Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7716447

INVESTIGATION OF CKD USING PROBABILISTIC MODEL AND DEEP LEARNING APPROACH

[1]Sapna Malhotra [2] Pooja Gupta, [3] Ramakant Prasad
[1,2,3]Asst. Professor, Gargi College, DU

Abstract: In recent time health industry need the data of various patients collected for the research work. Usually, doctor suggest the medicine one the basis of their past knowledge and diagnosis report of the person. This paper has to investigate the technique which has quite good to predict the CKD disease. The deep learning technique has been applying to find the accuracy of prediction disease. The data sets used for kidney disease has been taken from the UCI library. This investigation does work on various DL learning algorithm on MATLAB platform. This paper found After running MATLAB, it was discovered that the DT method has the greatest prediction value when compared to the other KNN and MLP algorithms. When DT and KNN are used instead of MLP, the execution time is much less. The accuracy, precision, and F-call value of the deep learning algorithms has been compared with exploratory bar graph where the MLP found the best for the CKD data set.

Keywords: Chronic renal disease (CRD), Deep Learning Algorithm, Probabilistic Model

I. Introduction

Chronic Kidney Disease (CKD) is a common medical condition that affects millions of people worldwide. Early detection and management of CKD are crucial for preventing further complications and improving patient outcomes. Probabilistic models and deep learning approaches have shown promise in the diagnosis and management of CKD. A probabilistic model can be used to estimate the probability of an individual having CKD based on various risk factors such as age, gender, blood pressure, diabetes, and smoking history. This type of model can be useful for identifying high-risk individuals and recommending further diagnostic tests or interventions. Deep learning approaches, on the other hand, can be used to analyse large amounts of medical data, such as electronic health records, laboratory results, and imaging studies, to identify patterns and predict outcomes. For example, a deep learning model can be trained to predict the progression of CKD based on patient data, such as kidney function tests and medical history. One challenge in using deep learning models for CKD is the need for large amounts of labeled data, which can be difficult to obtain in a clinical setting. However, recent advances in transfer learning and unsupervised learning have shown promise in overcoming this challenge. Overall, the use of probabilistic models and deep learning approaches in CKD diagnosis and management has the potential to improve patient outcomes and reduce healthcare costs by enabling early detection and personalized treatment plans. Chronic renal disease (CRD) (Álamo et al., 2011) is a serious public health problem that affects people all over the globe. Regular screening is recommended for the general population as well as for high-risk populations since early detection of chronic kidney disease enables for the development of effective treatments. When retinal microvascular symptoms (such as retinopathy, artery constriction, venular dilatation) are clinically evident, it implies that the retina might give additional screening information to supplement currently available ways of detecting chronic kidney disease. This may encounter issues such as:

- ✓ A high level of hypertension
- ✓ Hemolytic uremic syndrome
- ✓ Bones in poor condition
- ✓ A lack of proper nourishment
- ✓ Damage to the nerves

As a result of kidney illness, a person more likely to get heart and blood vessel problems. These issues may take a long time to develop. Chronic renal disease may be slowed down or even reversed if caught and treated early (Kohan & Barton, 2014). The progression of kidney disease may lead to renal failure, which necessitates dialysis or a kidney transplant in order to preserve one's health. Chronic kidney disease gets its name from the fact that the kidneys are steadily destroyed over time. This might lead to a build-up of wastes in the body. Additionally, CKD might lead to additional health issues (Jha et al. 2013).

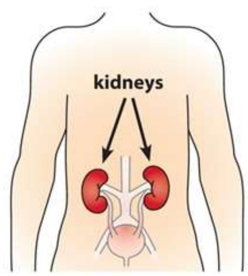

Fig. 1: A picture presented the kidney position in human body

It is the kidneys' primary role to remove excess water and wastes from the circulation to produce urine. The kidneys are responsible for keeping the salts and minerals in the blood, such as calcium, phosphorus, sodium, and potassium, in perfect balance. Chronic kidney disease (CKD) is a substantial contributor to worldwide morbidity and death, according to the Global Burden of Disease collaboration. CKD prevalence and mortality rose by 29.3 and 41.5 percent worldwide between 1990 and 2017 (Agudelo-Botero et al.,2020). More people in India died from renal failure between 2001–03 and 2010–13, a 38 percent rise. Many secondary and tertiary preventive opportunities in chronic kidney disease are overlooked in underdeveloped countries like India because of lack of resources. At this point in the illness process, patients frequently seek medical assistance because they are experiencing symptoms. CKD is only discovered in its earliest stages when the kidneys are tested for other medical reasons or, less often, as part of regular monitoring. The high frequency of CKD in India is due to a variety of factors. According to UNICEF data, 28 percent of new burns are under 2.5 kilogrammes. Pregnancy-induced hypovitaminosis A and other nutritional problems might result in a smaller kidney volume and a decreased eGFR at delivery. An increased incidence of congenital abnormalities of the kidneys and urinary system and obstructive or reflux nephropathy may be attributed to genetic inbreeding. There are several factors that contribute to kidney illness, including poverty, poor sanitation, pollution in the water supply, overpopulation, and known and undiscovered nephrotoxicants (such as heavy metals and plant poisons in folk treatments). Hypertension and diabetes, both of which are on the rise, are complicating matters further. By 2030, diabetes sufferers are predicted to outnumber those in all other countries combined in India (Ali et al. 2010). The alarming reality that only 1850 nephrologists service a population of 1.3 billion people is made worse by the fact that they are dispersed unevenly, with the majority of them concentrated in metropolitan areas. The lack of nephrology training places is exacerbated by the "brain drain" to wealthy nations. Global morbidity and death from chronic illnesses is on the rise. 4 out of every 5 chronic illness mortality in low- and middle-income nations were previously thought to be exclusive to rich countries. The number of fatalities from chronic illnesses in India is estimated to climb from 3.78 million in 1990 (40.4 percent of all deaths) to 7.63 million in 2020. Chronic kidney disease (CKD) and its risk factors have traditionally been overlooked in health-promotion efforts aimed at preventing chronic diseases. The burden of CKD is steadily rising over the globe. As of the end of 2004, there were 1,783,000 ESRD patients throughout the globe getting treatment, 77 percent of whom were on dialysis and 23 percent of whom had a functional renal transplant (RT), and this figure is growing by 7 percent per year. ESRD patients will number more than 2 million by the year 2010 if present trends continue. At 150 per million people (pmp), the incidence of ESRD in poor nations is lower than that observed in industrialised countries. Racially and ethnically diverse populations have been blamed for the gap in the prevalence of ESRD across affluent countries.

1.1 Symptoms of CKD

In most cases, there are no signs or symptoms of kidney disease in the early stages. It is possible that a persons have renal disease if they have a blood or urine test performed for another reason and the findings show that the kidneys may be in difficulty. Symptoms that may be present at a later stage include (Tsai et al., 2012):

- ✓ Tiredness
- ✓ Feet, ankles, or hands may become swollen
- ✓ Breathlessness
- ✓ feeling well.
- ✓ Pee with blood in it (urine)

1.2 Causes of CKD

An underlying medical problem places a significant burden on the kidneys, leading to chronic kidney disease. In many cases, it's a consequence of many issues.

1.3 CKD can be caused by:

High blood pressure: In the long run, this may place a burden on the kidneys' tiny blood veins, which can lead to renal failure.

Diabetes: The kidneys' small filters might be damaged if a person has too much glucose in blood.

High cholesterol: As a result, the blood veins feeding kidneys may get clogged with fatty deposits, making it more difficult for them to function effectively.

Kidney infections glomerulonephritis: A swelling of the kidneys.

Polycystic kidney disease: A genetic disorder characterised by the development of kidney cysts.

1.4 Tests for CKD

Diagnosing CKD is possible with the use of blood and urine testing (Paepe & Daminet, 2013). A dysfunctional kidney is indicated by elevated amounts of specific chemicals in the blood and urine, which are detected during these tests. A regular kidney disease screening programme may be advised if a person have a known risk factor for kidney disease, such as high blood pressure or diabetes, in order to identify the illness in its early stages.

Treatments for CKD

CKD cannot be cured; however, medication may alleviate symptoms and prevent the disease from worsening. The extent of illness will dictate the course of a person therapy.

Treatments include:

- ✓ Adapting the way of life to keep as healthy as possible
- ✓ Controlling conditions such as hypertension and excessive cholesterol with medication
- ✓ Advanced kidney disease may need dialysis, a therapy that duplicates some of the kidney's activities.
- ✓ with severe CKD, a kidney transplant may also be indicated.

1.5 Outlook for CKD

Chronic kidney disease (CKD) may be moderate, with no or few symptoms, or it can be quite severe, where the kidneys fail. Medicine and frequent check-ups are usually all that's needed to keep CKD under control for most individuals. It is estimated that only one in every 50 persons with CKD advances to renal failure. The risk of cardiovascular disease is higher for those with CKD, no matter how mild their condition is. Among them are heart attacks and strokes, which are part of a larger set of illnesses that affect the heart and blood arteries. People with renal disease are more likely to die from cardiovascular disease than the general population, although healthy lifestyle modifications and medication may help lower the risk.

II. Literature Review

Priyanka et al. (2022), One of the most common causes of renal failure is chronic "kidney disease," in which the kidneys are damaged over an extended length of time. A kidney may cease to function if the damage is severe enough. End-stage kidney disease, or "kidney shattering hopes," is a term for this (ESRD). Chronic renal disease patients may have a gradual decrease in kidney function over time, which may lead to a steady decline in kidney function. Specialists may then use this information to make a diagnosis of renal disease in their patients. By looking at the best accuracy aftereffect of AI computation continually, we are able to predict whether people with renal disease are going to progress into chronic kidney disease or not. The purpose is to research artificial intelligence-based solutions for accurately diagnosing chronic kidney disease (CKD) with the maximum degree of accuracy achievable. In order to find some of the data similarity, variable recognised proof, univariate investigation, bi-dimensional and multi-variate investigation as well as missing worth medications, the whole given dataset will be explored utilising the directed supervised machine learning strategy (SMLT).

Abdel-Fattah et al. (2022), Chronic kidney disease (CKD) is becoming more common among the population. People at risk for this illness should be identified and treated as soon as possible to avoid serious health consequences such as cardiovascular disease, greater risk of renal disease, and kidney failure in the later stages of life. Researchers in the medical field may reap significant benefits from the machine learning algorithm's capacity to consistently diagnose sickness in its early phases of development. As part of this study's ensemble learning technique, classification methods such as decision trees (DT), logistic regressions, Naive Bayes, Random Forest, and Gradient-Backed Tree Classifier were used, as well as other classification methods. Several distinct assessment procedures were used to confirm the results. These included accuracy and precision evaluations, recall evaluations, and the F1-measure. Feature selection has been accomplished via the use of three methods: full features, Relief-F, and Chi-squared. The cross-validation and testing results for each strategy have been determined using these three approaches, and the findings for each approach have been published. The SVM, DT, and GBT Classifiers were determined to be the most accurate at 100 percent when trained using the provided attributes. Overall, the qualities of the Relief-Chosen F are preferable than the features of the full F and the features of the chi-selected square.

Nishat et al. (2021), The term "chronic renal disease" refers to the gradual loss of kidney function over time. Early identification and treatment are necessary to provide a favourable prognosis and a long-life expectancy because of irreversible and undetectable deterioration up to one of the later stages of the illness. When it comes to illness diagnosis, machine learning algorithms have shown to be quite effective. For the identification of chronic renal disease, they plan to use several machine learning algorithms and compare their accuracies and performance factors. There have been eight supervised machine learning models constructed using the python programming language using the 'chronic renal disease dataset' from the University of California, Irvine's machine learning repository.

Krishnamurthy et al. (2021), The rising number of people with CKD, the high risk of developing end-stage renal disease, and the poor prognosis for morbidity and death make it a significant burden on the healthcare system. The goal of this research is to create a machine-learning model that utilises data from Taiwan's National Health Insurance Research Database to predict the beginning of chronic kidney disease (CKD) and the resulting prevalence in the general population six to twelve months in advance. Propensity score matching was used to identify a total of 18,000 patients with CKD and 72,000 patients without CKD diagnosis. A predictive model was developed based on information gleaned from each patient's two years of medical records, including demographics, medication usage, and co-morbidities. Using the model used in this work, policymakers might better estimate the prevalence of chronic kidney disease (CKD) in the community. Patient-centred care may be improved by using these models to keep tabs on those at risk, diagnose kidney disease early, and better allocate resources.

Jeong et al. (2020), The purpose of this study is to evaluate the classification performance of statistical models utilising renal data that is severely imbalanced in order to better understand how they work. The National Health Insurance Service of Korea's health examination cohort database is utilised to construct models, which are trained using a variety of machine learning techniques. The glomerular filtration rate (GFR) must be measured in order to diagnose chronic kidney disease (CKD) (CKD). When calculating it, the Modification of Diet in Renal Disease technique is employed, and the result is divided into five stages. There are six stages of chronic kidney disease (CKD) that may be identified based on estimated GFR. Categorization is accomplished by the use of two machine learning models: random forest (RF) and autoencoder (AE) (AE). For the purpose of determining the most accurate categorization of chronic kidney disease stages, the data is divided into ten-fold datasets, each of which has the same rate for each stage. When it comes to classification accuracy, the results demonstrate that RF and AE models beat multinomial and ordinal LR models. When a model is employed to mimic a highly imbalanced dataset, the precision of the model's performance may, on the other hand, mislead real-world performance. This occurs as a result of the fact that statistical models that categorise a minority group as a member of the majority group while maintaining high accuracy are available. Taking into consideration not just accuracy from the confusion matrix, but also sensitivity, specificity, precision, and the F-1 measure for each class separately, this difficulty in performance interpretation is solved. They discover that the AE model is the most accurate for accurately categorising CKD stages across all performance parameters.

Ma et al. (2020), Chronic kidney disease (CKD) is on the rise in today's scientific environment. Machine learning algorithms are becoming more significant in medical diagnosis because of their capacity to accurately classify patients with CKD. Algorithms for feature selection have recently become more important in determining the precision with which classification algorithms may be used. When it comes to diagnosing chronic renal failure on the Internet of Medical Things platform, HMANN has been recommended as a method for detecting early signs, segmentation, and diagnosis of the condition. In the pre-processing stage of the proposed approach, an ultrasound picture is taken and the kidney area of interest is segmented. The suggested HMANN approach for kidney segmentation provides excellent accuracy while minimising the amount of time it takes to outline the shape.

Khamparia et al. (2020), More than 10% of the world's population has been impacted by Chronic Kidney Disease (CKD) in recent years, and millions of people die each year as a result. Because of this, early identification of CKD may be helpful to patients in terms of extending their lives and minimising the cost of therapy. It is necessary to develop a multimedia-driven model that can assist detect the illness quickly and accurately before it gets out of hand. Researchers have utilised a variety of methods similar to traditional machine learning models in the past without including multimodal data-driven learning. For chronic kidney disease categorization, this study presents a unique deep learning framework that uses an autoencoder model that incorporates multimedia data, and a softmax classifier. Softmax classifiers are used to forecast the final class after an autoencoder has extracted the dataset's valuable characteristics. It has been tested on a dataset of 400 CKD patients with 25 variables from the UCI dataset, which is a binary classification issue involving 400 patients.

Macias et al. (2020), To enhance mortality prediction in end-stage renal disease (ESRD), they propose to integrate huge information gathered throughout the course of a patient's life with machine learning approaches. This study used a retrospective cohort of 261 patients, whose diagnoses, lab tests, and characteristics collected during haemodialysis sessions were all integrated for this study's analysis. Long short-term memory (LSTM) recurrent neural networks have a baseline performance that was determined using random forest (RF). Then, LSTMs were trained using a variety of variables, including those discovered by RF, as well as those selected by experts. It was shown that the factors discovered using RF had stronger predictive power than those selected using expert knowledge, but utilising all of them yielded the best results overall. Integrating three different data sources into the receiver operating characteristic curve is expected to result in an improvement of more than 4%. Using this method, one may accurately forecast death across a wide variety of time spans. Patients with End-Stage Renal Disease (ESRD) and the use of LSMTs are expected to enhance mortality prediction models by a large margin. In the end, the use of machine learning may lead to a paradigm shift in the examination of predictive indicators for mortality in ESRD patients.

Yashfi et al. (2020), CKD is a significant cause of death and disability. In 1990, it was ranked 27th, while in 2010 it was ranked 18th. 2013 saw the death of about a million individuals. People in underdeveloped nations are still being afflicted by CKD, despite this fact. They developed our system in Python, an interpreted high-level programming language. Random Forest and ANN were used to train the data using a 10-fold CV. Random Forest's 97.12 percent accuracy and ANN's 94.5 percent accuracy are comparable. Chronic renal disease may be detected earlier with the use of this approach.

Qin et al. (2019), Chronic kidney disease (CKD) is a worldwide health concern that has a high morbidity and death rate, and it may lead to the development of other illnesses. Patients may be unaware they have CKD since there are no evident symptoms in the early stages of the illness. Patients who are diagnosed with CKD at an early stage might benefit from medication that slows the disease's development. Due of their high speed and accuracy of identification, machine learning models may help doctors attain this aim more quickly. They suggest a machine learning approach to CKD diagnosis in this paper.

III. Chronic Kidney Disease in India

The prevalence of risk factors for CKD, such as diabetes, hypertension, renal stone disease, etc., was also examined in this research. As a result of their extrapolation, the authors came to the conclusion that the prevalence and incidence of ESRD in India would be respectively 785 and 160 pmp (Agarwal & Srivastava, 2009). Diabetic nephropathy accounted for 41% of the cases of CKD in this population-based investigation (Ayodele & Alebiosu, 2010). When a fresh survey was conducted in a nearby location, the prevalence of impaired kidney function was found to be 8.6 per thousand after screening 25,000 individuals, and then 13.9 per thousand after screening another 21,500 people in the same area. As a result of the first survey, 0.68 percent of the population was found to have some kind of renal disease (non-CKD). Additionally, this research looked at diabetes and hypertension, two of the most common risk factors for chronic kidney disease (CKD). There was one significant difference between this research and the one from North India, namely that serum creatinine was not measured in all participants. In addition, no mention was made of whether the initial test for creatinine was followed up with another to confirm the presence of CKD. Due to its reliance on ESRD patients examined at a single hospital and the assumption that all of the region's ESRD patients were presenting to that facility, this research was constrained by the potential of referral bias and population mobility. Despite the fact that the patients were assessed at a hospital, no risk factors for CKD were mentioned in this research. As a result, the study's major goal was to focus on individuals with end-stage renal disease (ESRD). A definite diagnosis of chronic glomerulonephritis or chronic interstitial illness is difficult to make if a patient has reached a point where a conclusive diagnosis cannot be established. Diabetes was shown to be the underlying cause of CKD in about 30% of patients in these hospital-based investigations. Approximately 0.8% of patients had kidney disease stage 3 or above, according to data gathered from 48 hospitals throughout India for another hospital-based research.

There is a pilot initiative initiated by a group of nephrologists and subsequently endorsed by the National Society of Nephrology that might shed light on the pattern of CKD patients who come to hospitals. When it was first launched in June 2005, the "Indian CKD Registry" was tasked with gathering information on the prevalence of kidney disease in India, as well as the types of disorders that may lead to the condition. In the beginning, just 10 centres participated, but now 152 hospitals have decided to participate, and paediatric nephrologists have also agreed to join. Until yet, around 30,000 participants' data has been entered into a database. Nearly 70% of the adult population is male, with an average age of 45–50 years. There were a large number of patients in the stage CKD 4–5 categories (70 percent). In about 30 percent of patients, diabetes mellitus was shown to be the primary cause of CKD; in 97 percent of cases, type 2 diabetes was responsible, and in 40 percent of cases, the duration of diabetes was less than 10 years. While just 0.7% of those with CKD in stage 1 were found to have cardiovascular disease (CVD), this increased to 43% in stage 5 of the disease. When a patient is admitted to the hospital for the first time, the register as it now exists has several limits in terms of the data it can gather. There has been no further investigation as of yet. Some doctors already treat patients before the data is entered into the register, affecting the accuracy of the results. The variables may be affected by this.

Apart from finding out how common kidney disease is in India, this study also aims to find out the prevalence of its causes and complications, characterise risks for kidney disease, develop models for predicting Indian-specific risks for kidney disease and its complications, develop educational materials in local languages to educate Indians about kidney disease and change the unjust system. There are some centres that profess to use a domiciliary method; however, this is more often a camp style. The Indian Society of Nephrology's Annual Conference has heard about the SEEK study for the last two years. A total of 6,000 adults from 21 sites at 53 community camps had been examined at the previous presentation. Ninety-three percent of individuals had their serum creatinine and urine tested. According to the established eGFR method, this research found a 17.4% prevalence of CKD, which is 17 times higher than prior community-based findings from India. Criteria to identify CKD, such as repeated creatinine tests to ensure that renal disease is chronic, were not included in the presentation. Diabetes and hypertension were also examined as potential CKD risk factors in this investigation. Until recently, no SEEK data has been made public. The truncated MDRD formula, which has not been verified in the Indian population, was used to compute eGFR in the vast majority of the papers cited above. However, it is the only approach available for evaluating GFR in Indian community-based research at this time. A growing number of people are developing chronic kidney disease, and this number is likely to climb even more in the future, given that the average life expectancy in the developed world is already 63 years, and is expected to rise even further in the future.

IV. Challenges in CKD Management in India

The availability and cost of therapy are two of the most essential considerations in the treatment of any illness. We should first examine the availability of RRT facilities in India before talking about how much they cost. There are now over 820 nephrologists in India, with a distribution of 35.5 percent each in the north, south, west, east, and central India, as well as 30, 23, 9, and 2.5 percent across the rest of the country. Hemodialysis units' number in the seven-hundred-and-ten range. Figure 1 depicts the distribution of units by state. There are around 2,500 dialysis stations in these hemodialysis units, with an average of three stations per unit (ranging from two to twenty-four). Dialysis facilities that conduct both transplant-oriented dialysis as well as maintenance hemodialysis are about 85 percent privately owned (MHD). In contrast to the private sector, which can afford to offer MHD, the public sector is limited to operating RT-oriented hemodialysis facilities because of financial constraints. A total of 5,000 patients in India are now receiving CAPD treatment, which began in 1990. The usage of CAPD is on the rise, but there is still a long way to go before it is widely accepted. For many individuals, CAPD is a last-resort treatment option rather than a treatment of first choice. This shows that a significant percentage of patients discontinue CAPD therapy for a variety of reasons, including death while undergoing treatment, method failure, discontinuation of therapy for financial reasons, and the receipt of RT. An educated hypothesis is that the most prevalent cause of death is a patient's death while he or she was undergoing treatment. Radiotherapy (RT) is the best therapeutic option for people with end-stage renal disease (ESRD). As well as saving money in the long term, it delivers a higher quality of life and a greater chance of survival. Nearly 3,500 RTs are performed each year at India's 172+ RT facilities, the majority of which are privately owned. Without a well-established donor programme, live donors are the primary supply of organs for transplantation in India, and sadly a considerable majority are not related. Private hospitals see a higher rate of unrelated RT than do public hospitals. Many of these unrelated transplants are illegal, therefore the outcomes and details are not published in peer-reviewed publications. Even though an organ transplant statute was approved in 1994, just around 700 cadavers have been transplanted in India to this day. Unrelated RT may be difficult to completely halt because of a lack of structured cadaver RTs and the absence of family donors in many cases (Agarwal & Srivastava, 2009).

V. Methodology

To carry out this experiment, the procedure described above will be used. A prediction DNN model and performance assessment are all part of this process. Figure 2 depicts the many components involved in data gathering, pre-processing, and selecting target variables.

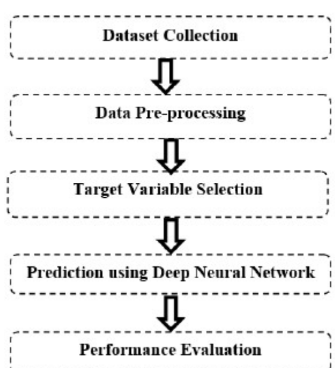

Fig. 2: Flow of Research Methodology

5.1 Mathematical Model for CKD

Let X = (X1, X2, ..., Xn) be the vector of risk factors, where X1 represents age, X2 represents gender, X3 represents blood pressure, X4 represents diabetes, and X5 represents smoking history.

Let Y be the binary variable representing the presence or absence of CKD. Assuming X and Y are independent, the probability of Y=1 given X can be modeled using logistic regression as:

$$P\ (Y=1 \mid X) = 1 / (1 + \exp(-\beta 0 - \beta 1X1 - \beta 2X2 - \beta 3X3 - \beta 4X4 - \beta 5X5))$$

Where, β0, β1, β2, β3, β4, and β5 are the coefficients to be estimated. Alternatively, the probability of Y=1 given X can be modeled using a Bayesian network, which represents the joint probability distribution of X and Y as a product of conditional probabilities:

$$P\ (X, Y) = P\ (Y \mid X)\ P(X) = P\ (Y \mid X1, X2, X3, X4, X5)\ P(X1)\ P(X2)\ P(X3)\ P(X4)\ P(X5)$$

Where, P (Y | X) is the conditional probability of Y given X, and P(Xi) is the marginal probability distribution of Xi.

5.2 Deep learning approach

Let X = (X1, X2, ..., Xm) be the vector of patient data, where X1 represents kidney function tests, X2 represents medical history, and so on.

Let Y be the variable representing the progression of CKD. A neural network can be used to predict Y based on X, where the output of the neural network is a continuous value that can be thresholded to obtain a binary prediction.

The neural network can be represented using the following equations:

$$h1 = f(W1\ X + b1)$$
$$h2 = f(W2\ h1 + b2)$$
$$\ldots$$
$$hk = f(Wk\ hk\text{-}1 + bk)$$
$$y = g(V\ hk + c)$$

where f is the activation function, W1, W2, ..., Wk and b1, b2, ..., bk are the weights and biases of the hidden layers, V and c are the weights and bias of the output layer, and h1, h2, ..., hk are the hidden layer activations.

The weights and biases of the neural network can be learned by minimizing a loss function L using techniques such as backpropagation and stochastic gradient descent:

$$L = 1/N \sum i=1N (y\_pred,i - y\_true,i)^2$$

Where, N is the number of patients in the training set, y_pred,i is the predicted value of Y for patient i, and y_true,i is the true value of Y for patient i.

The performance of the neural network can be evaluated using metrics such as accuracy, precision, recall, and F1 score.

5.3 Data Collection

As a data source, we've chosen the patient's medical record for renal illness. The information in this collection was gathered from the UCI library. A total of 334 patients' records are included, each with a unique set of 11 traits and parameters along with age and gender. Table 1 shows a binary categorization of CKD and non-CKD illness as a goal variable is included.

Table 1: Codes and full form**s**

Sr. No.	**Code**	**Full Form**
1	id	Patient ID
2	age	Age
3	bp	Blood Pressure
4	sg	Specific Gravity
5	al	Albumin
6	su	Sugar
7	rbc	Red Blood Cell
8	pc	pus cell
9	pcc	pus cell clumps
10	ba	bacteria
11	bgr	blood glucose random
12	bu	Blood Urea
13	sc	Serum Creatinine
14	sod	Sodium
15	pot	Pottasium
16	hemo	Hemoglobin

17	pcv	packed cell volume
18	wc	white blood cell count
19	rc	red blood cell count
20	htn	hypertension
21	dm	diabetes mellitus
22	cad	coronary artery disease
23	appet	appetite
24	pe	pedal edema
25	ane	anemia
26	classification	

5.4 Data Pre-processing

When it comes to data mining, pre-processing is the most critical step in the process, since raw data is riddled with errors, missing information, and inconsistencies that may lead to disastrous predictions using DL learning. It's the dataset number which need to be transformed to numeric values for Males and Females to make it easier for the computer to handle as string values are not understood by deep learning mechanism.

5.5 Prediction using DNN

An artificial neural network (ANN) is a subset of a biological neural network (BNN) (Ambroise et al. 2017) that mimics the structure and functions of a biological neural network (BNN) consisting of an input, weights, and activation function. When deciding on the number of neurons, the input data might be taken into consideration. Three layers of training data were applied to the model. The activation function for the output layer is Sigmoid. The model was optimised using the Stochastic Gradient algorithm.

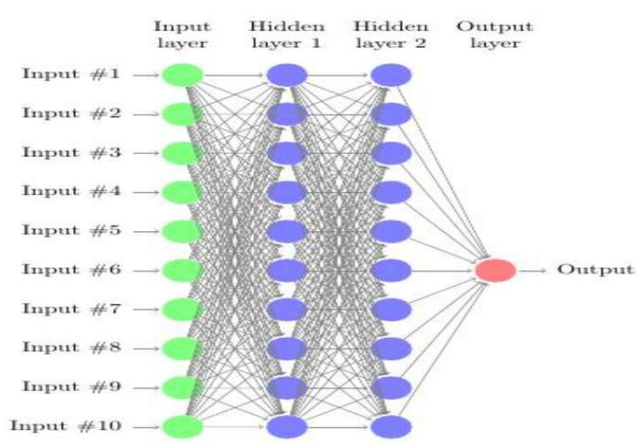

Figure 3: MLP

5.6 Evaluation of The Model

Accuracy, specificity, sensitivity, kappa statistic, precision, F1 score, ROC score, and recall are some of the performance metrics used in this study.

- **Accuracy-** It is used to determine how many of the projected data points were right.
- **Precision:** It is defined as the percentage of retrieved instances that are relevant.
- **Recall/ Sensitivity:** In the context of predictive analytics, recall is a statistic that measures how many right positive predictions were produced out of all possible positive predictions.
- **F1 Score:** Between accuracy and recall, this is the perfect balance. A score of f1 ranges from zero to one. A model's precision (the number of records it can properly classify) and robustness (the number of records it doesn't miss) are measured using this metric. The following is an example of how to represent an F1 score:

VI. Result

We test on three machine learning algorithms as KNN, DT and MLP. The accuracy, precision and F-call has been evaluated. This paper evaluates the deep learning algorithm test on the CKD data sets. The data taken from UCI library and put up the data in MATLAB to perform the Deep learning algorithms.

Table 2: CKD prediction outcome (Execution on MATLAB)

CKD Prediction				
	Accuracy	Precision	F-call	Time
KNN	74.2424	78.8889	62.0574	1.197
DT	95.1515	80	77.8788	0.7742
MLP	94.536	80	77.646	10.3515

The above result has been derived from MATLAB based on the proposed methogology. As this table shows that the DT has higest accuracy prediction for CKD data sets.

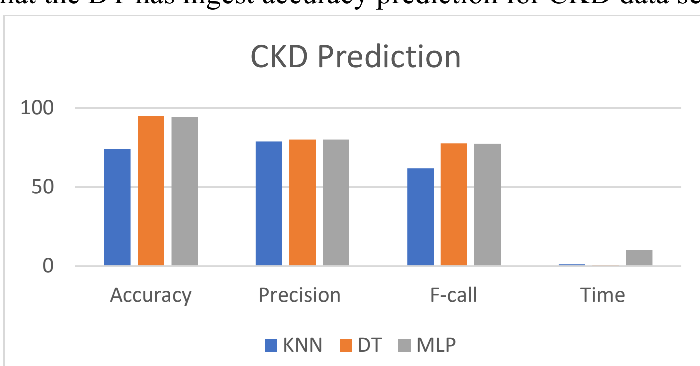

Fig. 3: CKD prediction outcome with accuracy parameter and time of execution

The above result found after the execution of MATLAB, the DT algorithm has highest prediction value as compared it to the other KNN and MLP. The execution of time is very less in DT and KNN as compared to MLP. The probability of findings through the deep learning algorithm is presented in accuracy, precision and F-call value.

VII. Conclusion and Future work

On the basis of the information or diagnostic report provided by patients, this article needs to conduct research into a method that is pretty effective at predicting CKD diseases. The data sets that were used for the analysis of CKD disease originated from the UCI library. The MATLAB platform is used in this inquiry to do work on a variety of machine learning algorithms. According to the results of this study When compared to the other KNN and MLP algorithms, it was revealed that the DT technique had the largest prediction value after running MATLAB. This was the case regardless of the input data. The amount of time needed to carry out an operation is significantly reduced when DT and KNN are used rather than MLP. When compared with an exploratory bar graph, the accuracy, precision, and F-call value of the deep learning methods were found to be worse, with the MLP coming out on top as the superior approach for the CKD data set.

References

Abdel-Fattah, M. A., Othman, N. A., & Goher, N. (2022). Predicting Chronic Kidney Disease Using Hybrid Machine Learning Based on Apache Spark. *Computational Intelligence and Neuroscience*, *2022*.

Agarwal, S. K., & Srivastava, R. K. (2009). Chronic kidney disease in India: challenges and solutions. *Nephron clinical practice*, *111*(3), c197-c203.

Agudelo-Botero, M., Valdez-Ortiz, R., Giraldo-Rodríguez, L., González-Robledo, M. C., Mino-León, D., Rosales-Herrera, M. F., ... & Dávila-Cervantes, C. A. (2020). Overview of the burden of chronic kidney disease in Mexico: secondary data analysis based on the Global Burden of Disease Study 2017. *BMJ open*, *10*(3), e035285.

Álamo, S. M., Esteve, C. G., & Pérez, M. S. (2011). Dental considerations for the patient with renal disease. *J Clin Exp Dent*, *3*(2), 112-9.

Ali, M. K., Narayan, K. V., & Tandon, N. (2010). Diabetes & coronary heart disease: current perspectives. *The Indian journal of medical research*, *132*(5), 584.

Ambroise, M., Buccelli, S., Grassia, F., Pirog, A., Bornat, Y., Chiappalone, M., & Levi, T. (2017). Biomimetic neural network for modifying biological dynamics during hybrid experiments. *Artificial Life and Robotics*, *22*(3), 398-403.

Ayodele, O. E., & Alebiosu, C. O. (2010). Burden of chronic kidney disease: an international perspective. *Advances in chronic kidney disease*, *17*(3), 215-224.

Jeong, B., Cho, H., Kim, J., Kwon, S. K., Hong, S., Lee, C., ... & Heo, T. Y. (2020). Comparison between statistical models and machine learning methods on classification for highly imbalanced multiclass kidney data. *Diagnostics*, *10*(6), 415.

Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., ... & Yang, C. W. (2013). Chronic kidney disease: global dimension and perspectives. *The Lancet*, *382*(9888), 260-272.

Khamparia, A., Saini, G., Pandey, B., Tiwari, S., Gupta, D., & Khanna, A. (2020). KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network. *Multimedia Tools and Applications*, *79*(47), 35425-35440.

Kohan, D. E., & Barton, M. (2014). Endothelin and endothelin antagonists in chronic kidney disease. *Kidney international*, *86*(5), 896-904.

Krishnamurthy, S., Ks, K., Dovgan, E., Luštrek, M., Gradišek Piletič, B., Srinivasan, K., ... & Syed-Abdul, S. (2021, May). Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan. In *Healthcare* (Vol. 9, No. 5, p. 546). Multidisciplinary Digital Publishing Institute.

Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*, *111*, 17-26.

Macias, E., Morell, A., Serrano, J., Vicario, J. L., & Ibeas, J. (2020). Mortality prediction enhancement in end-stage renal disease: A machine learning approach. *Informatics in Medicine Unlocked*, *19*, 100351.

Nishat, M. M., Dip, R. R., Faisal, F., Nasrullah, S. M., Ahsan, R., Shikder, M. F., ... & Hoque, M. A. (2021). A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, *18*(e6).

Paepe, D., & Daminet, S. (2013). Feline CKD: Diagnosis, staging and screening–what is recommended? *Journal of feline medicine and surgery*, *15*(1_suppl), 15-27.

Priyanka, D., Anusha, D., Anandhi, T., Indria, P., Brumancia, E., & Gomathi, R. M. (2022). Prediction of Chronic Kidney Disease by Best Accuracy Using Supervised Classification Machine Learning Approach. In *Advances in Data Science and Management* (pp. 265-272). Springer, Singapore.

Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2019). A machine learning methodology for diagnosing chronic kidney disease. *IEEE Access*, *8*, 20991-21002.

Tsai, Y. C., Chiu, Y. W., Hung, C. C., Hwang, S. J., Tsai, J. C., Wang, S. L., ... & Chen, H. C. (2012). Association of symptoms of depression with progression of CKD. *American journal of kidney diseases*, *60*(1), 54-61.

Yashfi, S. Y., Islam, M. A., Sakib, N., Islam, T., Shahbaaz, M., & Pantho, S. S. (2020, July). Risk prediction of chronic kidney disease using machine learning algorithms. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.