly relevant to predict CRD and provides suitable analysis, based on different methods such as YOLO CLASSIFICATION. These prediction models can provide relevant information to the clinical experts to judge the patient conditions more smartly and intelligently. The task is to analyze predictive models for chronic renal disease, including Yolo Classification, to predict whether a patient has Chronic Renal disease or not. Output is single and binary (0, 1). 1 for the patients who are affected from CKD and 0 for the patients without CKD. Edema, vomiting, blood pressure, GFR, creatinine, diabetes, cardiovascular-factor, urea, urine protein, age and gender are set of inputs on which the output depends on. A segment of the data should be used for training and part for testing of performance. Which of the models give better result will be examined at the end. proposed methodology is applied to build the mining models. It consists of six major phases: business perception, data understanding, data preparation, modeling, estimation, and deployment. Business understanding phase main aim is on understanding the objectives and requirements from a kidney disease. Data understanding phase uses the data and proceeds to understand the data (medical terminologies, tests etc), identify its standard, gain initial insights, and detect interesting subsets to form hypotheses for hidden information. Data preparation phase put up the final dataset that will be given into the modeling tools. This includes table, record, and attribute selection as well as data cleaning and modification. The modeling phase selects and put in three models as mentioned above, and adapt their parameters to optimal values. The evaluation phase estimates the model to ensure that it attains the business objectives.

The deployment phase specifies the tasks that are needed to use the model It is divided into three modules and the output is analyzed and obtained the result as expected.

- MODULE I: Clustering Module
- MODULE II: Feature Extraction and Selection Module
- MODULE III: Yolo ClassificationModule

MODULE I: CLUSTERING MODULE:

i. We engage Fuzzy C means Clustering which is suitable for clustering medical data.

ii. The data provided by the doctor and the patient has been given into then algorithm.

iii. It is now the work of the clustering algorithm. Data clustering is process of splitting data elements into classes or clusters so that objects in the same class are as similar as possible, and items in different classes are as dissimilar as possible

MODULE II: FEATURE EXTRACTION AND SELECTION MODULE:

It is also known as variable selection or attribute selection, is the process of choosing a subset of relevant features (variables, predictors) for use in model construction. Feature selection techniques are used for various purpose:

i. Shorter training periods

ii. Increased generalization by reducing overfitting.

MODULE III: YOLO CLASSIFICATION MODULE:

From the perspective of automatic learning kidney disease detection can be seen as a classification or clustering problem. On the other hand, we established a model on the vast set of presence and absence file data; we can reduce this problem to classification. For known families, this problem can be reduced to one classification only - having a limited set of classes, including the kidney disease sample, it is easier to identify the right class, and the outcome would be more precise than with clustering algorithms.

4. FLOW OF THE PROCESS

We employ machine learning algorithms and determine if the patient has chronic kidney disease at an earlier stage based on numerous factors available in the dataset. So automated tool which will use machine learning techniques to determine the patient's kidney condition that will be helpful to the doctors in prediction of chronic kidney disease and hence better treatment. The dataset consists of patients details who are already affected by CKD. Finally, after the model gets

trained when a random patient detail is given as input it compares and gives the result. For this we are using yolo classification algorithm which is expected to give better accuracy when compared to all.

5. CONCLUSION

On the idea of results and analysis over the given data, it is clear that Yolo gives best results as. The choice of variables is predicted on the Chi-Square. Yolo performance is higher and error rate is low. We can say that Yolo is best diagnostic model for the given data over chronic renal patients.

REFERENCES

[1] Nusrat Tazin , Shahed Anzarus Sabab, Muhammed Tawfiq Chowdhury 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec).

[2] N V Ganapathi Raju; K Prasanna Lakshmi; K. Gayathri Praharshitha 2019 International Conference on Intelligent Computing and Control Systems (ICCS)

[3] Jamshid Norouzi, Ali Yadollahpour, Seyed Ahmad Mirbagheri, Mitra Mahdavi Mazdeh, Seyed Ahmad Hosseini, "Predicting Renal Failure Progression in Chronic Kidney Disease Using Integrated Intelligent Fuzzy Expert System", Computational and Mathematical Methods in Medicine, vol. 2016.

[4] N. Bhaskar, M. Suchetha and N. Y. Philip, "Time Series Classification-Based Correlational Neural Network With Bidirectional LSTM for Automated Detection of Kidney Disease," in IEEE Sensors Journal.

[5] S. Fallmann and L. Chen, "Detecting Chronic Diseases from Sleep-Wake Behaviour and Clinical Features," 2018 5th International Conference on Systems and Informatics (ICSAI) .

[6] P. Chittora et al., "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," in IEEE Access, vol. 9, pp. 17312-17334, 2021.

[7] G. R. Vásquez-Morales, S. M. Martínez Monterrubio, P. Moreno-Ger and J. A.Recio García, "Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning," in IEEE .

[8] G. Chen et al., "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," in IEEE .

[9] M. Manonmani and S. Balakrishnan, "An Ensemble Feature Selection Method for Prediction of CKD," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020.

[10] 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