RESEARCH ARTICLE OPEN ACCESS

Chronic Kidney Disease (Ckd) Prediction Using Data Mining Techniques

Mrs J Sarada MCA,M.Tech,M.Phil,(Ph.D) [1], Jashuva Nanagari [2]

[1] Associate Professor, Department of Computer Applications

[2] Student, Department of Computer Applications

[1],[2] Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRACT

In the past decade, the rapid growth of digital data and global accessibility through the modern internet has seen a massive rise in machine learning research. In proportion to it, the medical data has also seen a massive surge of expansion. With the availability of structured clinical data, researchers have attracted scores to study clinical disease detection automation with machine learning and data mining. Chronic Kidney Disease (CKD), also known as the renal disorder, has been such a field of study for quite some time now. Therefore, our research aims to study the automated detection of chronic kidney disease using several machine learning classifiers with clinical data. The purpose of this research work is to diagnose kidney disease using a number of machine learning algorithms such as the Support Vector Machine (SVM) and the Bayesian Network (BN) and to select the most effective one to assess the extent of CKD patients. The amount of expertise in the medical field in relation to CKD is limited. Many patients have to wait a long to get their test results. The experience of medical staff is declining in value. Upon retirement, new employees replace them. It helps professional doctors or medical staff in their diagnosis of CKD. This paper's primary purpose is to present a clear view of Chronic Kidney Disease (CKD), its symptoms, and the process of early detection that may help humanity be safe from this life-threatening disease.

Keywords: - CKD, SVM, BN

I. INTRODUCTION

A global health problem which is steadily growing is Chronic kidney disease (CKD). It is a chronic condition associated with increased morbidity and mortality, a high risk of many other diseases including cardiovascular disease, and high health care costs. Over two million people worldwide receive dialysis or kidney transplant treatment to stay alive, yet this number may represent only 10% of people who need treatment to live [9]. The majority of the 2 million people who receive treatment for kidney failure are in only five relatively wealthy countries, which represent 12% of the global population. By comparison, only 20% of the world's population is treated in about 100 developing countries, and they represent almost half the global population. Annually, more than one million people in 112 lower-income countries die from untreated kidney failure, due to the huge financial burden of dialysis or kidney transplantation treatment [9].

Thus, there is significant importance in the early detection, controlling, and managing of the disease. It is necessary to predict the progression of CKD with reasonable accuracy because of its dynamic and covert nature in the early stages, and patient heterogeneity. CKD is often described by severity stages. Clinical decisions are influenced by the stage, whether a patient is progressing, and the rate of progression. Also, defining the disease stage is quite crucial as it gives several indications that support the determination of required intervention and treatments.

Therefore, data mining can play a major role in extracting hidden data from the large patient medical and clinical dataset that physicians frequently collect from patients to obtain insights about the diagnostic information, and to implement precise treatment plans. Data mining can be defined as the process of extracting hidden data from a large dataset. Data mining techniques are applied and used widely in various contexts and fields. With data mining techniques we could predict, classify, filter and cluster data. The goal or prediction attribute refers to the algorithm processing of a training set containing a set of attributes and outcomes.

Machine learning algorithms have been used to predict and classify in the healthcare field. Yu et al. [17] have used the Support Vector Machine Algorithm to classify and predict diabetes and pre-diabetes patients, and the results show that SVM is useful to classify patients with common diseases. Similarly, Magnin et al. [19] have classified Alzheimer's disease by using a Support Vector Machine (SVM) to analyze whole-brain anatomical magnetic resonance imaging (MRI) for a set of patients, and the results shows that SVM is a promising approach for Alzheimer's disease early detection. Dessai et al. [18] have done heart disease prediction using the Probabilistic Neural Network Algorithm, Decision tree Algorithm, and Naïve Bayes Algorithm, and PRNN provides the best results compared with other algorithms for heart disease prediction. Cao et al. [20] have done prediction of HBV-induced liver cirrhosis using the Multilayered Perceptron (MLP) Algorithm and the results shows that the MLP classifier gives satisfactory prediction outputs for liver disease, mostly in HBV-related liver cirrhosis patients.

II. LITERATURE SURVEY

Primary Care Detection of Chronic Kidney Disease in Adults with Type-2 Diabetes

This US, multicenter, observational study assessed the CKD prevalence in adult patients with type-2 diabetes mellitus (T2DM) and characterized the proportion of detected and undiagnosed CKD in the primary care setting using the following: a clinician survey; a patient physical exam and medical history; a single blood draw for estimated glomerular filtration rate (eGFR) and glycosolated hemoglobin (HbA1c); urine dipstick for protein; urine albumin-creatinine ratio (ACR); two patient quality of life questionnaires; and a 15-month medical record review. The study consisted of 9339 adults with T2DM and 466 investigator sites. Of the 9339 enrolled, 9307 had complete data collection for analysis. The 15-month retrospective review showed urine protein, urine ACR, and eGFR testing were not performed in 51.4%, 52.9% and 15.2% of individuals, respectively. Of the 9307 patients, 5036 (54.1%) had Stage 1–5 CKD based on eGFR and albuminuria; however, only 607 (12.1%) of those patients were identified as having CKD by their clinicians. Clinicians were more successful in diagnosing patients with Stage 3–5 CKD than Stages 1 and 2. There were no differences in clinicians' likelihood of identification of CKD based on practice setting, number of years in practice, or self-reported patients seen per week. Awareness or patient self-reported CKD was 81.1% with practitioner detection versus 2.6% in the absence of diagnosis. Primary care of T2DM demonstrates recommended urine CKD testing is underutilized, and CKD is significantly under-diagnosed. This is the first study to show CKD detection is associated with awareness.

The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease

This randomised double-blind trial included 9270 patients with chronic kidney disease (3023 on dialysis and 6247 not) with no known history of myocardial infarction or coronary revascularisation. Patients were randomly assigned to simvastatin 20 mg plus ezetimibe 10 mg daily versus matching placebo. The key prespecified outcome was first major atherosclerotic event (non-fatal myocardial infarction or coronary death, non-haemorrhagic stroke, or any arterial revascularisation procedure). All analyses were by intention to treat. This trial is registered at ClinicalTrials.gov, NCT00125593, and ISRCTN54137607.

Population Health for CKD and Diabetes: Lessons From the Indian Health Service

Despite extensive clinical guidelines, innovative efforts to improve care, and well-funded efforts to raise awareness, limited progress has been made in reducing the burden of kidney disease in the United States, and the prevalence continues to increase worldwide. The Indian Health Service and the Centers for Disease Control and Prevention recently reported a 54% decrease in the incidence of kidney failure among American Indian and Alaska Native people with diabetes. This decrease in end-stage renal disease incidence was associated with a population health approach to diabetes care based in the community and the primary clinical setting. The effort focused on integrating better care for kidney disease within the context of routine diabetes care. Although the American Indian population and the Indian Health Service may be unfamiliar to many clinicians and health

system administrators, the demonstration that simple evidence-based interventions implemented in a comprehensive and consistent way can reduce the burden of end-stage renal disease suggests that population-based approaches to chronic disease offer significant potential benefits. Large pragmatic trials may offer the best way to rigorously test this hypothesis.

Practical Approach to Detection and Management of Chronic Kidney Disease for the Primary Care Clinician

A panel of internists and nephrologists developed this practical approach for the Kidney Disease Outcomes Quality Initiative to guide assessment and care of chronic kidney disease (CKD) by primary care clinicians. Chronic kidney disease is defined as a glomerular filtration rate (GFR) <60 mL/min/1.73 m(2) and/or markers of kidney damage for at least 3 months. In clinical practice the most common tests for CKD include GFR estimated from the serum creatinine concentration (eGFR) and albuminuria from the urinary albumin-to-creatinine ratio. Assessment of eGFR and albuminuria should be performed for persons with diabetes and/or hypertension but is not recommended for the general population. Management of CKD includes reducing the patient's risk of CKD progression and risk of associated complications, such as acute kidney injury and cardiovascular disease, anemia, and metabolic acidosis, as well as mineral and bone disorder. Prevention of CKD progression requires blood pressure <140/90 mm Hg, use of angiotensin-converting enzyme inhibitors or angiotensin receptor blockers for patients with albuminuria and hypertension, hemoglobin A1c ≤7% for patients with diabetes, and correction of CKD-associated metabolic acidosis. To reduce patient safety hazards from medications, the level of eGFR should be considered when prescribing, and nephrotoxins should be avoided, such as nonsteroidal anti-inflammatory drugs. The main reasons to refer to nephrology specialists are eGFR <30 mL/min/1.73 m(2), severe albuminuria, and acute kidney injury. The ultimate goal of CKD management is to prevent disease progression, minimize complications, and promote quality of life.

III. EXISTING SYSTEM

To build a model using data mining techniques to predict if a patient does indeed have CKD by reviewing and analyzing symptoms and various health parameters; Using Data Mining tools to classify those data and compare the results acquired through different techniques. As a result, a preeminent health concern, such as chronic kidney disease. And renal failure is not diagnosed in this population. As chronic kidney disorder often does not show any symptoms. This consequently has flourished the private sector but at the price of high medical costs. They worked with more than 5 attributes. The classification algorithm's performance has been compared based on accuracy, precision, and total execution time for the prediction of Chronic Kidney disease.

Disadvantages

- Low performance
- Low efficiency

IV. PROPOSED SYSTEM

Chronic Kidney Disease (CKD), also known as the renal disorder, has been such a field of study for quite some time now.The kidney is essential for the filtering and purification process of our blood. Without at least one functioning kidney, death is imminent and inevitable within a few days. Therefore, our research aims to study the automated detection of chronic kidney disease using several machine learning classifiers with clinical data. The purpose of this research work is to diagnose kidney disease using a number of machine learning algorithms such as the Support Vector Machine (SVM) and the Bayesian Network (BN) and to select the most effective one to assess the extent of CKD patients. The amount of expertise in the medical field in relation to CKD is limited. Many patients have to wait a long to get their test results. The experience of medical staff is declining in value. Upon retirement, new employees replace them. It helps professional doctors or medical staff in their diagnosis of CKD. In this project primary purpose is to present a clear view of Chronic Kidney Disease (CKD), its symptoms, and the process of early detection that may help humanity be safe from this life-threatening disease.

Advantages

- Increases accuracy
- Improves performance
- Improves efficiency

V. IMPLEMENTATION

1. Data Collection
2. Classification
3. Clustering
4. Prediction

➢ **Data Collection:**

In data collection phase different student's information collected as in the form of data sets. The information includes their daily activities, results of each exam, teaching performance and health issues etc. from the relevant sources can analyses the student performance.

➢ **Classification:**

Classification is the most common data mining technique, which employs a set of pre classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network - based classification algorithms. The data classification process involves learning and classification. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier - training algorithm uses these pre -classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

➢ **Clustering:**

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes.
Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification.

➢ **Prediction**

Predicting students' performance in a selected education academy based on previously recorded students' behavior and activities. In this module can predict which students are going to dropout in the upcoming examinations in earlier and also identify the reason for dropout in educations. Regression technique can be adapted for prediction. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what want to predict.

VI. RESULTS

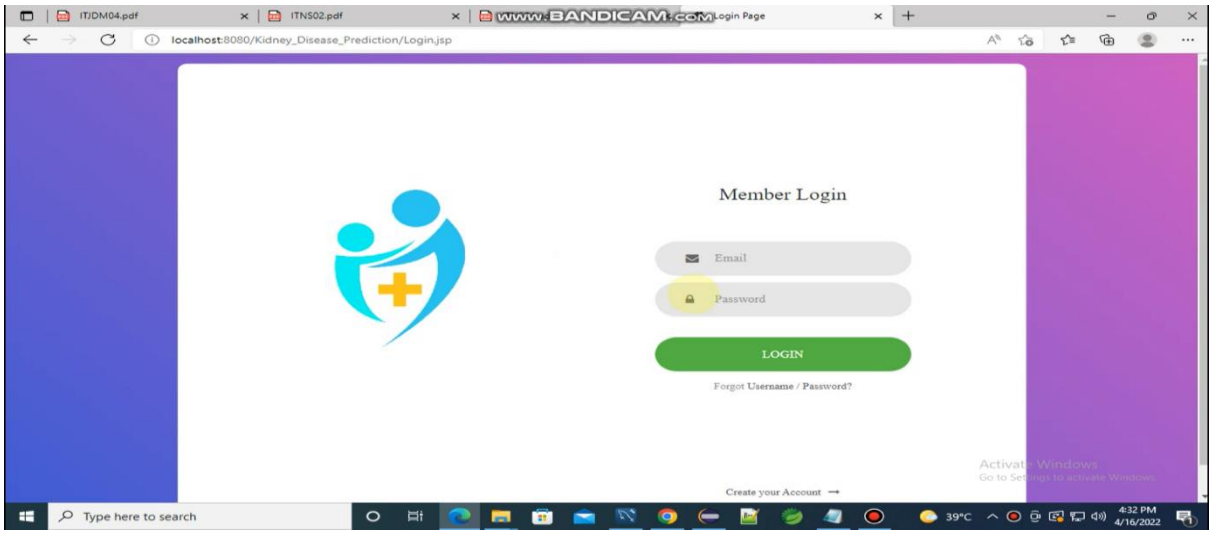


Fig 1 Home Page

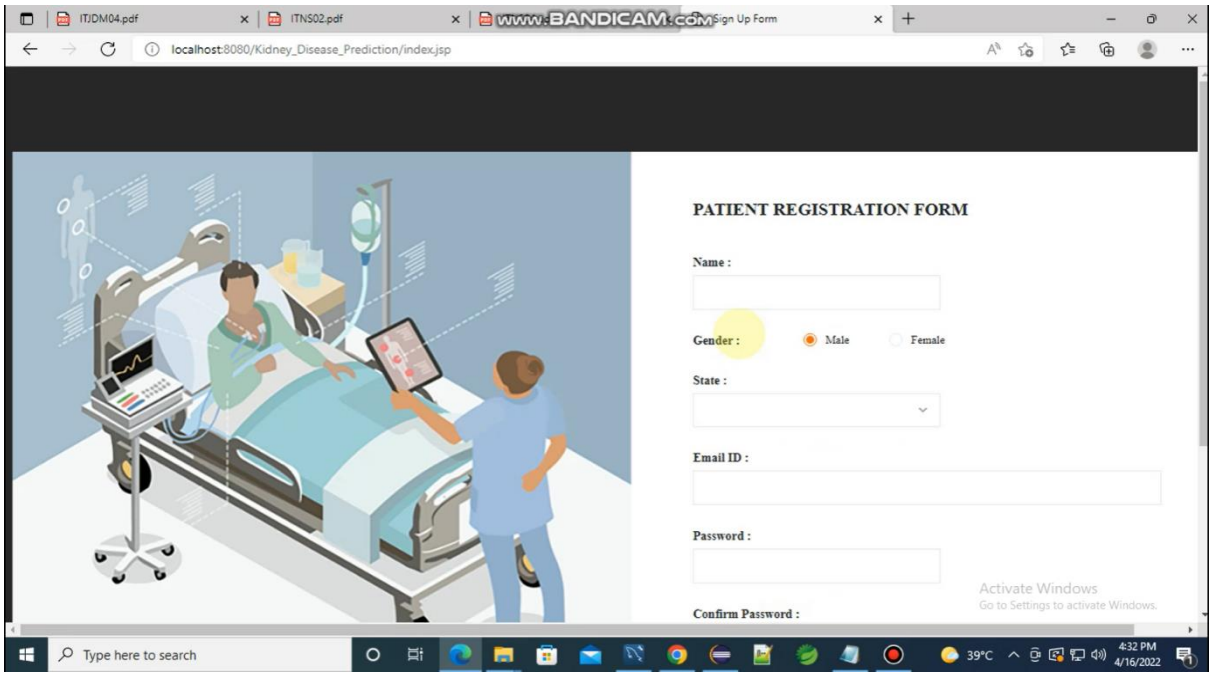


Fig 2: Registration Page

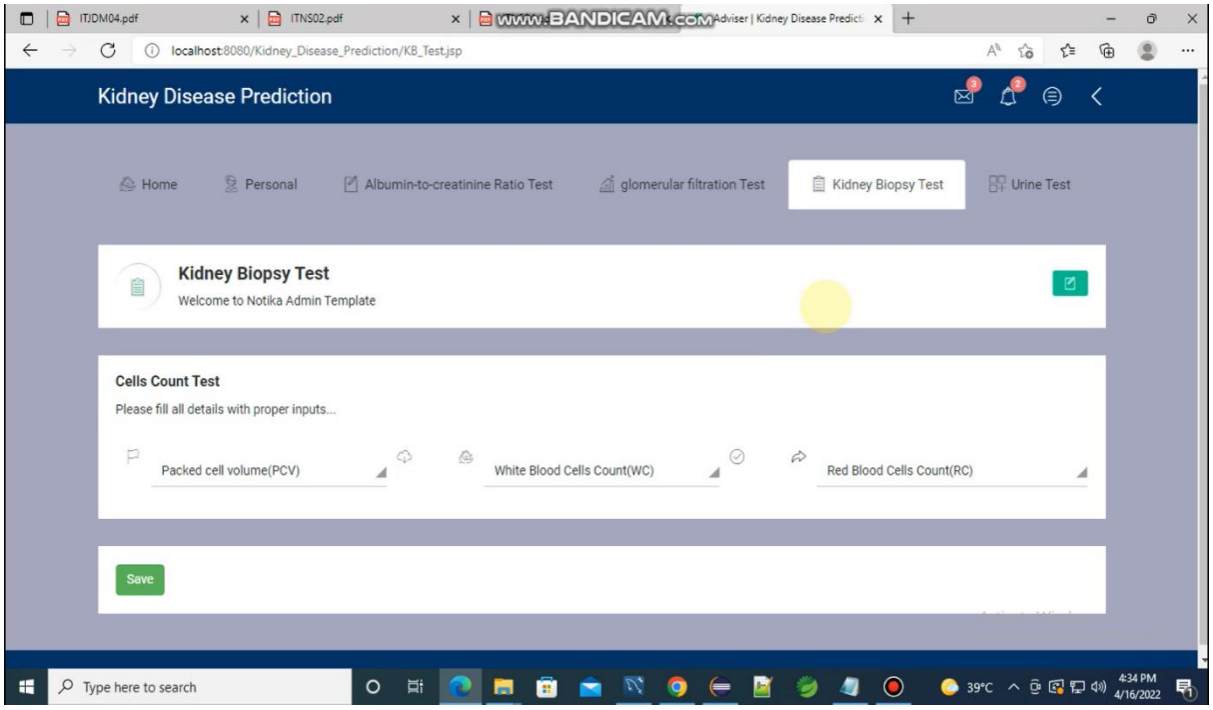


Fig 3: Prediction

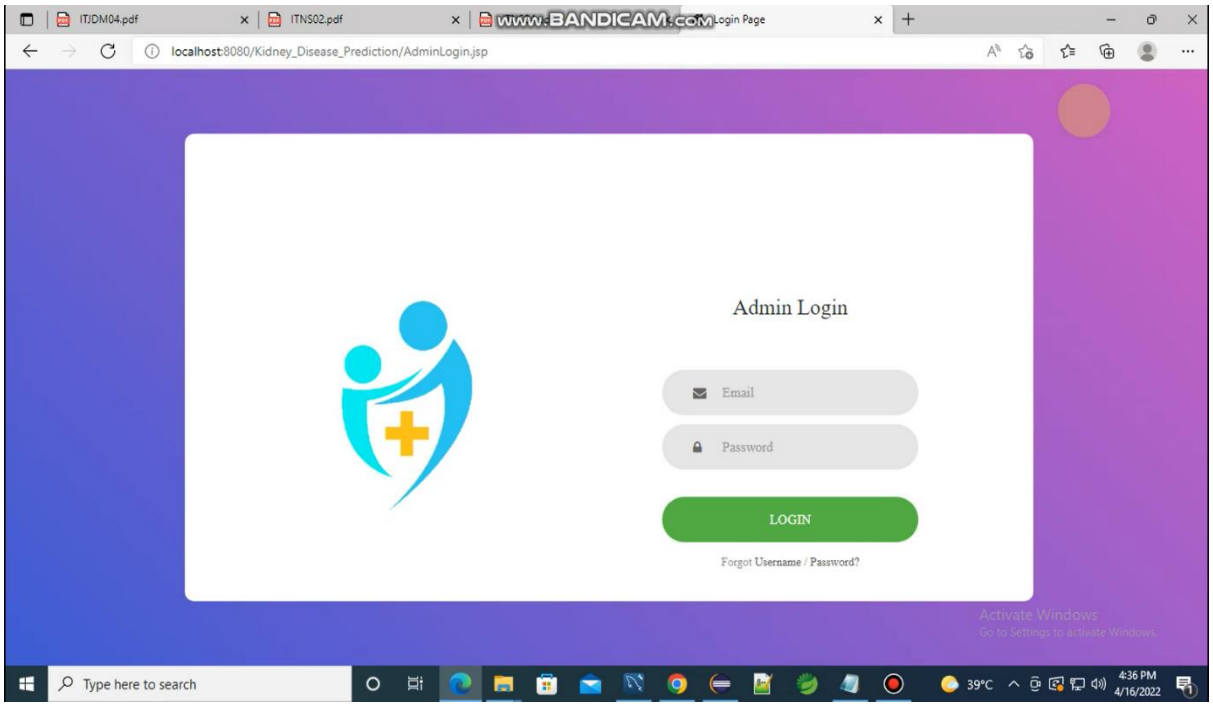
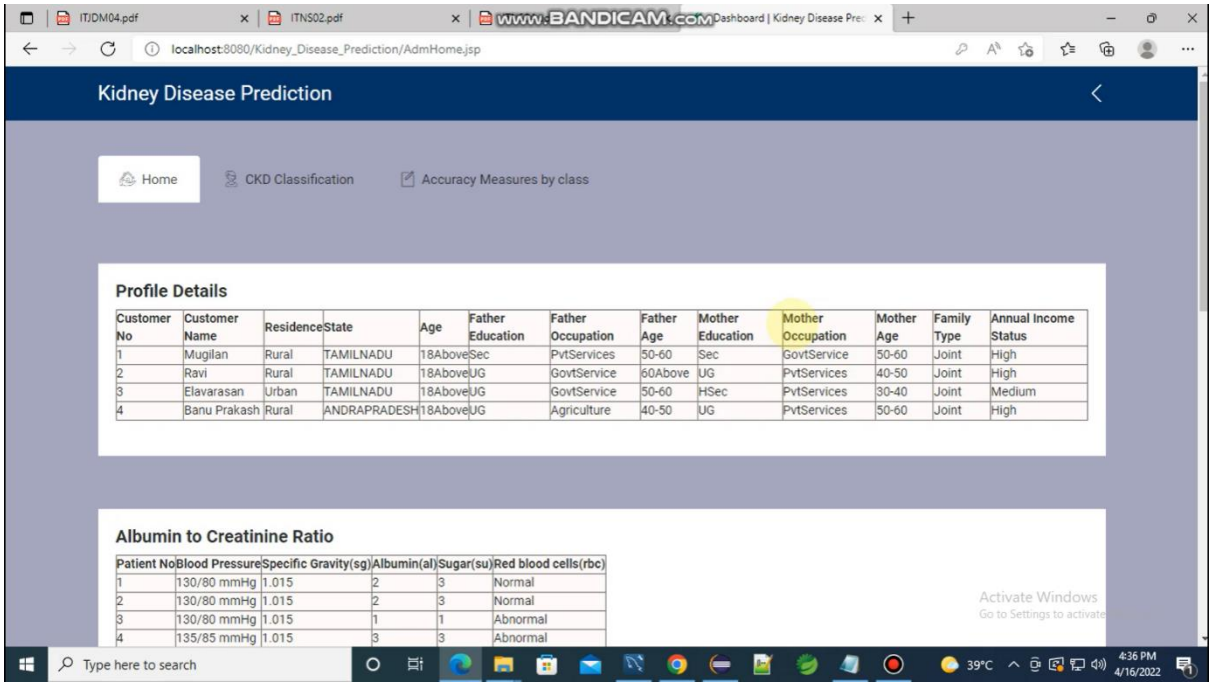


Fig 4: Admin page



Customer No	Customer Name	Residence	State	Age	Father Education	Father Occupation	Father Age	Mother Education	Mother Occupation	Mother Age	Family Type	Annual Income Status
1	Mugilan	Rural	TAMILNADU	18Above	Sec	PvtServices	50-60	Sec	GovtService	50-60	Joint	High
2	Ravi	Rural	TAMILNADU	18Above	UG	GovtService	60Above	UG	PvtServices	40-50	Joint	High
3	Elavarasan	Urban	TAMILNADU	18Above	UG	GovtService	50-60	HSec	PvtServices	30-40	Joint	Medium
4	Banu Prakash	Rural	ANDRAPRADESH	18Above	UG	Agriculture	40-50	UG	PvtServices	50-60	Joint	High

Patient No	Blood Pressure	Specific Gravity(sg)	Albumin(al)	Sugar(su)	Red blood cells(rbc)
1	130/80 mmHg	1.015	2	3	Normal
2	130/80 mmHg	1.015	2	3	Normal
3	130/80 mmHg	1.015	1	1	Abnormal
4	135/85 mmHg	1.015	3	3	Abnormal

Fig 5 : Prediction Page

VII. CONCLUSION

In conclusion, the use of data mining methods for data risk analysis is very important in the health sector because it first gives the power to fight diseases and therefore saves people's lives by reversing treatment. In this work, we used a number of learning algorithms to assess patients with chronic renal failure (ckd) and patientsclassification has proven its performance in predicting the best results in terms of accuracy and minimum execution time.

REFERENCES

1. Centers for Disease Control and Prevention. National Chronic Kidney Disease Fact Sheet. In: U.S. Department of Health and Human Services, ed. Atlanta, GA: Centers for Disease Control and Prevention; 2017.
2. Szczech LA, Stewart RC, Su HL, et al. Primary care detection of chronic kidney disease in adults with type-2 diabetes: the ADD-CKD Study (awareness, detection and drug therapy in type 2 diabetes and chronic kidney disease). PloS one 2014;9:e110535.
3. United States Renal Data System. 2016 USRDS annual data report: Epidemiology of kidney disease in the United States. In. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2016.
4. Golestaneh L, Alvarez PJ, Reaven NL, et al. All-Cause Costs Increase Exponentially with Increased Chronic Kidney Disease Stage. American Journal of Managed Care 2017;23:S0.
5. USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. Bethesda, MD; 2013.
6. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C-y. Chronic Kidney Disease and the Risks of Death, Cardiovascular Events, and Hospitalization. New England Journal of Medicine 2004;351:1296-305.
7. Matsushita K, Coresh J, Sang Y, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. The Lancet Diabetes & Endocrinology;3:514-25.
8. Narva A. Population Health for CKD and Diabetes: Lessons From the Indian Health Service. American Journal of Kidney Diseases.
9. National Kidney Foundation. K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification. Am J Kidney Dis 2002;39:S1-S266.
10. Vassalotti JA, Centor R, Turner BJ, Greer RC, Choi M, Sequist TD. Practical Approach to Detection and Management of Chronic Kidney Disease for the Primary Care Clinician. The American Journal of Medicine;129:153-62.e7.
11. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018;71:e13-e115.
12. Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. Jama 2003;289:2560-72.
13. James PA, Oparil S, Carter BL, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the eighth joint national committee (jnc 8). Jama 2014;311:507-20.
14. Taler SJ, Agarwal R, Bakris GL, et al. KDOQI US Commentary on the 2012 KDIGO Clinical Practice Guideline for Management of Blood Pressure in CKD. American Journal of Kidney Diseases;62:201-13.
15. Breyer JA, Bain RP, Evans JK, et al. Predictors of the progression of renal insufficiency in patients with insulin-dependent diabetes and overt diabetic nephropathy. Kidney International 1996;50:1651-8.