M.KUMARASAMY
COLLEGE OF ENGINEERING
NAAC Accredited Autonomous Institution
Approved by AICTE & Affiliated to Anna University
ISO 9001:2015 & ISO 14001:2015 Certified Institution
Thalavapalayam, Karur – 639 113.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NALAYATHIRAN PROJECT

TITLE : EARLY DETECTION OF CHRONIC DISEASE USING MACHINE LEARNING

DOMAIN : APPLIED DATA SCIENCE

FACULTY MENTOR : Dr. S. SUJANTHI M.E., PH.D.,

INDUSTRY MENTOR : SOWJANYA, SANDEEP DOODIGANI

PROJECT MEMBERS :

REGISTER NO	NAME	ROLE
19BCS4001	S. AARTHI NIVETHA	LEADER
19BCS4007	M. AKCHAYA	MEMBER
19BCS4112	S. SNEHA	MEMBER
19BCS4401	A. POOJA	MEMBER

EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

ABSTRACT:

Chronic Kidney Disease is a serious lifelong condition that induced by either kidney pathology or reduced kidney functions. Early prediction and proper treatments can possibly stop, or slow the progression of this chronic disease to end-stage, where dialysis or kidney transplantation is the only way to save patient's life. In our project, we examine the ability of several machine-learning methods for early prediction of Chronic Kidney Disease. This matter has been studied widely; however, we are supporting our methodology by the use of predictive analytics, in which we examine the relationship in between data parameters as well as with the target class attribute. Predictive analytics enables us to introduce the optimal subset of parameters to feed machine learning to build a set of predictive models.

LITERATURE SURVEY:

Andressa C.M. da Silveira et al [1] proposed a system that uses the methodology helps to alleviate such issue, investment in early prediction is necessary. The purpose of this study is to assist the early prediction of CKD, addressing problems related to imbalanced and limited-size datasets. We used data from medical records of Brazilians with or without a diagnosis of CKD, containing the following attributes: hypertension, diabetes mellitus, creatinine, urea, albuminuria, age, gender, and glomerular filtration rate. We present an oversampling approach based on manual and automated augmentation. We experimented with the synthetic minority oversampling technique (SMOTE), Borderline-SMOTE, and Borderline SMOTE SVM. We implemented models based on the algorithms: decision tree (DT), random forest, and multi-class AdaBoosted DTs and the k-nearest oracles- union, k-nearest oracles-eliminate, and META-DES for dynamic ensemble selection. We analyzed the models' performances using the hold-out validation.

Problem identified:

When dealing with imbalanced and limited-size datasets, the evaluation of resampling and validation methods is essential to verify the stability of ML models. However, one of the main limitations of this study is the usage of the gridSearchCV tool to find the best parameters for each algorithm. We faced processing limitations, mainly for the ensemble models, because the parameter search was conducted for each ML model. The usage of gridSearchCV with 5 folds for the DT model is one example of such a situation. We handled 960 candidates, resulting in 4800 adjustments. However, when using the META-DESmodel, we handle 8640 candidates, resulting in 43,200 adjustments for the ensemble model, presenting a higher processing cost to adjust the parameters.

Rayan Alanazi et al [2] described a system for all over the world, chronic diseases are a critical issue in the healthcare domain. According to the medical statement, due to chronic diseases, the death rate of humans increases. The treatments given for this disease consume over 70% of the patient's income. Hence, it is highly essential to minimize the patient's risk factor that leads to death. The advancement in medical research makes health-related data collection easier. The healthcare data includes the demographics, medical analysis reports, and the history of disease of the patient. The diseases caused could be varied based on the regions and the living habitats in that region. Hence, along with the disease data, the environmental condition and the living habitat of the patient should also be recorded in the data set.

Problem identified:

Nowadays, humans face various diseases due to the current environmental condition and their living habits. The identification and prediction of such diseases at their earlier stages are much important, so as to prevent the extremity of it. It is difficult for doctors to manually identify the diseases accurately most of the time. The goal of this paper is to identify and predict the patients with more common chronic illnesses. This could be achieved by using a cutting-edge machine learning technique to ensure that this categorization reliably identifies persons with chronic diseases. The prediction of diseases is also a challenging task. Hence, data mining plays a critical role in disease prediction.

Reshma S et al [3] proposed a system in which chronic Kidney Disease also recognized as Chronic Renal Disease, is an uncharacteristic functioning of kidney or a failure of renal

function expanding over a period of months or years. Habitually, chronic kidney disease is detected during the screening of people who are known to be in threat by kidney problems, such as those with high blood pressure or diabetes and those with a blood relative Chronic Kidney Disease(CKD) patients. So the early prediction is necessary in combating the disease and to provide good treatment. This study proposes the use of machine learning techniques for CKD such as Ant Colony Optimization(ACO) technique and Support Vector Machine(SVM) classifier. Final output predicts whether the person is having CKD or not by using minimum number of features.

Problem identified:

Chronic Kidney Disease (CKD) is considered as an important threat for the society with respect to the health in the present era. Chronic kidney disease can be detected with regular laboratory tests, and some treatments are present which can prevent development, slow disease progression, reduce complications of decreased Glomerular Filtration Rate(GFR) and risk of cardiovascular disease, and improve survival and quality of life.

Tauja K J et al [4] described a system that the investigation proposes the utilization of Machine learning techniques like Support Vector Machine (SVM), Naive bayes, Random Forest, Decision Tree classifier. Presently, there are numerous individuals on the planet experiencing chronic kidney infections around the world. Because of the few danger factors like food, climate and expectations for everyday comforts numerous individuals get infections abruptly without comprehension of their condition. Diagnosing of persistent kidney illnesses is by and large intrusive, exorbitant, tedious and frequently hazardous. That is the reason numerous patients arrive at late phases of it without treatment,particularly in those nations where the assets are restricted. Last yield predicts if the individual is having CKD by utilizing least number of highlights. In this project, Naive Bayes, Random Forest, Support Vector Machine and Decision Tree are employed for thedisease detection.

Problem identified:

According to the World Health Organization, chronickidney disease has become a major concern in developing countries (WHO). CKD is a kidney disease that may be treated in the early stages but leads to renalfailure in the later stages. Chronic renal disease claimed the lives of 753 million people worldwide in 2016, including 336 million men and 417 million women. Itis classed as a "chronic" illness since the kidney infection develops gradually andlasts a long time, affecting the kidney's function. The amassing of side-effects inthe blood prompts the rise of other medical issues, which are related with a few indications, High and low circulatory strain, diabetes, nerve damage, and bone problems are all factors that contribute to cardiovascular disease. Diabetes, pulse,and cardiovascular disease (CVD) are all risk factors for CKD patients. Incidentalconsequences impair the apprehensive and invulnerable framework in CKD patients, especially of the late stages of the disease. Patients in agriculturalcountries may come at a late stage, necessitating dialysis or kidney transplants.

Deepika Bidri et al [5] proposed a system where the main causes are damaged blood vessels of the kidneys due to High Blood Pressure and Diabetes. The CKD is also called a chronic kidney failure where according current medical statistics the 10% of the population worldwide is affected by CKD. There were approximately 58 million deaths in the year of 2005 worldwide. According to the World Health Organization (WHO), 35 million attributed to chronic diseases. Currently it is estimated that one in five men, and one in four women aged 65 through 74 are going to be affected by CKD worldwide. Diagnosing CKD usually starts with clinical data, lab tests, imaging studies and finally biopsy. In this study, by using the machine learning techniques, we are proposing cheap, simple and non-invasive tests that can be performed easily. By this strategy, we hope to produce "down- staging" (increasing in the proportion of CKD detected at an early stage) of the disease to stages that are more amenable to curative treatment.

Problem identified:

There are many people who are suffering from chronic kidney diseases worldwide. Due to the several risk factors like food, environment and living standards many people get diseases suddenly. Diagnosing of chronic kidney diseases is generally

invasive, costly, time-consuming and often risky. That is why many patients reach late stages of it without treatment, especially in those countries where the resources are limited. Therefore, the early detection strategy of the disease remains important, particularly in developing countries, where thediseases are generally diagnosed in later stages. Finding a solution for the above-mentioned problems and riding out from disadvantages became a strong motive to conduct this study. Chronic Kidney Disease is one of the types of kidney disease, which results in a gradual loss of kidney function. This phenomenon can be observed over a period of months or years due to several living conditions of patients. The goal is to build a real time application by usingthe machine learning techniques (Naive Bayes and KNN algorithms), to detect the CKD at an early stage.

SUMMARY OF LITERATURE SURVEY:

Author	Year	Title	Algorithm used	Limitations
Andressa C.M.da Silveira	2022	Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms	Decision tree (DT), random forest, and multi-class AdaBoosted DTs	Leads to processing limitations, mainly for the ensemble models
Rayan Alanazi	2022	Identification and Prediction of Chronic Diseases Using Machine Learning Approach	Convolutional neural network (CNN), K-nearest neighbor (KNN)	Identify and predict the patients with more common chronic illnesses
Reshma S	2020	Chronic Kidney Disease Prediction using Machine Learning	Chronic kidney, SVM, Ant colony optimization	Slow disease progression, reduce complications of decreased Glomerular Filtration Rate(GFR)
Tauja K J	2019	Detection of Chronic Kidney Disease Using Machine Learning Techniques	CKD, Decision Tree, SVM, Random Forest, Naive Bayes	The strength of the data is not higher because of the size of the dataset
Deepika Bidri	2018	Early Prediction of Chronic Kidney Disease by using MachineLearning Techniques	Naive bayes; K-Nearestneighbor; Machine learning	Leads to low accuracy

REFERENCE

1. Andressa C.M. da Silveira, Exploring Early Prediction of Chronic Kidney Disease Using MachineLearning Algorithms, January 2022

2. Rayan Alanazi, Identification and Prediction of Chronic Diseases Using Machine Learning Approach, February 2022

3. Reshma S, Chronic Kidney Disease Prediction using Machine Learning, July 2020

4. Tauja K J**.** Detection of Chronic Kidney Disease Using Machine Learning Techniques, March 2019

5. Deepika Bidri. Early Prediction of Chronic Kidney Disease by using Machine Learning Techniques, September 2018