|  |  |
| --- | --- |
| **PREDICTION OF CHRONIC KIDNEY DISEASE USING SVM** | PRESENTED BY:  PRACHETAA S  NANDHINI SRI V R  DEEPAK SURYA M  BHAVADHAARANY R  ISHRATH TASNEEM K |

**1 INTRODUCTION**

1.1 OVERVIEW

This is an ambitious, student-led, initiative to plan, design and build a model which provides an efficient and faster way to predict whether the patient is suffering from the major kidney problem known as CHRONIC-KIDNEY DISEASE.

CHRONIC KIDNEY DISEASE (CKD)is a major medical problem and can be cured if treated in the early stages. This model is built in with different conditions of people (i.e with ckd and without ckd).Thus , it would help in predicting the conditions of new suspects in a much faster way.

1.2 PURPOSE

The attributes collected from various medical tests are used here to measure the severity of the problem. The attributes may include the person's age factor, red blood cells count, hemoglobinlevel, blood pressure, etc. In this prediction model, we feed those values of attributes collected as input.

And this trained model would deliver the feedback whether the patient is suspected to have the disease or not. Thus, this would help in working to cure the disease and would be able to predict the survival rate of the patient after illness.

**2.LITERATURE SURVEY**

2.1 EXISTING PROBLEM

Healthcare professionals are able to infer the acuity and severity of a case without superfluous or redundant documentation, but auditors may not have this ability. Adequate documentation for every service date helps to convey patient complexity during a medical record review. Although the problem list may not change dramatically from day to day during a hospitalization, the auditor only reviews the service date in question, not the entire medical record.

Hospitalists should be sure to formulate a complete and accurate description of the patient’s condition with an analogous plan of care for each encounter. Listing problems without a corresponding plan of care does not corroborate physician management of that problem and could cause a downgrade of complexity. Listing problems with a brief, generalized comment (e.g. “DM, CKD, CHF: Continue current treatment plan”) equally diminishes the complexity and effort put forth by the physician.

Greater recognition and support for CKD may require that the disease no longer be viewed as one continuous disease state. The major problem with this disease is it is hard to recognize till it reaches advanced state. Early CKD stages require less complex care and generate lower costs. In contrast, late-stage CKD is every bit as complex and costly as other major chronic diseases. Health authorities may not recognize and fund CKD care appropriately until late-stage CKD is defined clearly as separate and distinct from earlier stages of disease. In this review, we describe the burden of chronic diseases, consider the challenges and barriers and propose processes to improve late-stage CKD care. In particular, we recommend the need for improved continuity of care, enhanced use of information technology, multidisciplinary care, timely referral to nephrologists, protocol use and improved patient engagement.

2.2 PROPOSED SOLUTION

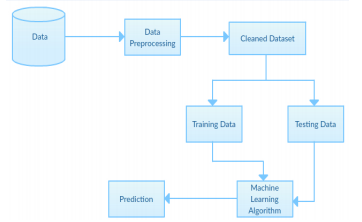
Physician documentation should always:

* Identify all problems managed or addressed during each encounter;
* Identify problems as stable or progressing, when appropriate;
* Indicate differential diagnoses when the problem remains undefined;
* Indicate the management/treatment option(s) for each problem; and
* Note management options to be continued somewhere in the progress note for that encounter (e.g. medication list) when documentation indicates a continuation of current management options
* Data credits should be given.
* The data documented by the physician should be safe and persistent monitoring should be undergone in a quick manner and at regular intervals.

Using this model which is trained with a multiple and variety of inputs would be more helpful in quicker diagonising the patient's condition. The basic test for diagnosis is blood test, urine test. Along with those the physician uses your age, gender, etc. The necessary data can be collected and the values can be documented. Those values can be fed to this model as input and with the output predicted, we can conclude the severity or status of the patient, that he/she should undergo hemodialysis or renal transplantation or peritonial dialysis or continue with meds or cured. Thus ,it becomes more efficient and we can avoid loss of data as well as delay in predictions.

**3.EXPERIMENTAL ANALYSIS**

3.1 BLOCK DIAGRAM



3.2 SOFTWARE DESIGNING

STEPS INVOLVED IN BUILDING THIS MODEL:

* Import the required packages that is numpy, pandas and matplotlib.pyplot.
* A proper dataset must be collected and should ensure that there are no null values
* Import the dataset using pandas
* Ensure that there are no null values andif there is any fill them with either mean or mode.
* Make sure that the ID column is dropped as it is not necessary for processing further.
* Import Label Encoding method from sklearn.preprocessing and make sure that the dataset values of type either or (present, not present, normal, abormal,yes no) are encoded to 0 and 1.
* Seperate the independant and dependant values and store them in X and Y respectively.
* Check for outliers using Boxplot method and remove if any.
* The outliers are removed using IQR (Inter Quartile Range).
* Import Standard Scalar from sklearn and fit the scaled values in x.
* From sklearn import train\_test\_split method for further process.
* Split the dataset for training and testing using train\_test\_split method and give 80% of data for training the model whereas the rest can be used for testing.
* Now choose SVM algorithm from sklearn which is the required and suited algorithm for the given problem statement.
* From SVM we import SVC, using kernel='rbf'.
* Now we fit the training data to the algorithm.
* Predicted data is now stored in a new variable named y\_pred and testing is done for the correctness of model.
* To find the accuracy we import accuracy\_score method from sklearn.metrics.
* The y\_test and y\_pred data are fed and the accuracy is checked.
* Here ends the model building method. The next step we do is the application bulding.
* In application building we create an html file where the user (physician) enters the documented values of attributes such as rbc, heomoglobin level, blood pressure, age, etc and can witness the output.
* For the backend part to function we write a python code named(app.py) in which we import the necessary web framework named FLASK, for the web page to function.
* In this flask, we introduce the files with “ .save” extensions which is obtained by importing the package named joblib.
* Now run the .py file in anaconda prompt, where we get the ip address to our website, which should be opened through our browser.
* The obtained ip address will direct us to our website where we can feed our values and obtain the predictions.

**4.Experimental Investigation**

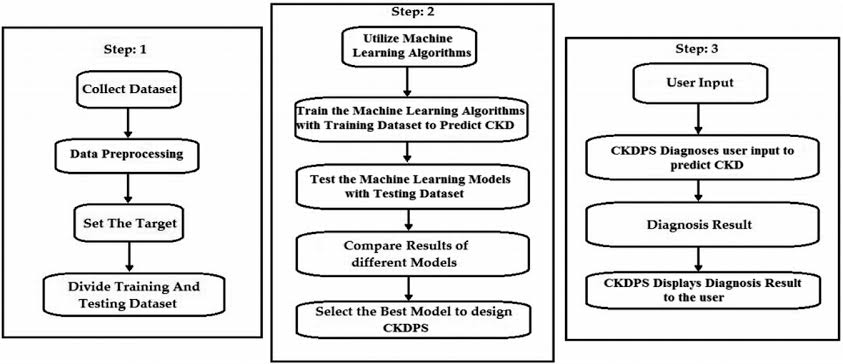
The main aim of this project is to identify a patient who is either tested positive or negative for chronic kidney disease which is a major medical problem and can be cured if treated in the early stages. This model is built in with different conditions of people (i.e with ckd and without ckd).Thus , it would help in predicting the conditions of new suspects in a much faster way.

Data like

* Age
* BP (Blood Pressure)
* SG (Specific gravity)
* AL (Albumin)
* SU (Sugar)
* RBC (Red Blood Cells)
* PC (Pus Cell)
* PCC (Pus Cell Clumps)
* BA (Bacteria)
* BGR (Blood Glucose Random)
* BU (Blood Urea)
* SC (Serum Creatinine)
* SOD (Sodium)
* POT (Potassium)
* HEM (Hemoglobin)
* PCV (Packed Cell Volume)
* WC (White Blood Cell Count)
* RC (Red Blood Cell Count)
* HTN (Hypertension)
* DM (Diabetes Mellitus)
* CAD (Coronary Artery Disease)
* Appet (Appetite)
* PE (Pedal Edema)
* ANE (Anemia)

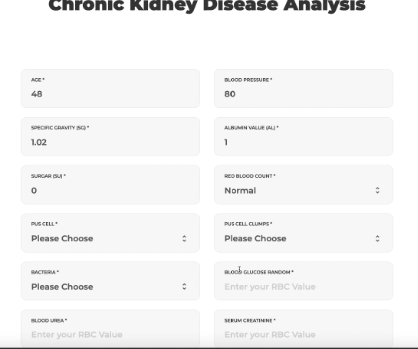
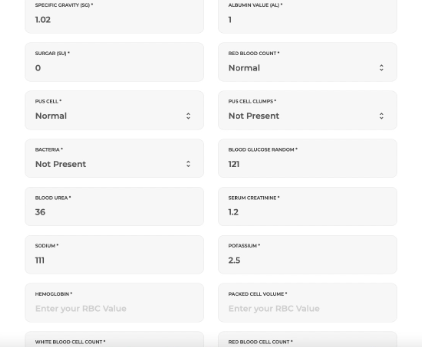
are collected from the patients and then based on the observations and analysis, the model classifies them as either CKD(Chronic Kidney Disease) or not CKD(NO Chronic Kidney Disease).This classification analysis is no complicated task, because it simply collects the data which is fed already and based on the pre-processed facts or algorithms, the model is able to predict the output that is it classifies the processed input as either 0 or 1 based on the certainty. SVM algorithm is used in this model which yields 97% accuracy rate. The performance of the model is certain and efficient without any perplexity. On Experimental Analysis the model is obvious with the collected data and fabricates the result of about ninety seven percent correctness. The model is trained with 80% of the data which is why the predictions are true to the natural behaviour when a new suspect is diagonised. As the model adapted the nature of the data which is widely ranged, the prediction which is treated on analysis will be efficient even when a large number of data is fed.

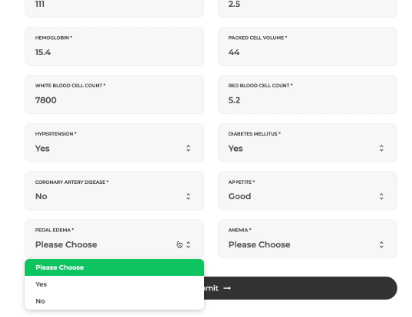
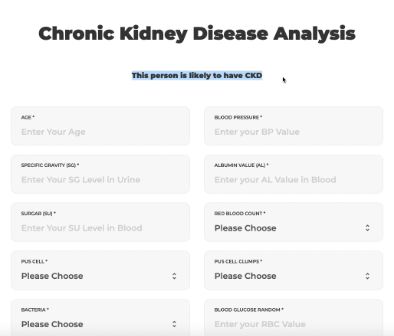
**5.FLOW CHART**



**6.RESULT**

This model is being built using Support Vector Algorithm which comes under Supervised Machine Learning that can be used in classification and regression problems. Since this model is a classification problem, the SVM algorithm yields about 97% accuracy .The result of each classifier has been evaluated using different evaluation metrics and validated against overfitting values. The experiments are conducted using Python 3.7 programming language through the Jupyter Notebook web application. Several libraries from Sciket-learn have been used, which is a free software for the machine learning library in Python

**7.ADVANTAGES AND DISADVANTAGES**

**ADVANTAGES**

* Faster and Reliable.
* Efficient output with maximum accuracy.
* Predicts with certainity.
* Simpler concept.
* Can perform even with large number of suspects.
* Lesser perplexity.
* Earlier deduction reduces high risks.
* Can prevent loss of data.
* It is a proposal for best prediction framework for CKD.

**DISADVANTAGES**

* Requires server which processes heavy loads.
* Requires stable network connection.
* May lead to overfitting.
* The model should be trained properly with a balance between CKD and not CKD values, else there may be chances of predictions with depreciated accuracy.

**8.APPLICATIONS**

Machine Learning is a buzzward for today's technology and it is growing rapidly day by day. We are using Machine Learning in many ways for our daily lives. This aids greatly for simplifying our day to day tasks by using one of its efficient algorithms. Likewise various real time applications are

* Medical diagnosis.
* Stock market Trading.
* Educational Institutions
* Online Fraud Detections etc.

This model can be efficiently used in Healthcare Systems such as Diagnosis which aids in rapid growth of medical technology. This model identifies positive and negative cases of CKD which helps in early detection of the disease and reduces risks. Also predicts with certainty.

**9.CONCLUSION**

In this project, we have analysed 25 different attributes related to CKD patients and predicted accuracy using Support Vector Machine algorithm which is a supervised algorithm. From the results analysis, it is observed that the SVM gives accuracy of 97%. The advantage of this system is that, the prediction process is less time consuming. It will help the doctors to start the treatments early for the

CKD patients and also it will help to diagnose more patients within a less time period. Limitations of this study are the strength of the data is not higher because of the size of the data set and the missing attribute values. To build a machine learning model targeting chronic kidney disease with overall accuracy of 99.99%, will need millions of records with zero missing values.

**10.FUTURE SCOPE**

This work will be considered as basement for the healthcare system for CKD patients. Also, extension to this work is the implementation of deep learning since deep learning provides high-quality performance than machine learning algorithm.

**11.BIBLIOGRAPHY**

* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5087766/>
* <https://www.ajkd.org/article/S0272-6386(06)01914-7/abstract#back-bib2>
* <https://www.sciencedirect.com/science/article/abs/pii/S0010482519301258>
* <https://link.springer.com/article/10.1007/s10916-017-0703-x>