RESEARCH ARTICLE OPEN ACCESS

Early Prediction of Lifestyle diseases using Machine Learning and Django Web Application Framework

SMD Adnan Sami*, P.Nikhil Kumar **,Pusuluru Rohith Kumar***,Divya S Patel ****,Preethi J*****, Dr. Hasan Hussain S******

*(Computer Science and Engineering/Presidency University, Bengaluru
Email: 201910102143@presidencyuniversity.in)
** (Computer Science and Engineering /Presidency University, Bengaluru
Email: 201910101341@presidencyuniversity.in)
*** (Computer Science and Engineering /Presidency University, Bengaluru
Email: 201910102004@presidencyuniversity.in)
**** (Computer Science and Engineering /Presidency University, Bengaluru
Email: 201910100434@presidencyuniversity.in)
***** (Computer Science and Engineering /Presidency University, Bengaluru
Email: 201910101359@presidencyuniversity.in)
****** (Computer Science and Engineering /Presidency University, Bengaluru
Email: hasan.hussain@presidencyuniversity.in)

----------**********----------

Abstract:

Most of the population is exaggerated by the everyday lifestyle-related diseases almost most of the nation globally counting India also. The daily adheres to we participate in encourage the emergence of lifestyle diseases. The primary origin of it is how everyone spends their life. It covers illnesses that everybody has heard of, such as heart disease, high blood pressure, chronic renal disease, and liver illness and many more. We also encountered a lot of these folks, many of whom passed away because they were unaware of their illness and did not receive treatment in a timely manner. In this project, we created a machine learning application for predicting chronic kidney disease and liver disease. This effort raises the individual's awareness of their health so that they may receive the necessary care in a timely manner, perhaps saving many lives. With the use of a web application, we applied a variety of machine learning models to predict diseases including CKD and liver illness. The high quality dataset is made available via the University of California, Irvine (UCI). Support Vector Machines (SVM), Logistic Regression (LR), Random Forest Classifier (RF), Naive Bayes Classifier(NBC), Decision Tree Classifier (DT), K-Nearest Neighbor(KNN), and Adaptive Boosting (AdaBoost) are a few examples of machine learning (ML) algorithms that are employed. The parameters Accuracy, Precision, Recall, F1-Score, and Support have been examined for all algorithms. Various exploratory data analysis techniques, including frequency distribution, feature correlation, feature distribution plots, and classification and prediction plots, were then applied. Algorithm performance evaluation and comparison studies are calculated, and the best accuracy and other particular metrics are found. With an accuracy of 98.75%, ADA Boost & SVM get the highest rating of all algorithms.

Keywords **—Machine Learning,Chronic Kidney Disease, Liver Disease, Predictions, Django Framework**

----------**********----------

I. INTRODUCTION

Unhealthy lifestyle habits are responsible for most of the day to day Lifestyle diseases. These diseases are often chronic in nature and tend to develop over a long period of time due to the cumulative effects of certain lifestyle factors. The major lifestyle diseases include Chronic Kidney Disease, Liver Disease, cardiovascular diseases, type 2 diabetes, obesity, certain types of cancer, and respiratory diseases. The rise in lifestyle diseases can be attributed to the rapid changes in our modern way of life, including sedentary lifestyles, poor dietary choices, tobacco and alcohol consumption, and increased stress levels. Cardiovascular diseases, such as heart disease and stroke, are the leading causes of death globally. The people who follow the unhealthy food are more affected because the food has high cholesterol and also lack of physical activity, tobacco use, and excessive alcohol consumption. Type 2 diabetes is another common lifestyle disease characterized by high blood sugar levels. It is closely linked to obesity, poor diet, and physical inactivity. The disease affects the body's ability to properly utilize insulin, a hormone responsible for regulating blood sugar levels. Obesity itself is considered a lifestyle disease and is a major risk factor for numerous health conditions. Sedentary lifestyles, excessive calorie intake, unhealthy eating habits, and lack of exercise contribute to the development of obesity. Certain types of cancer, such as lung, colorectal, and breast cancer, have been associated with lifestyle choices. Factors like tobacco use, poor diet, lack of physical activity, and excessive alcohol consumption increase the risk of developing these cancers. Preventing and managing lifestyle diseases requires a holistic approach that focuses on making healthy choices. Avoiding tobacco and excessive alcohol consumption, managing stress, and maintaining a healthy weight are key strategies for reducing the risk of lifestyle diseases. Early detection, regular medical check-ups, and appropriate medical interventions are also important for effective disease management.Promoting awareness about the importance of a healthy lifestyle, implementing public health policies to discourage unhealthy behaviours, and providing access to affordable healthcare are essential steps in combating lifestyle diseases and reducing their burden on individuals and society as a whole.

Several factors contribute to the development of lifestyle diseases. These factors can be categorized into modifiable and non-modifiable factors. Modifiable factors are those that individuals have control over and can modify to reduce their risk of developing lifestyle diseases. Non-modifiable factors, on the other hand, are beyond an individual's control. Here are some of the key factors affecting lifestyle diseases.

Unhealthy Diet or Poor dietary choices, such as consuming a diet high in saturated and transfats, sugar, salt, and processed foods, can increase the risk of developing lifestyle diseases. Inadequate intake of fruits, vegetables, whole grains, and lean proteins can also contribute to the development of these diseases. Lack of Physical Activity increase the risk of obesity, cardiovascular diseases, and other lifestyle diseases. Physical inactivity can lead to weight gain, poor cardiovascular health, and metabolic imbalances.Tobacco use, including smoking cigarettes and using smokeless tobacco products, is a major risk factor for various lifestyle diseases, including cancer, heart disease, and respiratory diseases. Excessive alcohol consumption can also contribute to liver disease, cardiovascular problems, and certain types of cancer.Being overweight or obese significantly increases the risk of developing lifestyle diseases, including diabetes, heart disease, certain types of cancer, and joint problems. Obesity is often the result of an imbalance between calorie intake and energy expenditure. Chronic stress can contribute to the development of lifestyle diseases. Excessive stress levels can lead to unhealthy coping mechanisms such as overeating, smoking, and alcohol abuse, which can increase the risk of developing diseases like obesity, heart disease, and substance abuse disorders. While lifestyle choices

play a significant role in the development of lifestyle diseases, genetic factors also contribute. Some individuals may have a genetic predisposition that makes them more susceptible to certain diseases. However, genetic factors alone do not determine disease development, and lifestyle choices still play a crucial role.

Socioeconomic factors, such as income level, education, and access to healthcare, can influence the prevalence and management of lifestyle diseases. Limited access to healthy food options, recreational facilities, and healthcare resources can contribute to the development and progression of these diseases.It is important to note that while some factors are beyond an individual's control, such as genetic predisposition, taking steps to modify modifiable factors like diet, physical activity, and tobacco/alcohol use can have a significant impact on reducing the risk of lifestyle diseases and improving overall health.

In this paper, we have worked on two disease prediction methodologies such Chronic Kidney Disease and Liver Disease and developed the web application to predict the same.

II. RELATED WORKS

Due to the dynamic and covert character of CKD and liver disease in its early stages as well as patient anomaly, it is crucial to accurately forecast the disease's progression. Depending on the stage, CKD requires medical therapy. If not, it is critical to identify the infection's organizational structure since it provides some clues. It supports the certainty of essential prayers and treatments.An important application area for intellectual intelligence systems is medical care. In order to uncover hidden information from the extensive patient medical and treatment dataset that doctors regularly acquire from patients in order to get knowledge about the symptomatic data and to carry out accurate treatment plans, data mining may then play a significant role. Data mining is a technique for locating hidden information in a big dataset. Data mining techniques are interconnected and widely applied in many situations and fields. We can anticipate, categorize, filter, and cluster data using data mining techniques.

In order to diagnose and treat CKD, Qin et al. offered data assertion and a sample diagnostic. Logistic regression, random forest, support vector machine, K-nearest neighbour, naive Bayes classifier, and feed-forward neural network are six classifier algorithms that are used to determine the accuracy of a diagnosis. On a dataset of 40000 cases, Vasquez-Morales et al. created a neural network model with a 95% accuracy for risk prediction of the development of chronic kidney disease. For the purpose of calculating the patient's risk, Chen et al. used SVM and KNN models. Naive Bayes had the highest accuracy of 91%, according to Padmanaban et al.'s study on early liver detection in diabetes patients using the Weka tool. Gunarathne and colleagues developed a model using several machine learning classifiers' algorithms and analyses, and they forecasted the multiclass decision forest approach has an accuracy of 99.1% and was the best match for the CKD dataset.*Polat et al.* used algorithm for CKD prediction with Two-approach Wrapper and filter with the SVM algorithm. The results of all techniques were compared and it was found that SVM gave the highest accuracy with filtered subset evaluator, i.e. 98.5%. *Sujata Drall et al.* worked on CKD prediction using Naïve Bayes and K-Nearest Neighbor. *Almasoud et al*. they applied the filter feature selection method on attributes and the gradient boosting Algorithm achieved the highest accuracy of 99.1%.

III. PROPOSED METHODOLOGY

The main objective of the project is to predict Chronic Kidney Disease and Liver Disease as shown in Figure 1 is based on full features and important features of CKD dataset and liver dataset using Machine Learning Algorithms. In this perception, Logistic Regression, Support Vector Classifier, Random Forest Classifier,

Decision Tree Classifier, Naive Bayes Classifier, K-Nearest Neighbour, Adaptive Boosting (ADABoost) and Artificial Neural Network are applied. Performance Evaluation and comparison study of algorithms are calculated and finding the best accuracy and other specific metrics. Exploratory data Analysis on CKD Dataset.

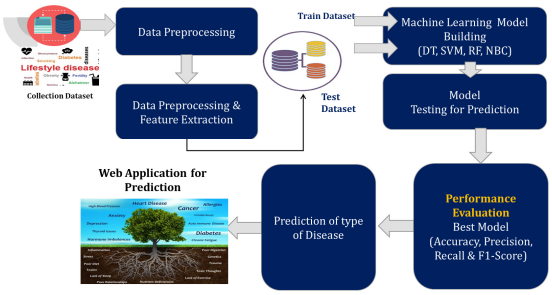

Figure 1: Proposed System of Lifestyle Disease prediction

IV SYSTEM IMPLEMENTATION

The purposed System Implementation developed for lifestyle disease predictions by developing and positioning on-going support and maintenance of the system within the Performing Organization (the transition). The implementation can classified into different modules of project and are listed as Data Collection, Data Preprocessing, Feature Selection / Extraction, Model Training and Testing, Machine Learning Model Deployment, Performance Evaluation, Comparison Study of The Model and some Exploratory Data Analysis (EDA) on dataset.

A. Chronic Kidney Disease Prediction

1) Dataset:In this project we have used the dataset based on the CKD patients as shown in the figure 1. Many researchers had also used this dataset. This dataset is being provided by the UC Irvine Machine Learning Repository and it is available on the UCI website. This dataset contains 400 patients database and14 attributes with 1 target attribute. The target attribute has labelled in two-class to represent Diseased or Not. The dataset was collected from various hospitals in 2015. It contains also missing value.

Attribute	Value Used
Age	Discrete Integer Values
Blood Pressure	Discrete Integer Values
Albumin	Nominal Values
Red Blood cells	Nominal Values(Normal, Abnormal)
Pus cell	Nominal Values(Normal, Abnormal)
Pus cells clumps	Nominal Values(Present, Not-Present)
Serum creatinine	Numeric Values
Haemoglobin	Numeric Values
White blood cell count	Discrete Integer Values
Red blood cell count	Numeric Values
Anaemia	Nominal Values(Yes, No)
Classification	Nominal Values(CKD, Not CKD)
Appetite	Nominal Values(Good, Poor)
Packed cell volume	Discrete Integer Values

Table 1: CKD Attributes and values

2) Feature Selection:Feature selection is needed for trained each machine learning classifier because without removing unnecessary attributes from the dataset result may be affected. The classifier algorithm with feature selection gives better performance and reduces the execution time of the model. For this process, three different feature selection methods were used in this research.

3) ***Machine Learning Algorithms:***Machine Learning Classification algorithms are deployed and performed best for CKD prediction and provided insight evaluation model values on the dataset of kidney disease. The confusion Matrix of each algorithm provides the detailed view of values over the Dataset.

B. Liver Disease Prediction

1) Dataset:This data set contains 416 liver patient records and 167 non liver patient records collected from North East of Andhra Pradesh, India. The "Dataset" column is a class label used to divide groups into liver patient (liver disease) or not (no disease). This data set contains 441 male patient records and 142 female patient records. Any patient whose age exceeded 89 is listed as being of age "90".

Age	Gender	Total_Bili	Direct_Bil	Alkaline_I	Alamine_A	Aspartate	Total_Pro	Albumin	Albumin_	Dataset
65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	1
62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
62	Male	7.3	4.1	490	60	68	7	3.3	0.89	1
58	Male	1	0.4	182	14	20	6.8	3.4	1	1
72	Male	3.9	2	195	27	59	7.3	2.4	0.4	1
46	Male	1.8	0.7	208	19	14	7.6	4.4	1.3	1
26	Female	0.9	0.2	154	16	12	7	3.5	1	1
29	Female	0.9	0.3	202	14	11	6.7	3.6	1.1	1
17	Male	0.9	0.3	202	22	19	7.4	4.1	1.2	2
55	Male	0.7	0.2	290	53	58	6.8	3.4	1	1
57	Male	0.6	0.1	210	51	59	5.9	2.7	0.8	1
72	Male	2.7	1.3	260	31	56	7.4	3	0.6	1
64	Male	0.9	0.3	310	61	58	7	3.4	0.9	2
74	Female	1.1	0.4	214	22	30	8.1	4.1	1	1
61	Male	0.7	0.2	145	53	41	5.8	2.7	0.87	1
25	Male	0.6	0.1	183	91	53	5.5	2.3	0.7	2
38	Male	1.8	0.8	342	168	441	7.6	4.4	1.3	1
33	Male	1.6	0.5	165	15	23	7.3	3.5	0.92	2
40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
51	Male	2.2	1	610	17	28	7.3	2.6	0.55	1
51	Male	2.9	1.3	482	22	34	7	2.4	0.5	1
62	Male	6.8	3	542	116	66	6.4	3.1	0.9	1
40	Male	1.9	1	231	16	55	4.3	1.6	0.6	1

Figure 2: Liver Dataset

2) **Machine Learning Approaches:**The machine learning algorithms are used to classify the model and various algorithms are used during the study of liver disease predictionand made the comparison study at the end to find the best fitted model and finally predicting that, patient is having Liver disease or No Liver disease. Here we have discussed few algorithms we used in our study.

V.EXPERIMENTAL RESULTS

The experiment is carried on two different datasets, one is on CKD dataset and second one is on Liver disease dataset. Here we have used machine leaning algorithms for our study as prediction algorithms. The predictions are shown by using the web applications by giving the input to the machine learning implemented models. Here we input the values for both apps and get the results.

The implementation of CKD is as shown in Figure 3. Here we need to provide the inputs such as Specific Gravity, Hypertension level, diabetes, Albumin, Appetite, RBC and Pus Cell.

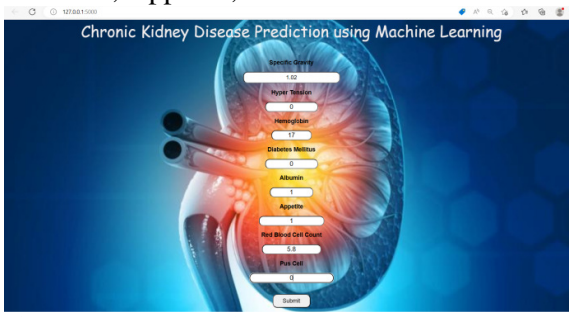

Figure 3: CKD Prediction

Figure 4: CKD prediction with no CKD

As above figure 4 shows patient has no CKD symptoms as per the inputs provided.

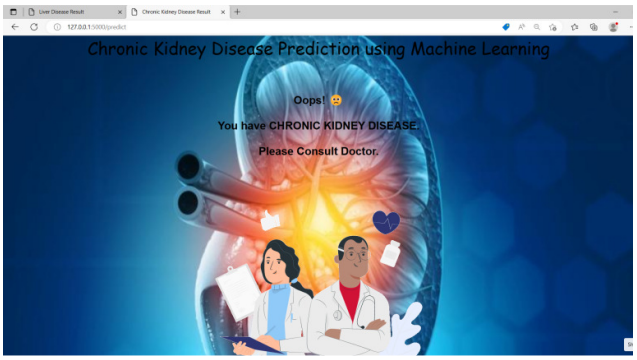

Figure 5: CKD Prediction with patient has CKD

The figure 5 shows that patient is having CKD and need the further assistance from doctors.

The second lifestyle disease is Liver disease prediction as shown in figure 6.

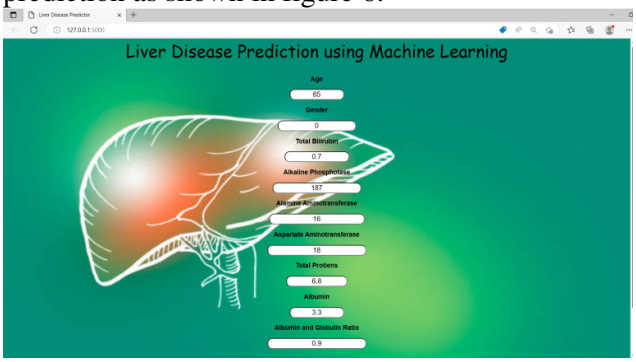

Figure 6: Liver Disease Prediction using ML

The parameters are Age, gender, Alkaline Phosphate, Aspartate, Total proteins, albumin and Globulin Ratio. The next Figure 7 shows that the patient is not having Liver disease.

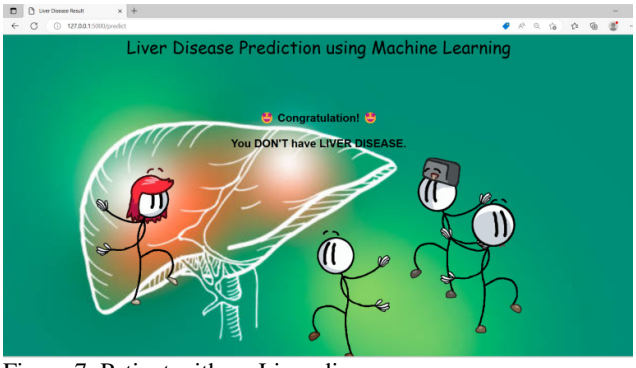

Figure 7: Patient with no Liver disease

Figure 8 shows the person is having the Liver disease and needs immediate action.

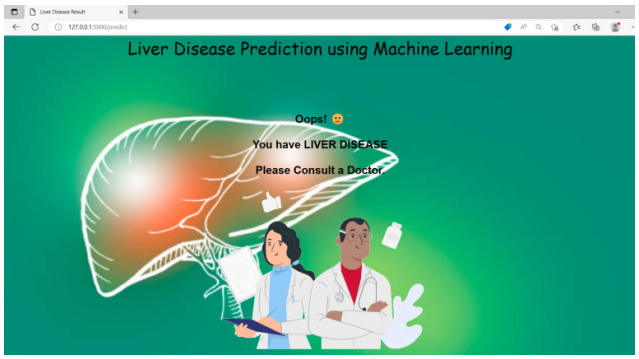

Figure 8: Patient with Liver disease

The algorithm accuracy can be found by using various parameters such as Precision, Recall, F1 score and Accuracy.
The experimental values as show as in Figure 9 and figure 10. As per the result for CKD, Random Forest is having the highest accuracy of 82%. In case of Liver disease Random Forest and AdaBoost is having 98.75% of accuracy.

The web application using Django provides the best platform to check the lifestyle diseases. The results of the predictions are good enough for the patients to check with the real time lab testing results.

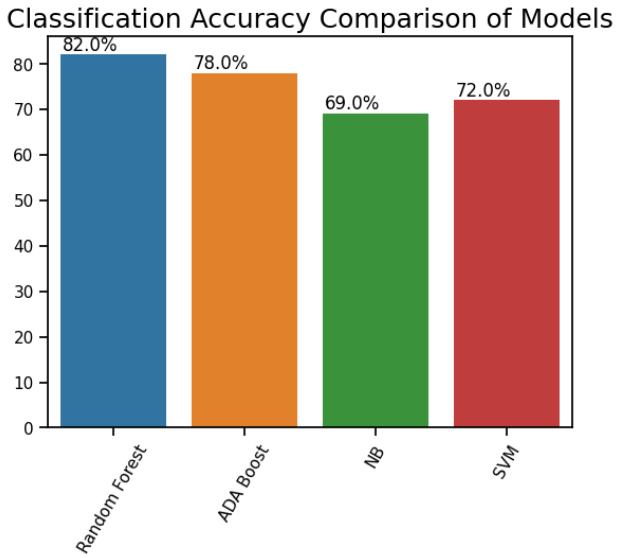

Figure 9: CKD ML algorithm comparison

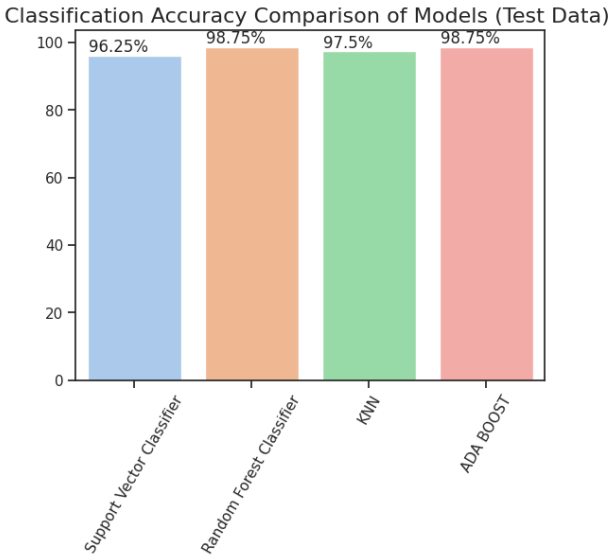

Figure 10: Liver Disease Algorithm Comparison

IV. CONCLUSIONS

Various machine learning algorithms are applied on CKD& Liver disease prediction. We have made a brief literature survey on numerous articles to get the get of research gap and the area to enhance the work .We have totally used four classification algorithm , among that Random Forest and AdaBoost proving the best accuracy of 99% in evaluating and predicting both CDK & No-CKD and liver disease or not analysis. We have used the concept of exploratory data analysis to plot various graphical representations. We developed a web application to demonstrate the Chronic Kidney Disease and Liver Disease.

REFERENCES

[1] Linta Antony, Sami Azam, Ryana Quadir "A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction", *IEEE ACCESS, PP 126481 – 126501 VOLUME 9 2021*

[2] Davide Chicco 1, Christopher A. Lovejoy 2,3, And Luca Oneto, "A Machine Learning Analysis of Health Records of Patients With Chronic Kidney Disease at Risk of Cardiovascular Disease", IEEE ACCESS, PP 165133 – 165144 VOLUME 9 2021

[3] Jiongming Qin 1, Lin Chen 2, Yuhua Liu 1, Chuanjun Liu 2, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease", *IEEE ACCESS, PP 20991 – 21002 VOLUME 8 2020*

[4] Bilal Khan, Rashid Naseem, Fazal Muhammad, Ghulam Abbas, And Sunghwan Kim 5 "An Empirical Evaluation Of Machine Learning Techniques for Chronic Kidney Disease Prophecy" ", *IEEE ACCESS, PP 55012– 55022165144 VOLUME 8 2020.*

[5] Jivan Parab , Marlon Sequeira , Madhusudan Lanjewar , Caje Pinto , And Gourish Naik, "Backpropagation Neural Network-Based Machine Learning Model for Prediction of Blood Urea and Glucose in CKD Patients", *IEEE Journal Translational Engineering in Health & Madicines, PP 4900608– 4900608 VOLUME 9 2021.*

[6] Md. Rashed-Al-Mahfuz , Abedul Haque , Akm Azad3, Salem A. Alyami , Julian M. W. Quinn, And Mohammad Ali Moni, "Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening*", IEEE Journal Translational Engineering in Health & Madicines, PP 4900511 4900511 VOLUME 9 2021.*

[7] Pankaj Chittora , Sandeep Chaurasia, Prasun Chakrabarti, Tulika Chakrabarti, Zbigniew Leonowicz, "Prediction of Chronic Kidney Disease - A Machine

[8] Learning Perspective", IEEE ACCESS, PP 165133 – 165144 VOLUME 9 2021

[9] Ramesh Chandra Poonia , Mukesh Kumar Gupta , Ibrahim Abunadi, Amani Abdulrahman Albraikan, Fahd N. Al-Wesabi , Manar Ahmed Hamza and Tulasi B, "Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease", *Healthcare 2022, 10, 371.*

[10] Shahinda Mohamed Mostafa Elkholy , Amira Rezk , And Ahmed Abo El Fetoh Saleh, "Early Prediction of Chronic Kidney Disease Using Deep Belief Network", *IEEE ACCESS, PP 135542 – 135549 VOLUME 9 2021.*

[11] Rezuanur Rahman Dip, Sarker Md. Nasrullah, " A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms", *EAI Endorsed Transactions on Pervasive Health and Technology 10 2021 - 11 2021 Volume 7 | Issue 29 | e1*

[12] Pallavi Sharma, Dr. Gurmanik Kaur, " A survey on Support Vector Machine in Chronic Kidney Disease Prediction", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 8, Issue 7, July 2021

[13] Anna Laura Herzog, Holger K. von Jouanne-Diedrich, Christoph Wanner, "COVID-19 and the kidney: A retrospective analysis of 37 critically ill patients using machine learning", *PLOS ONE, 2021*

[14] Mafizur Rahman1, Linta I slam2, Masud Rana3, "A Predictive Analysis of Chronic Kidney Disease by Exploring Important Features", *International Journal of Computing and Digital Systems, 2022.*

[15] Mahima Jayanth1, Sangeetha N2, Radhika Kulkarni3, "A Generic Real Time Application for CKD and its Stages Prediction", *International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 06 | June 2021.*

[16] Muhammad Nabeel1, Shumaila Majeed 2, Mazhar Javed Awan1, " Review on Effective Disease Prediction through Data Mining Techniques", *International Journal on Electrical Engineering and Informatics - Volume 13, Number 3, September 2021*

[17] Yen-Ling Chiu1, Mao-Jhen Jhou, Tian-Shyug Lee, Chi-Jie Lu, "Health Data-Driven Machine Learning Algorithms Applied to Risk Indicators Assessment for Chronic Kidney Disease", *Risk Management and Healthcare Policy, pp- 4401- 4412 2021*

[18] Siddhartha Krishna, Maanvi Keesara and Mary Subaja EasyChair, "Comparative Study on Chronic Kidney Disease Detection Using Tree-Based Models", *2022*

[19] Cheng Li , "Risk prediction of chronic disease using machine learning and rebalancing methods"

[20] Donghwan Yun1,2, MD; Semin Cho2, MD; Yong Chul Kim2, MD, PhD; Dong Ki Kim2, "Use of Deep Learning to Predict Acute Kidney Injury After Intravenous Contrast Media Administration: Prediction Model Development Study", *JMIR Med Inform 2021 | vol. 9 | iss. 10 | e27177 | p. 1*

[21] Sunil Belur Nagaraj, PhD1*, Michelle J Pena, PhD1, Wenjun Ju, PhD2, Hiddo L Heerspink, "Machine Learning based Early Prediction of End-stage Renal Disease in Patients with Diabetic Kidney Disease using Clinical Trials Data".

