Predicting Chronic Kidney Disease Using KNN Algorithm

V Mareeswari¹, Sunita Chalageri² & Kavita K Patil³

¹,²,³Department of Computer Science and Engineering, ACS College of Engineering, Bangalore

DOI: https://doi.org/10.34293/acsjse.v1i2.10

Abstract - Chronic kidney disease (CKD) is a world health issue, and that includes damages and can't filter blood the way it should be. Since we cannot predict the early stages of CKD, patients will fail to recognize the disease. Pre-detection of CKD will allow patients to get timely facility to ameliorate the progress of the disease. Machine learning models will effectively aid clinicians' progress this goal because of the early and accurate recognition performances. The CKD data set is collected from the University of California Irvine (UCI) Machine Learning Recognition. Multiple Machine and deep learning algorithms used to predict the chronic kidney disease.

Keywords: Chronic Kidney Diseases; Disease Registration System; Disease results conclusions.

I. INTRODUCTION

Chronic kidney disease (CKD) is defined as the current situation of the kidney being damaged, usually defined as the abnormality of test blood, imaging in kidney or urinalysis.

Or Estimated Glomerular Filtration Rate (eGFR) that has less than 60 ml/min per 1.73 for two or three months. It is predicted that there are nearly 500 million patients with the CKD in the whole world, making up 7-10% of the total population, which became the global public health problem. Diagnosing CKD will start with the clinical data, imaging studies, lab tests and the biopsy. Even though biopsy is the diagnosing standard tests, it also has a lot of disadvantages, they are being costly, invasive, consuming time and also risky.

For example, when the biopsy is done, the patient might have infection, the surgery scare and would be misdiagnosing. Imaging studies like mammogram, sonogram, and MRI of the Kidney have been used for so many years till now to detect the disease, but to use them have some limits more expressly to exposure effects of radiation. Also, being it risky the data given by imaging is not sufficient for the diagnosing.

GFR the CKD patients will reduce approximately 1% per year without having the valid treatment. For the best prognosing of the patients, interventions were been advocated, for example, Angiotension were advocated for example, Angiotension2 receptor antagonists (ARBs) was to reduce ex-creation of urinary proteins and down slows the progression of CKD.

Early inventions on risk analysis and factors refer to the nephrologists specialist are critical in delaying the disease. It only not affect the probiscis after being initialized dialysis, but as the quality of the life and the economy of burdens of the public. In the different study the researchers trying to predicts the Long terms kidney Transplantations overview. They have played comparing work on ANN and Linear Regression Algorithm.

The comparison analysis was implemented on the progressive metrics as in specificity, sensitivity and accuracy, while the study of kidney transplantation prediction of rejection of the kidney, which were based on the nine training and verifying dataset. The results showed that ANN algorithm can be useful being supportive. In the summary predicting kidney the
process defined problem. In the ability to predict the kidney rejection. It was 37% for Logistic Regression of 61% KNN. The ability of pre guiding of no rejection was 67% for Logistic Regression comparing for 84% of KNN by literature review, identified studies are less performeds in CKD by using SVMs and there is many studies how SVM were been used. In these instances studies researchers were using SVM. With a classification model in the diagnosing and detecting benign tumour and malignant on the MRI features, feature of ultrasound, feature of mammographic. It will result the try and create the test for the classifier of SVM by the CKD dataset. Improve order in the accuracy, classification of positive, negative of the performance in the Support Machine Vector that has been evaluated. The flexible algorithm classifier been proposed in the learning statistical effective method for pattern recognition on which it is finding hyperplane optimal to different classes input data of mapping of input data to the higher dimensional space. SVM used many application like the recognition of object and detection of face.

To achieve intervention early for chronic diseases, system registration were essential. potential CKD patients could detect and identify, being received the nephrology-specialized in the time of the care. or number of CKD may be overlooked on the patients or missed the time to get treated early for interventions. Investigating CKD prevalence and portion of patients by establishing CKD system registration.

Research contributes the decision support as the tool in the early, that tool is handed well as an emergency condition. doctor needed little more information by the work related by the patients to make decision on her/his coming treatment. This significant research to be prepared in the emergency because of the condition that is to be faced. they are hospital infrastructure to damage, being lack of information for example lost of the medical report. research are worked on how kidney treat the diseases under various emergency situations.

In this case, we did propose diagnosing support system of the CKD using the Support Vector Machine(SVM) in the algorithm for classification and the clustering. The idea of the basic algorithm SVM looks so optimal in the use of two pieces of separate class called hyperplanes. Researchers have been done the application in the SVM for the classification over the various field in the affective computing. The decision of the developments in the decision system diagnosing kidney diseases system support in the predicting whether the patients in the diseases been entering the phase of CKD or Not.

For improving the doctor’s condition on the judgement of the patients. supporting diagnosing need of the system in the various altitude. The approach have the system that will use the technique of the machine learning by the classification. several previous study in the field like Modifies Hybrid Cuckoo.

II. RELATED WORK (SURVEY)

Prediction of Intelligent Diagnostic and system classification for CKD. This introduces intelligent prediction and arrangement for density based feature selection(DFS) supported feature selection. The intelligent system that proposed will eliminates which isn't relevant or which is redundant feature by the DFS within the early to classifier construct. The one proposed the framework have three phases of the name preprocessing feature selection
and therefore the classification support of the choice that's diagnosing the system of CKD using the vector machine network.

Standard test is diagnosed by renal not being ordered is Biospy, this is numerous disadvantages, It is very expensive, it will take and dangerous, For example: when biopsy is performed, infection is being recognized, surgery fear and misdiagnosis. (breast imaging, ultra sound and kidney MRI) these methods are useful since years to find chronic kidney disease, but it has its own limitation. specifically, radiation exposure, also being at risk, information by imaging diagnostics is enough to find CKD. capability to predict renal rejection was 38% for linear regression and 61% for ANN, predict of power of losing was 70% and for LR and 60% for LR comparatively ANN. patience from some hospital in the world admitted between jan 12 and Dec 2012 were there through data system.

Data Collections and Statistical Analysis
1) Data Collections and Selections: In 2012, serum creatinine stage and routeurineanalysis clinical results of inpatients were collected through the hospital's data systems. result for patients withins the same ID number were selected. In 2012, for patients who were diagnosed with similar symptoms multiple times in the hospital, only the laboratory test resultt at the time of the first admission were recorded, and the remaining results were excluded, and the inpatients who tested the serum creatinine level. The results of the urine test were shown as follows

2) CKD diagnosis: Patient data, that is, general information such as gender and age, is accessed and updated via the hospital's data system. Using the CKDEPI formula, this method was used to calculate the eGFR for all patients. Based on the rules obtained from hospital information, patients with eGFR of 60 ml / min or without positive leukocytes on equivalent tests. The amount of patients diagnosed with CKD was shown in B.

3) Based on the distribution of inpatient CKD medicines admitted to the kidneys, the numbers of CKD patients was large, and the tops 5 departments were calculated.

CKD was obtained from UCI Machine Learning Storage System. A total of 401 cases were included, of which 249 were patients with CDK and the remaining 149 were CKD-free. The variable of interest indicates whether the patient has been diagnosed with a CDK. There are 24 clinical features, the rest of which are some of the 25 attributes that can be target attributes. It has several features, namely a history, physical examination and clinical examination. Consistent with the nature of the attribute, the target attributes was classified as negative ("without disease") and positive ("presince of diseases").

III. CHALLENGES
The prevalence of chronic kidney disease (CKD) has increased rapidly over the last two decades as the populations ages, resulting in increased diabetes, hypertension, and obesity. 14. A smalll percentage of those patients reach end-stages renal diseases (ESKD), which requires renal replacment therapys. Over the last decades, the amount of A is now over 20000.5 From 2009 to 2020, the cumulative costs of providing treatments to new and existing ESKD patients in Australia is estimated to reaching $ 12 billion. 5 Significant
efforts to expand the use of low-cost home dialysis treatments and transplants, but the best and most cost-effective treatments for ESKD are to apply these methods to meet more demand. There are considerable barriers to. Therefore, it is necessary to focus on switching to prevention of the exact purpose in CKD treatment. Not only is CKD's burden associated with many of which may be ready justed by medical professionals. There is another issue where we have to participate in the issues of various actors such as government agencies rather than for-profit organizations and industries. The Dialysis is required in patients who have reached 13ESKD. Then those who missed timely investigation and prevention opportunities and received treatment from vital signs were achieved in only 59%. Most patients in the early stages of CKD are trained (managed) by medical doctors. Training and education to improves physician awareness and knowledges of CKD is important to prevent CKD. This program must be successful has raised awareness over the years before has educated health care and public institutions a lot about CKD.

We started with a health education program from Kidney Health, Australia, improved our CKD management guidelines with common practices, and launched a recognition campaign called Kidney Health Week, which was successfully accepted. Despite 10, thank you for going so far and achieving it. This area requires joint efforts with medical professionals, government agencies and nonprofits. The secondary challenge of CKD prevention is associated with deficiencies in current screening methods. Existing CKD testing programs rely on the measurement of proteinuria and GFR. Factors such as physical activity, postures, and timings of urine collections can affect proteinuria. Laboratories measurements of creatinine and urinary proteins are also quite different. The supported creatinine-based formula estimates GFR, which induces significant fluctuations in the estimates. Currently used in Australia, the official is the CKD Epidemiology Collaboration formula for estimating GFR. The formula for CKD mechanical cooperation is less based as the GFR is higher than the formal awareness of kidney disease is deformed.

<table>
<thead>
<tr>
<th>Department</th>
<th>CKD cases</th>
<th>Ratio of CKD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency</td>
<td>1651</td>
<td>7.88</td>
</tr>
<tr>
<td>Breast</td>
<td>852</td>
<td>4.07</td>
</tr>
<tr>
<td>Urology</td>
<td>838</td>
<td>4.00</td>
</tr>
<tr>
<td>General</td>
<td>737</td>
<td>3.52</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>703</td>
<td>3.36</td>
</tr>
<tr>
<td>total</td>
<td>4781</td>
<td>22.83</td>
</tr>
</tbody>
</table>

Top five Departments with The Largest Number of CKD Patients
IV. **PROBLEM STATEMENT**

- A biopsy is a standard diagnostic test that is expensive, invasive, time consuming, and in some cases dangerous. For an example: When a biopsy is performed, the patients can experience infections, fear of surgery and misdiagnosis.
- Imaging tests such as kidney mammography, ultrasound and MRI have been used for many years to detect diseases. However, there are some restrictions to using them. More clearly, it is the exposure effect of radiation. The information provided by the imagination of danger is not sufficient for diagnosing CKD.
- The ability to predict renal rejections (sensitivity) was 35% for LR and 60% for ANNs. The abilities to predict non-rejections (specificity) was 63% for LR compared to 87% for ANNs.

V. **PROPOSED SYSTEM**

From January 2012 to December 2012, all patients from Guangdong State Hospital admitted to the hospital in January were registered in the hospital data system. For diagnosis, most early-stage patients with CKD receive medical training (management). Physicians, training and education are very important for CKD prevention in order to improve the awareness and knowledge of CKD physicians. These programs have been successful.

Educate medical services and public and public service providers by spreading awareness over the past few years. Lots of about CKD.

Kidney Health Australia needs to start and implement a health education program. Improved guidelines for managing CKD with current common practices and an early recognition campaign for Kidney Health Week were successfully received 10 but among the successes of the field and health care professionals to date. Thanks to the joint efforts in, we need government agencies and non-profit organizations. A second challenges in CKD preventions relate to deficiencies in current screenings methods. Currently, the CKD screenings process relies on measurements of both proteinuria and GFR. Factors such as physical activity, posture, and timing of urine collections can affect proteinuria. Still, there are only significant differences in laboratories measurement of urine protein and creatinine. GFR is believed to be a supported creatinine-based regimen. This causes significant fluctuations in the estimate. The formula currently used in Australia is the CKD epidemiologic collaboration for estimating GFR. Compared with dietary changes in the prescribing of kidney disease, the prescribing of CKD epidemiologic collaboration is less biased as the GFR is higher, and there is more than a differentiation in the prognosis of the Australian population.

Nevertheless, none of these calculations take B into account. It was. Test results for all patients with the same ID numbers were selected. For instance if The patients only booked laboratories test results from the first admission on several hospitalisations in 2012, and the rests of the results were discarded. The number of hospitalized patients tested for serum creatinine levels and urine analysis was denoted by A.

2) CKD diagnosis: General patient information such as age and gender is collected through system information. The eGFR for all patients was calculated using CKDEPI's formula. Consistent with the given guidelines, patients with an eGFR of 58 ml/min and abnormal
findings in urine analysis were diagnosed (confirmed) with CKD. Irregularurinalysiis was defined as proteinuria or hematuria without positive white blood cells on an equivalent test. The number of patients diagnosed with CKD is displayed in B.

E. 3) Prevalance Calculutions: Hospital CKD prevalance was a calculated using formula below.

VI. DATASET ELUCIDATION
• Collected datasets from the portal of UCI Web
• Next, we processed dataset
• Next, I converted the processed data set into useful records.
• Next, you should always train your dataset using different DL and ML algorithms.
• Forecast File
• Next, float the result and accuracy graph.

VII. PROPOSED CNN
LOG gets the deflection and the load of all predictors based on linear regression. If the overall sums of the results of all predictors exceeds a certain thresholds, the sample category is categorized as Not CKD or CKD RF, and the predictors and random training samples are sampled into a decision tree. Generate. Each decision tree Not CKD was trained to search for boundaries that maximize the difference between CKD and CKD. Prediction determines the final decision of all trees within the diagnosis of the disease. Set predictive variables for the sample Set the surface of the selection in multidimensional space to divide the different still life samples by SVM. The Naïve Bayes classifiers calculates the conditionall probability of Not CKD and CKD samples at each other measurements intervals, samples of the interval. The closest training sample calculates the distance between training samples and the test sample, searches from KNN and determines the diagnostic category through voting. Due to its complex structure, FNN was able to analyze non-linear relationships in the dataset, and the sigmoid activations function was used for unnikchun and chulryokchun.
VIII. MODEL SETUP
1. Data Pre-Processing

Regression-based model: LOG
Tree-based model: RF
Decision plane-based model: SVM
Distance-based model: KNN
Probability-based model: NB
Neural network: FNN

Python: A high-level, parse, and general-purpose programming language. Features and Structured Programming Languages and OOP

NumPy: NumPy is a scientific package that supports Python. Supports the addition of large multidimensional arrays and matrices.

Pandas: Pandas is a fast, powerful, flexible and easy-to-use open source data analysis and manipulation tool.

Used to work with tabular data such as data stored in spreadsheets or databases.

Matplotlib: Matplotlib is a comprehensive library for creating animated interactive visualizations in Python.

Create publication-quality plots in a few lines of code.
Scikitlearn: The scilearn library contains many efficient tools for machine learning and modeling, including classification, regression, clustering, dimensionality reduction, and more.

Precision (A) Commonly used as a measure of classification method. However, compliance rate values are no less susceptible to fluctuations in the number of correct decisions than the compliance rate and recovery rate. Eq. 1 To check the accuracy of prediction.

\[
\text{Accuracy} = \frac{TP + TN}{\text{Total no of Samples}}
\]  

Precision (P) It is the conditional probability that document d is classified under a class(ci). It measures the ability of classifiers to place a sample document under the correct class as opposed to all documents which are placed in that class, both correct and incorrect. Eq.2 is used for calculating precision

\[
\text{Precision} = \frac{Tp}{TP + FP}
\]  

Recall (R) A decision is taken to determine the probability whether document d should be classified under class (ci). Eq. 3 is used for calculating Recall.

\[
\text{Recall} = \frac{Tp}{TP + FN}
\]  

F-Measure (F) Precision and recall are combined to produce a single metric known as F-measure, which is the weighted harmonic mean of precision and recall. Eq. 4 below is used to find out F-Measure.

\[
F = \frac{2 \times P \times R}{P + R}
\]

IX. REFERENCES


