International Journal of Intelligent Information Technologies
Volume 20 • Issue 1

Intelligent Decision Support for Identifying Chronic Kidney Disease Stages:
Machine Learning Algorithms

V. Shanmugarajeshwari, Kalasalingam Academy of Research and Education, India*
https://orcid.org/0000-0001-6202-8517

M. Ilayaraja, Kalasalingam Academy of Research and Education, India

ABSTRACT

The decision tree classification algorithm is becoming increasingly important in machine learning (ML) technology. It is being used in a variety of fields to solve extremely complicated issues. DTCA is also utilised in medical health data to identify chronic kidney disorders such as cancer and diabetes utilising computer-aided diagnosis. Deep learning is an intelligent area of machine learning in which neural networks are used to learn unsupervised from unstructured or unlabeled data. For CKD, the DL employed the deep stacked auto-encoder and soft-max classifier techniques. Kidney illness is another condition that can lead to a variety of health problems. Random forest, SVM, C5.0, decision tree classification algorithm, C4.5, ANN, neuro-fuzzy systems, classification and clustering, DSAE, DNN, FNC, MLP are used in this study to predict and identify an early diagnosis of CKD patients using various machine and deep learning algorithms using R Studio and Python Colab software. The many stages of chronic kidney disease are identified in this paper.

KEYWORDS

Chronic Kidney Disease, CKD Stages, Computational Decision Support System, Decision Tree Classification Algorithm, Machine and Deep Learning Algorithms

INTRODUCTION

Chronic kidney disease, or simply CKD, is one of the fastest-growing noncommunicable illnesses, contributing to a significant increase in death and sickness. By 2020, CKD will have affected 803 million people worldwide, with 663 million men and 526 million women (Lakshmanaprabu et al., 2019). It is also a huge public health issue in India, which is home to 17% of the world's population (Ahmed et al., 2014).

About 10% of the adult residents worldwide has chronic kidney disease (CKD), making it one of the top 20 causes of mortality globally. Normal kidney function is disrupted by CKD. The rising

DOI: 10.4018/IJIIT.334557 *Corresponding Author

prevalence of chronic kidney disease necessitates the development of reliable methods for making accurate prognoses at an early stage. The innovation of this work is the creation of a diagnostic system for chronic renal disease (Sivasankar et al., 2019). Cloud computing and Internet of Things environments have lately seen widespread usage in a number of healthcare applications due to their ability to integrate monitoring devices like sensors as well as medical equipment for maintaining tabs on remote patients. Instead of relying on limited storage and processing resources, the huge volume of information formed by healthcare IoT gadgets might be analyzed in a CC context, leading to improved healthcare delivery. Simultaneously, early detection of CKD becomes crucial to drastically lowering mortality rates (Gupta et al., 2022). The problem of significant prognostic variation in chronic disorders affects clinical support systems. Large numbers of people dying from conditions like chronic kidney disease (CKD) are attributed in large part to this lack of knowledge. For this reason, accurate identification of this condition is a major focus area for healthcare providers (Al-Chalabi et al., 2023).

Analytical environments are employing a variety of ways to improve the value of health-related problem prediction by developing and exploring healthcare data records. Data from health-care records is mostly visual, and it comes from a wide range of sources around the world, including sensor equipment, photographs, and text in the form of electronic records. This disparity in data collection and representation methods points to several trials in the handling procedure and analysis of the original data. To analyze various types of documents, a diverse range of procedures is required (Reddy & Aggarwal, 2015). The kidneys' job is to filter blood and pass it through a filter. It removes unnecessary blood and maintains electrolyte and hydration balance. It strains blood and produces urine, which is produced by the kidney's two bean-shaped structures. Every kidney contains a million nephrons (units of measurement).

The kidneys' operations are to pass through a filter of the blood. It eliminates unwanted blood to regulate the stability of electrolytes and fluid. It strains blood, they create urine, which two bean-shaped structure of the kidney. Every one kidney surrounds a million things of unit so-called nephrons (Aditya et al., 2020).

FACTORS OF CKD

The following are some of the factors which lead to CKD, the main cause is diabetes and others are hypertension, smoke, fatness, heart illness, family record, alcohol, and age problem.

Symptoms

Some of the warning sign is listed down, that could be variations to urinary function, plasma in the urine, bulge & pain, severe tiredness and weakness.

Types: Acute and Chronic

- Acute Prerenal Kidney Failure - Suddenly decreases blood flow
- Acute Intrinsic Kidney Failure - Straight injury to the kidney's foundations unexpected damage in kidney
- Chronic Prerenal Kidney Failure - Gradually decreases blood flow
- Chronic Intrinsic Kidney Failure - Direct damage to the kidneys causes a gradual loss in kidney function (Lakshmanaprabu et al., 2019).

CKD is a worldwide health crisis. In 2019, the World Health Organization agree to fifty-eight million deaths and 35 million recognized to chronic kidney disease. The world level 850 million people now predicted to have kidney diseases from many causes, CKD causes at least 2.4 million deaths world wide-reaching per year sixth fastest-growing cause of disease and death. Dialysis is a fashion of life for many patient's pain with kidney sicknesses in India. The medical record of Government of Tamil

Nadu, India, every one year 2.2 Lakh fresh patients affected by final point renal disease or end-stage renal disease. According to GBD learning, kidney disease was hierarchical 27th 1990 but rose to 18th in 2010 and 9th in 2019. Motivations on the development and use of machine learning algorithms for classical methods using other machine learning approaches to achieve high accuracy.

CKD Factors Increase the Risk

Figure 1 depicts some of the factors that contribute to CKD: diabetes is the most common cause, but others include hypertension, obesity, heart disease, family history, alcohol, and age.

CKD Early Signs Symptoms

Figure 2 depicts some of the warning signs, which include changes in urinary function, plasma in the urine, bulge and pain, extreme weariness, and weakness.

Acute and Chronic Are the Two Types of CKD

- Acute_Prerenal_Kidney_Failure APKF
- Acute_Intrinsic_Kidney_Failure AIKF

Figure 1. Factors of CKD

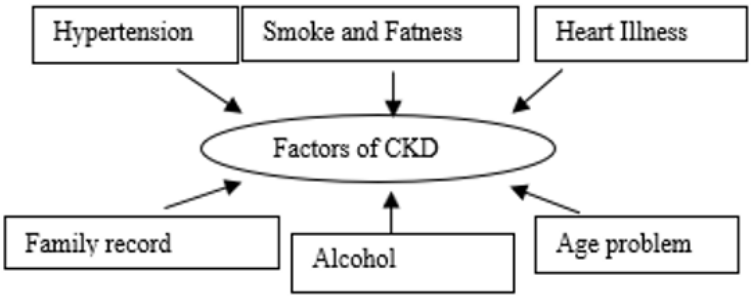
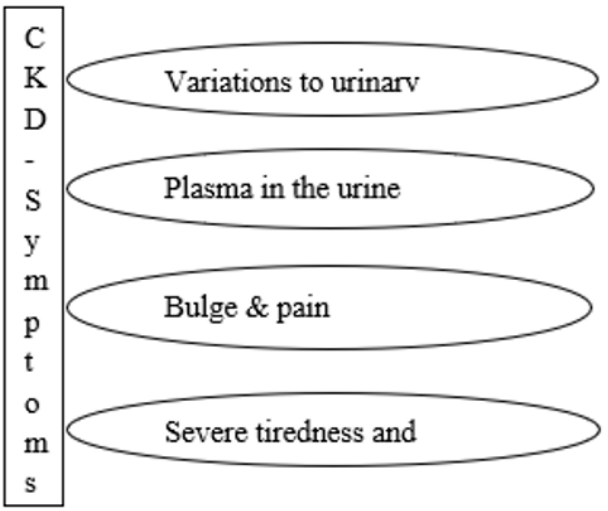


Figure 2. Symptoms of CKD



- Chronic_Prerenal_Kidney_Failure CPKF
- Chronic_Intrinsic_Kidney_Failure CIKF (Reddy & Aggarwal, 2015)

CKD is a global health emergency. In 2019, the World Health Organization estimates that there will be 58 million deaths worldwide, with 35 million of them due to chronic renal disease. CKD kills at least 2.4 million people every year, making it the sixth fastest-growing cause of disease and death on the planet. Dialysis is a way of life for many patients in India who suffer from renal diseases. According to the medical records of the Government of Tamil Nadu, India, 2.2 lakh new patients with end-stage renal disease are diagnosed every year. Kidney disease was ranked 27th in 1990 by GBD Learning, but rose to 18th in 2010 and 9th in 2019. Involves the development and implementation of machine learning algorithms in order to attain high accuracy for traditional methods employing other machine learning methodologies.

Figure 4 depicts the numerous factors that influence the risk level for chronic renal disease when patient data is analyzed using healthcare data analytics. Different machine learning and deep learning techniques can be used to classify a model for kidney disease risk prediction, and the results can then be evaluated in terms of model accuracy, specificity, and sensitivity.

LITERATURE REVIEW

This review of collected works refers to a serious overview. The analyses were carried out on a variety of outline topics. The foundation of fundamental knowledge, which enables everyone to develop advanced knowledge and thought for advanced study endurance.

The remainder of the paper is organized as follows, where Section 2 provides an overview of autoencoders, stacked autoencoders, and the SoftMax classifier. The definition of the dataset and the

Figure 3. Types of CKD

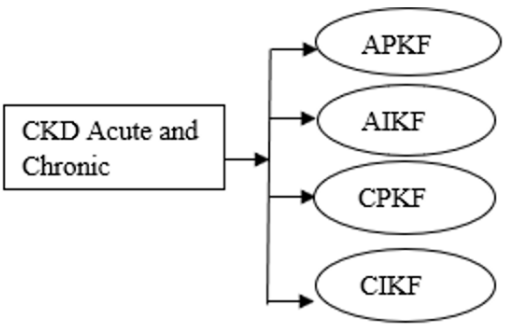
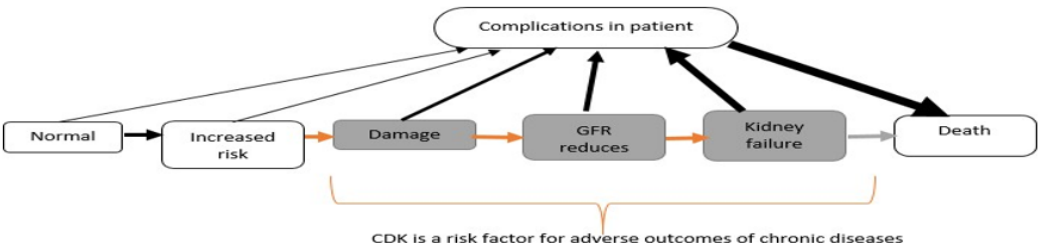


Figure 4. Factors that influence the risk level for chronic kidney disease according to healthcare data analytics



deep neural architecture procedure (a combination of stacked autoencoder and classifier SoftMax) are discussed in Section 3. The findings as well as the network output from the suggested network are explained in Section 4. Section 5 contains the conclusion and future scope.

The decision tree model and DSAE have the highest accuracy (99.04%) compared to the other models. This research provides preliminary evidence that ML methods, and specifically the DT and DSAE, support vector machine (SVM), may enhance CKD prediction. The findings of this research have the potential to enhance CKD prediction models, which in turn might improve patient outcomes and lessen the strain on healthcare systems via earlier detection and treatment of the illness. An effective home-based approach for detecting CKD was suggested in (Bhaskar et al., 2023). The suggested technology is meant to diagnose kidney illness from saliva samples, and it comes at a time when non-invasive methods of disease diagnosis are gaining favor. The non-invasive method of collecting saliva samples has contributed to the rise in popularity of salivary diagnostics in recent years. Salivary biomarkers for kidney disease monitoring and detection are the subject of an experimental study. To diagnose CKD, they determined the urea concentration in the patient's saliva. This article also details how predictive analysis based on machine learning methods and data analytics may be used in the context of distant healthcare administration. Samples were accurately categorized at a rate of 97.1% using the suggested health monitoring system. This approach may provide improved healthcare services in areas where internet access is widespread by providing real-time decision assistance on a remote monitoring platform.

FINDINGS

From the literature review as a result of this research, it is clear that the clinical performance of healthcare decision support can be evaluated by covering, and machine learning approaches can be rated by various algorithms and benchmark datasets. Our study was divided into three stages based on this survey. It works with large datasets and makes use of the R and Python programming languages. This research paper proposes an approach for classifying structures using several machine learning algorithms that has a high level of accuracy. The planned study effort will be effectively implemented in R and Python with a GUI environment in the future.

DESCRIPTION OF THE DATA SET: DATA COLLECTING INFORMATION

Chronic kidney disease is associated with the long-term disease affecting dialysis patients. It can also predict CKD based on the pertinent factors. Thirty-three attributes are included in the dataset that can be used to predict CKD. To create a system that works with both numerical and nominal data types. The data size is approximately 1,03,200 records collected from Velammal Hospital, Madurai. Pre-processing, attribute feature selection approaches, cataloguing, and classification algorithms are among the 33 aspects that can be applied to chronic kidney data utilizing quality assessment using machine and deep learning algorithms.

Main CKD Monitoring Characteristics

Urine albumin and eGFR are the two most essential characteristics. A blood test called the GFR (glomerular filtration rate) determines how successfully affected patients' kidneys filters patient blood. In addition to monitoring CKD, a urine test for albumin concentration can be helpful. In order to diagnose CKD, an eGFR and a test of albuminuria or proteinuria are required. Creatinine, eGFR, urine albumin-to-creatinine ratio, and urine protein-to-creatinine ratio are the key features of laboratory testing for CKD with appropriate reporting units. High blood pressure and heart rate are two other things to watch out for.

Early detection of this condition and the identification of risk factors will aid in the prevention of further disease progression and the reduction of patient health issues. The glomerular filtration

Table 1. Dataset description format: Attributes of chronic kidney disease dataset

S. No	Attribute_Name	Attribute_Type	Attribute_Code	Possible_Values
1.	Age	Numeric	age	E, VG, G, F, P
2.	Age Group	Numeric	ageg	E, VG, G, F, P
3.	Sex	Nominal	Sex	E, VG, G, F, P
4.	Systolic Blood Pressure	Numeric	sysbp	E, VG, G, F, P
5.	Diastolic Blood Pressure	Numeric	diabp	E, VG, G, F, P
6.	Specific Gravity	Numeric	sap	E, VG, G, F, P
7.	Albumin	Numeric	alb	E, VG, G, F, P
8.	Sugar	Numeric	sug	E, VG, G, F, P
9.	Red Blood Cell	Nominal	rbc	E, VG, G, F, P
10.	Pus Cell	Nominal	pcell	E, VG, G, F, P
11.	Pus Cell Clumps	Nominal	pcellc	E, VG, G, F, P
12.	Bacteria	Numeric	bac	E, VG, G, F, P
13.	Blood Glucose Random	Numeric	bgr	E, VG, G, F, P
14.	Blood Urea	Numeric	blu	E, VG, G, F, P
15.	Serum Creatine	Numeric	sercr	E, VG, G, F, P
16.	Sodium	Numeric	sdi	E, VG, G, F, P
17.	Potassium	Numeric	pota	E, VG, G, F, P
18.	Hemoglobin	Numeric	hg	E, VG, G, F, P
19.	Packed_Cell_Volume	Numeric	p_c_v	E, VG, G, F, P
20.	White_Blood_Cell_Count	Numeric	w_b_c_c	E, VG, G, F, P
21.	Red_Blood_Cell_Count	Numeric	r_b_c_c	E, VG, G, F, P
22.	Hypertension	Nominal	hyptn	E, VG, G, F, P
23.	Diabetes Mellitus	Numeric	diam	E, VG, G, F, P
24.	Appetite	Nominal	app	E, VG, G, F, P
25.	Pedal Edema	Nominal	peed	E, VG, G, F, P
26.	Low Density Lipoprotein	Numeric	ldl	E, VG, G, F, P
27.	smoking status	Numeric	smo	E, VG, G, F, P
28.	Alcohol Drinking	Numeric	alc	E, VG, G, F, P
29.	Anemia	Nominal	ane	E, VG, G, F, P
30.	Coronary Artery Disease	Nominal	Coad	E, VG, G, F, P
31.	Estimated Glomerular Filtration Rate	Numeric	egfr	E, VG, G, F, P
32.	CKD Level	Numeric or Nominal	ckd	E, VG, G, F, P
33.	Class	Numeric or Nominal	Class	E, VG, G, F, P

rate is an excellent indicator of kidney function (GFR). The amount of renal function a patient has is determined by the GFR (Glomerular Filtration Rate) measurement. As kidney disease progresses, it's worth begins to dwindle. It's calculated based on a patient's blood count, age, race, gender, and

other criteria. CKD is divided into five stages or levels based on the GFR value. Table 2 shows the various stages of CKD as well as the GFR level.

STACKED AUTO-ENCODER

Before to get into what a stacked autoencoder (SAE) is, need to understand what an autoencoder is (AE). An autoencoder is a deep learning architecture that is comparable to the Artificial Neural Network (ANN) and is based on unsupervised learning (Aditya et al., 2020). It is used for input encoding and decoding.

Figure 5 displays a rudimentary autoencoder (AE) network with an input, secret, and output layer that resembles an Artificial Neural Network (Liu et al., 2017). In an autoencoder, each layer has a specific number of neurons. In general, the input and output layers have the same number of neurons, while the hidden layer has less neurons, making it easier than feed-forward neural networks. Some forms of AE include sparse autoencoders, zero-biased autoencoders, denoising autoencoders, contractive autoencoders, and convolutional autoencoders (Kannadasan et al., 2019).

Figure 5 depicts a general AE procedure. Let's start with the assumption that the autoencoder's input is x. To obtain the secret code, this input vector is encoded using the encoder portion, which is delivered as an input to the decoder section. The decoder component tries to reconstruct the real input from the message, x '. The primary purpose of employing an autoencoder is to extract critical characteristics while lowering data size and obtaining noise-free data (Pujari & Hajare, 2014).

Equation (1) is used in the encoding process to obtain code c, while Equation (2) is used in the decoding phase to recreate the input data, i.e. x '. Then, as shown in eq. (3), the back-propagation technique can be used to measure the error and fine-tune the network to get the rebuilt output closer to the input data. The basic idea behind the strategy described above is to find key patterns in the data provided (Adam et al., 2019)

$$C = F\left(w^t x + b\right) \tag{1}$$

$$x' = F\left(wc + b'\right) \tag{2}$$

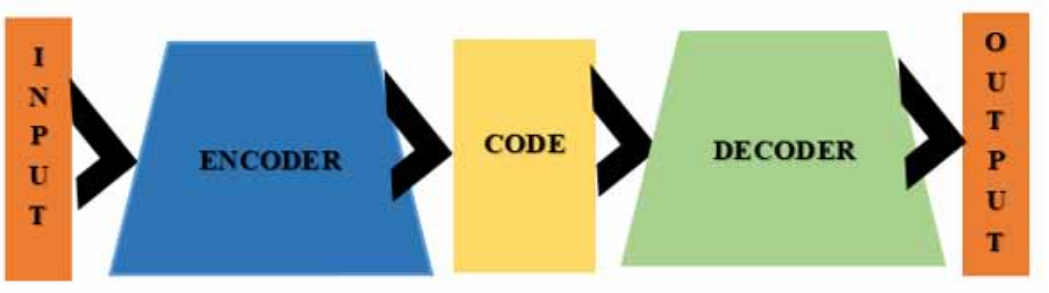
$$e = min\sum_{i=1}^{n}\left(x' - x\right)^2 \tag{3}$$

The input and output layers in Figure 5 demonstrate the design of an auto-encoder. F signifies the activation functions employed, b denotes the bias value, and w denotes the weights in the input and hidden layers in Equations (1) and (2). The variable x' specifies the regenerated form of input x when using c. Regularization can be used to reduce overfitting, which is also present in ANN as defined in equation (4).

Table 2. Five stages of CKD

Stage	Description	GFR (Glomerular Filtration Rate)
		(mL/min/1.73 m2)
1	Kidney function is normal	> = 90
2	Kidney damage is mild	60–89
3	Kidney damage is moderate	30–59
4	Kidney damage is severe	15–29
5	Established kidney failure	<= 15

Figure 5. Auto-encoder architecture



$$\min\left[\sum_{i=1}^{n}\left\{\left(x'-x\right)^{2}\right\}+rL\left(w\right)\right] \tag{4}$$

Within Equation (4), L (w) denotes the weight change parameter, whereas denotes the regularization parameter. The values of these limits are determined by a hit-or-miss approach.

Figure 6 shows a stacked autoencoder structure with two autoencoders grouped in a cascade way, which helps to reduce data dimensions by picking only the necessary features from the data. These functions can then be fed into a SoftMax classifier for classification purposes. The SoftMax layer is inserted after the stacked autoencoder to perform input data classification. It's a probability-based linear classifier that calculates the probability distribution of n distinct inputs. It simply employs the relevant attributes, which helps to increase classification efficiency.

$$F\left(x_{i}\right)=\frac{Exp\left(x\ \ i\right)}{\sum_{j=0}^{k}Exp\left(x\ \ j\right)}\text{, where i = 0, 1, …, k …} \tag{5}$$

The soft-max function is denoted by the equation (5). It computes the exponential value of the supplied input xi, as well as the sum of the exponential values of all input values and the fractional value of the soft-max function's output product. In a multi-classification situation where the output class must have the highest probability of all other classes, the soft-max function returns the probabilities of each categories.

Figure 6 depicts the process of combining auto-encoder 1 and auto-encoder 2 into a stacked auto-encoder with a soft-max classifier.

METHODOLOGY

The initial goal is to use the Chronic Kidney Disease dataset to make an early diagnosis of CKD patients with a high-risk level. This goal is important in current research because many people around the world suffer from this condition.

CKD Methodologies Phases are,

Phase 1: Deep Learning Algorithms
Phase 2: Machine Learning Algorithms
Phase 3: CKD Prediction Performance

Figure 6. Process of stacked auto-encoder with soft-max classifier

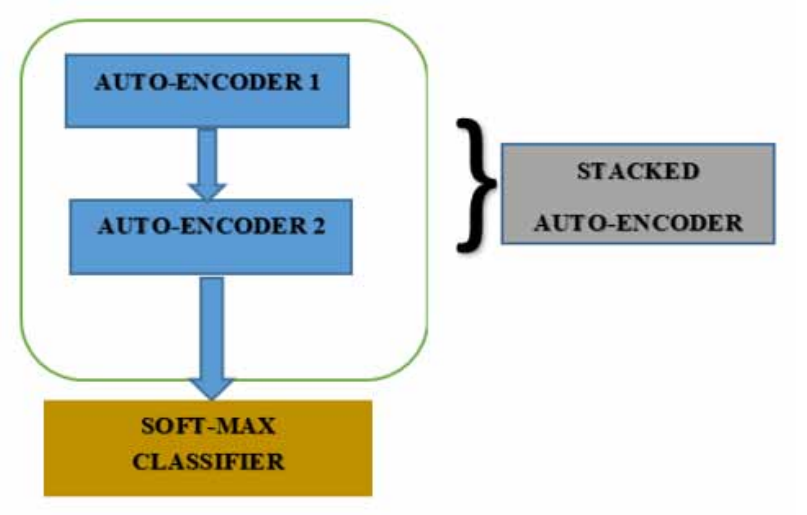
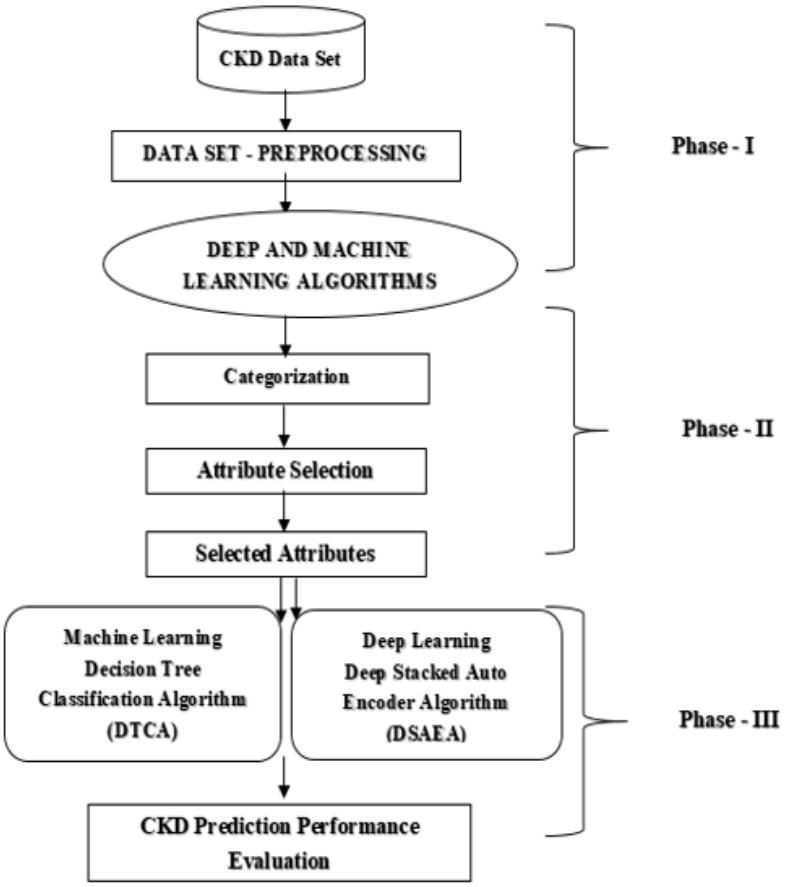


Figure 7. CKD methodology framework



Phase 1: Deep Learning Algorithms

The patient data is assessed with the collaborative healthcare data analytics block diagram of CKD for the deep stacked auto-encoder network, as shown in Figure 8.

$$Accuracy = Number\ of\ correctly\ classified\ ckd\ samples \Big/ Total\ number\ of\ ckd\ sample \tag{6}$$

$$Specificity = Number\ of\ true\ nagatives \Big/ Number\ of\ true\ nagatives + Number\ of\ false\ positives \tag{7}$$

$$Precision = Number\ of\ true\ positives \Big/ Number\ of\ true\ positives + Number\ of\ false\ positives \tag{8}$$

$$Recall = Number\ of\ true\ positives \Big/ Number\ of\ true\ positives + Number\ of\ false\ negatives \tag{9}$$

Accuracy = (20+4030)/20+0+0+4030=1=100
Precision = 20/20+0=1=100
Recall = 20/20+0=1=100
F_{1} Score = 2x 1x 1/ 1+1 = 1=100

Figure 8. CKD methodology framework: Phase I

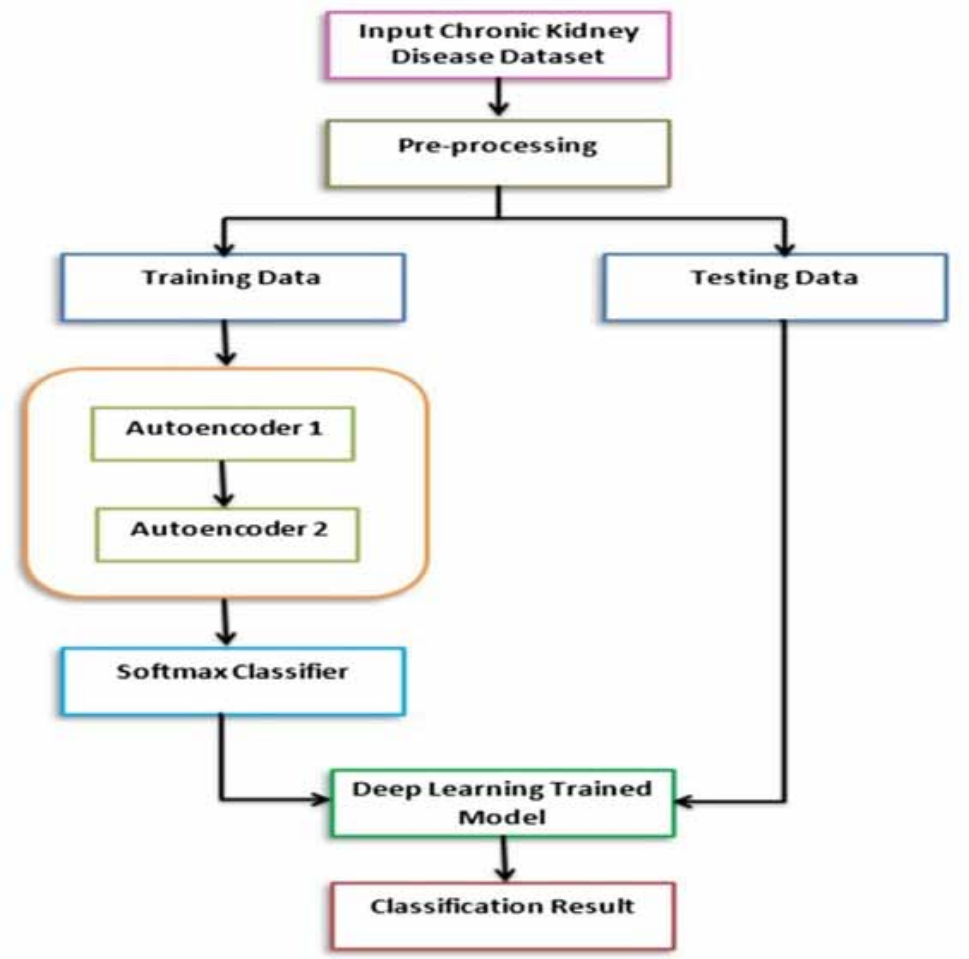


Table 3. Pseudocode for deep stacked auto-encoder network

```
ckdhiddenSize = 10;
ckdautoenc_1 = ckdtrainAutoencoder (X, ckdhiddenSize,); ckdfeatures_1 = encode (ckdautoenc_1, X);
ckdhiddenSize = 10;
ckdautoenc_2 = ckdtrainAutoencoder (ckdfeatures_1, ckdhiddenSize,); cdfeatures_2 = encode (ckdautoenc_2,
ckdfeatures_1);
ckdsoft_net = ckdtrainSoftmaxLayer (ckdfeatures_2, T,'LossFunction','crossentropy'); view(ckdsoft_net)
ckddeep_net = ckdstack (ckdautoenc_1, ckdautoenc_2, ckdsoftnet); ckddeep_net = train (ckddeep_net, X, T);
```

Figure 9. Visualization of auto-encoder network deep stacked

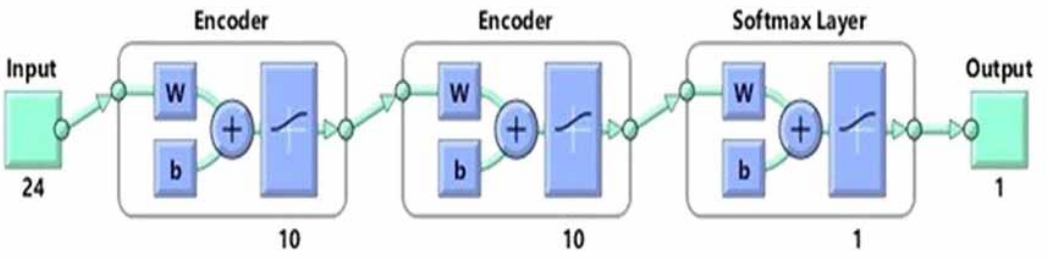


Table 4. Evaluation metrics of CKD

Evaluation Metrics of CKD	Proposed Model for Deep Stacked Autoencoder (%)
CKD_Accuracy	100
CKD_Specificity	100
CKD_Precision	100
CKD_Recall	100
CKD_F1 – score	100

Table 5. Predicted and actual result of CKD

Predicted and Actual Result of CKD		Actual Results	
Predicted results	Evaluation	Positive	Negative
	Positive	200(TP)	0(FP)
	Negative	0(FN)	103000(TN)

$$F1-score = 2 * \frac{\left(Precision * Recall\right)}{\left(Precision + Recall\right)} \tag{10}$$

In this Figure 11, blood tests to establish the estimated glomerular filtration rate (eGFR) and a urine test to evaluate albumin level are utilized to diagnose chronic kidney disease using different classification machine learning algorithms. The Random Forest decision tree classification algorithm, when compared to other classification algorithms, is superior at classifying early stage studies with less time and higher accuracy.

Figure 10. Mapping of the deep stacked auto-encoder network

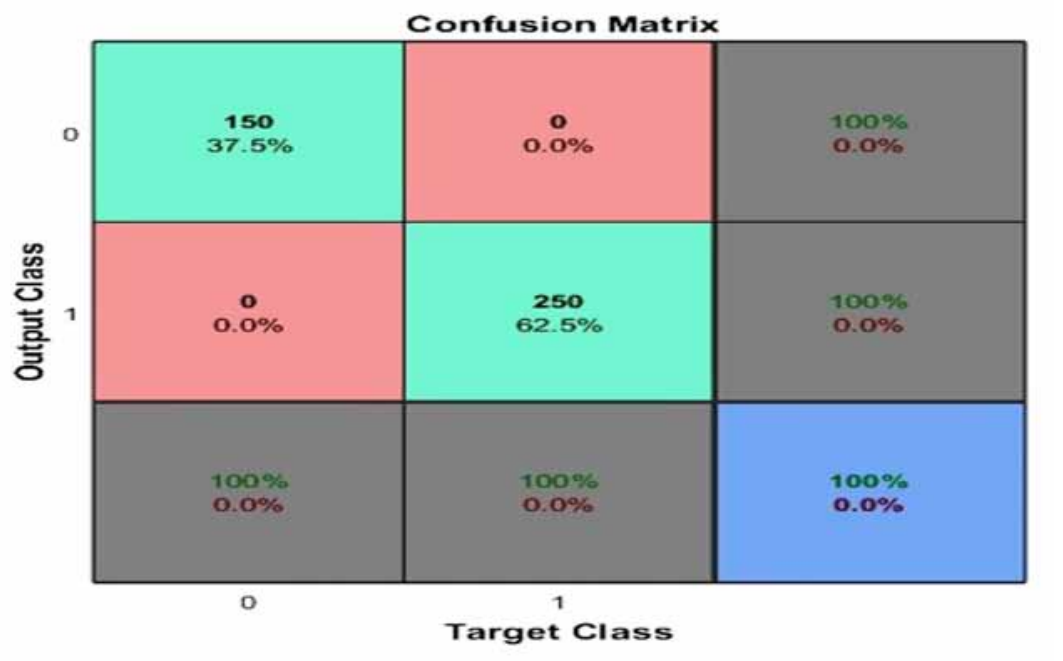
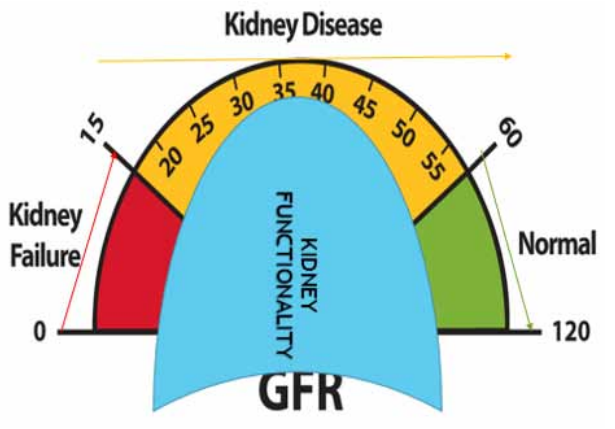


Figure 11. Kidney stages using glomerular filtration rate



Phase 2: Machine Learning Algorithms

Decision Tree Classification Algorithm

Accuracy =100
Accuracy = 100
TP + TN = Correct Prediction
FP + FN = Incorrect Prediction

Figure 12. Methodology framework for classification process of CKD machine learning algorithms

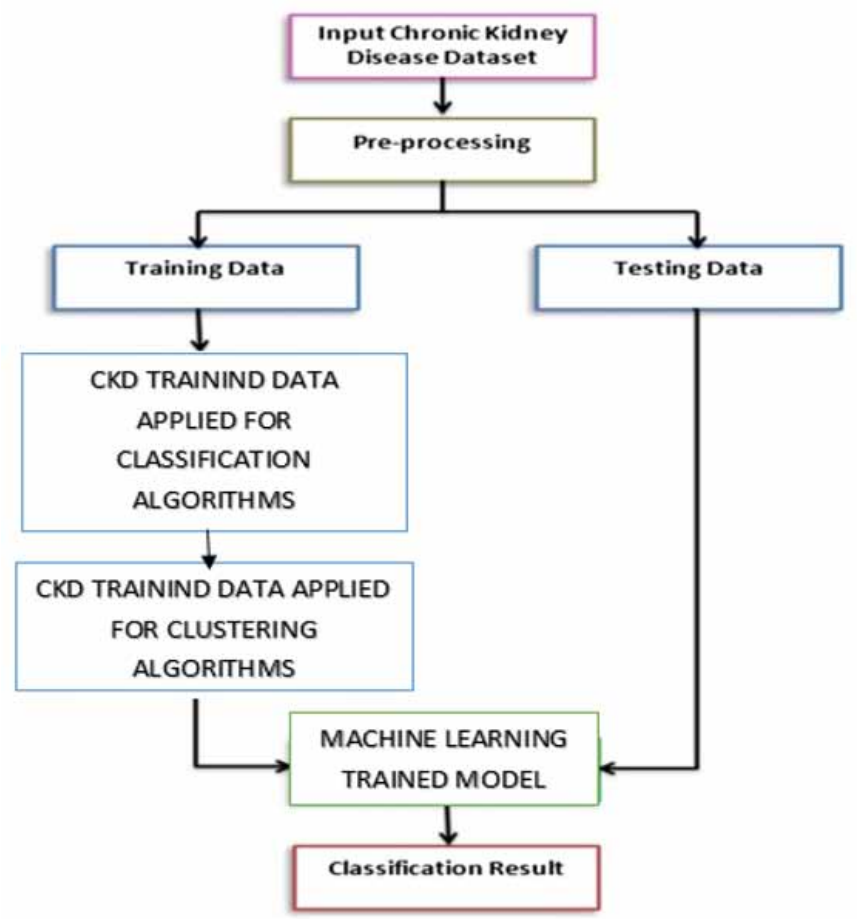


Table 6. Pseudocode for decision tree

```
Input: Chronic kidney data with selected features
Output: Classified data for Decision Tree
Step1: Apply S to D to find a splitting criterion
Step2: If (t is not a leaf node)
Step3: Create children nodes of t
Step4: Partition D into children partitions
Step5: Repeat on each partition
Step6: End
```

TP stands for True Positive, FP for False Positive, FN for False Negative, and TN for True Negative in the table above.

If a predictive model is created, the accuracy of the model is determined by the accuracy and recall values of the classification matrix, and it is critical to test its correctness.

Precision is the percentage of significant recovered instances. It's measured in terms of,

Figure 13. Decision tree classification algorithm using R based on CKD

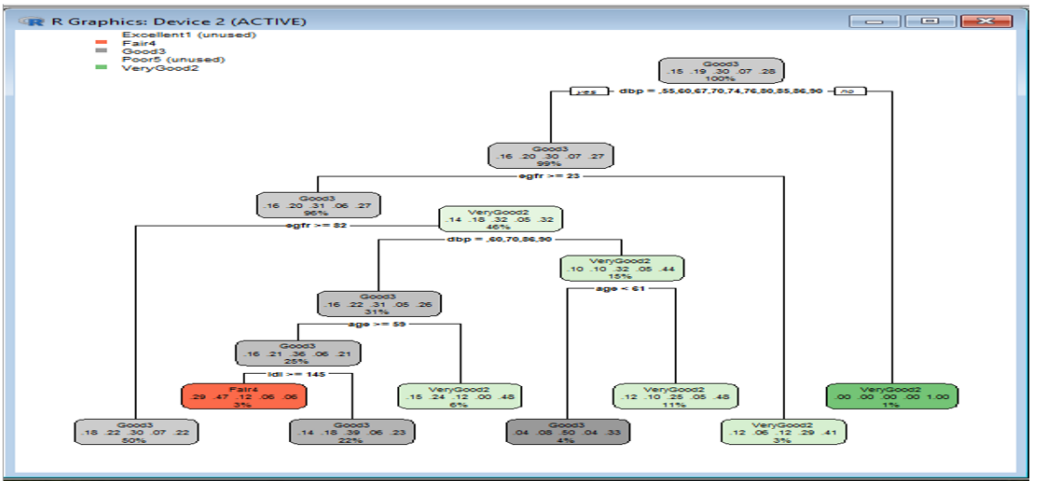
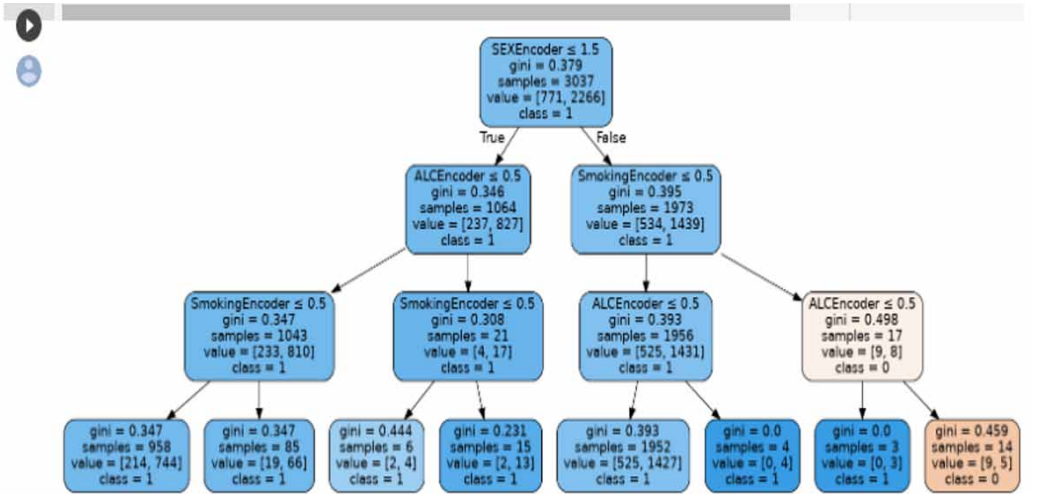


Figure 14. Decision tree classification algorithm using Python based on CKD



$$\text{Precision} = \frac{\text{T P}}{\text{T P} + \text{F P}}$$

The recall is a tiny proportion number of related retrieved instances. It is commonly stated as a percentage. It's measured in terms of,

$$\text{Recall} = \frac{\text{T P}}{\text{TP} + \text{F N}}$$

The aforementioned calculations can be used to evaluate the model's performance. The system's ability is used to make accurate forecasts. Table 9 summarizes the findings of an accuracy assessment.

Table 7. Accuracy performance

CKD Performance	Predicted Class	
ACTUAL CLASS	TP	FN
	FP	TN
Accuracy (%)		Value

Table 8. Confusion matrix

CKD Performance	Predicted Class	
	Predicated CKD	Predicated Not CKD
ACTUAL CLASS	200	0
	0	103000
Accuracy (%)		100

Table 9. Accuracy in CKD performance Using DCTA

Accuracy of CKD	100%
Sensitivity of CKD	100%
Specificity of CKD	97%

Figure 15. Accuracy performance using various machine learning algorithms

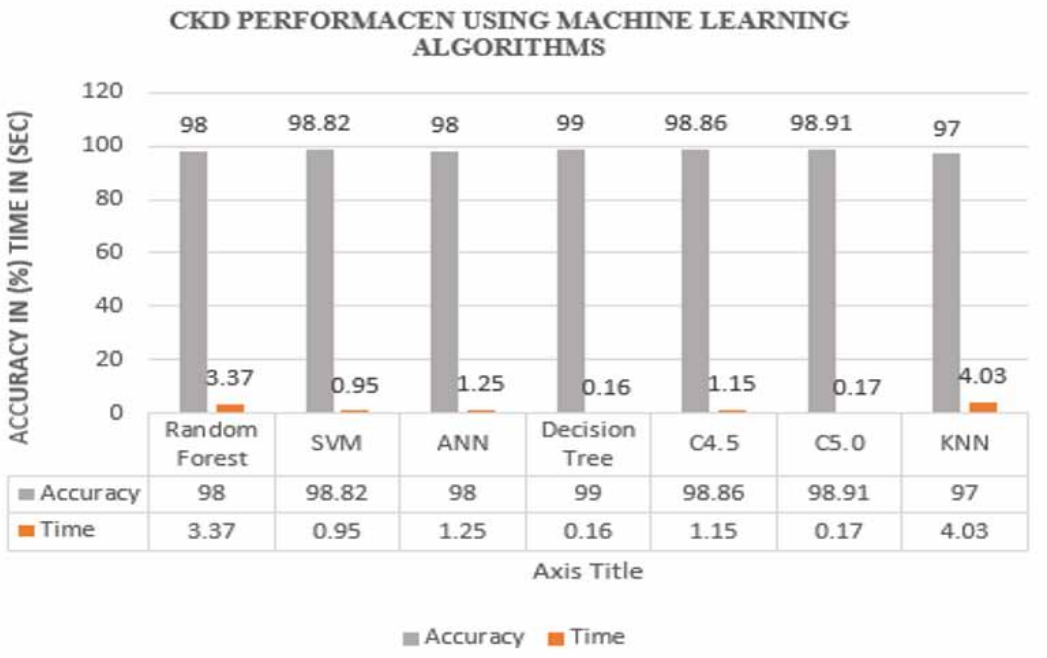
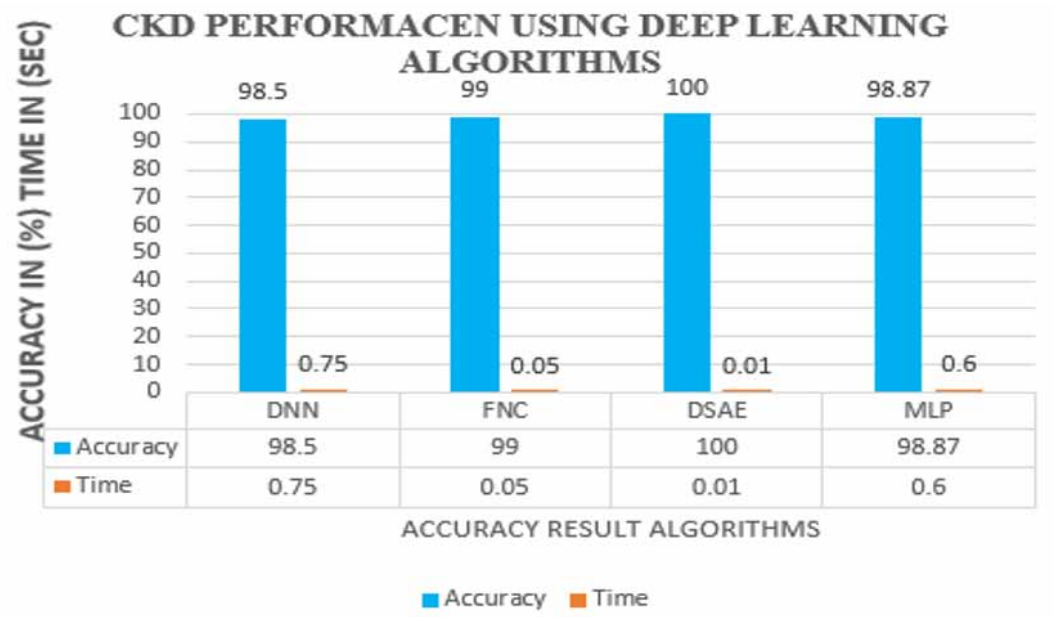


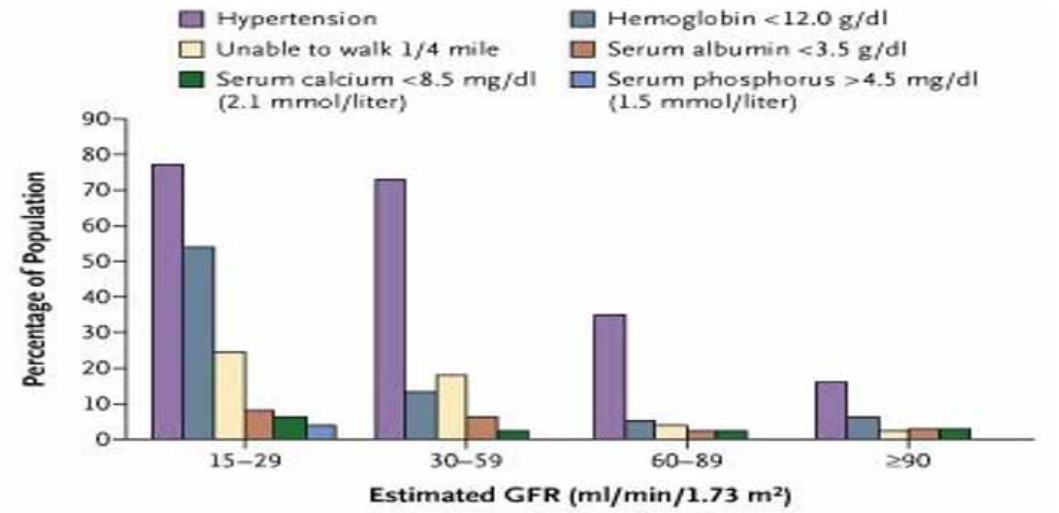
Figure 16. Accuracy performance using various deep learning algorithms



Glomerular Chronic Kidney Disease

Chronic kidney disease (CKD) is a syndrome that affects millions of people all over the world but it's a well risk factor for cardiovascular morbidity and mortality. CKD is defined as an average glomerular filtration rate (eGFR) of less than 60 mL/min/1.73 m2 and is defined as a glomerular filtration rate (eGFR) of less than 60 mL/min/1.73 m2. The purpose of this study was to use an actuarial concept of stratified hyperfiltration for age and sex to identify the cardiovascular risk involved with GHF in healthy people.

Figure 17. Level performance using glomerular chronic kidney disease



The early phases of CKD are identified in this Figure 17 based on the. It is the most advanced algorithm. As a result, the suggested system employs the random forest classifier to perform the prediction operation. The random forest decision tree classifier in the proposed system categorizes stages of chronic renal disease into specified categories such as normal (Excellent), mild (Very Good), moderate (Good), severe (Fair), and end-stage (Fair) (Poor or Failure).

Figure 18. CKD stages for DSAE algorithm

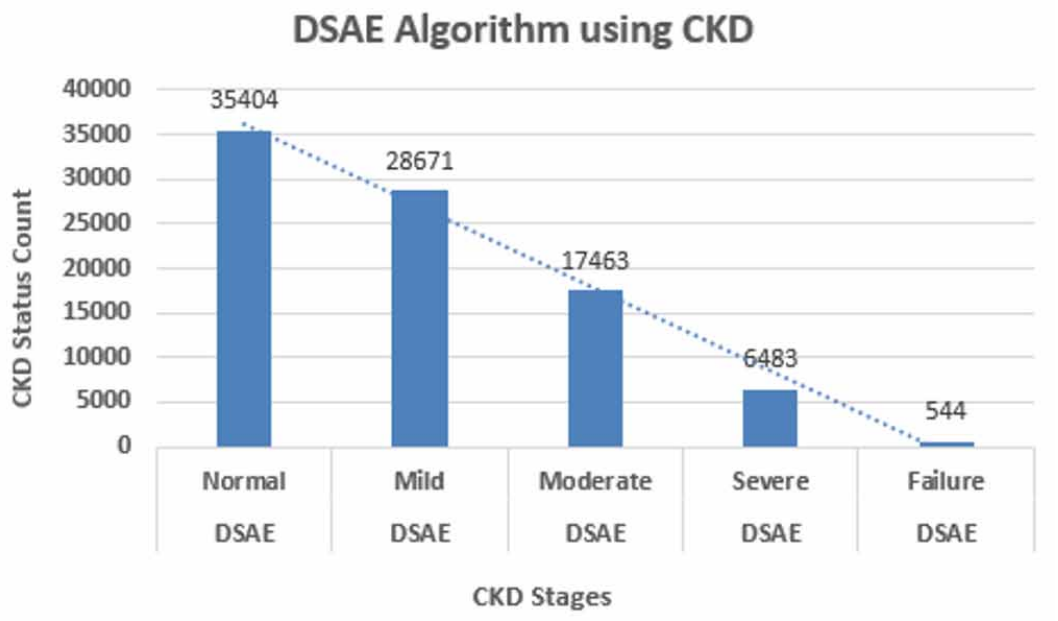


Figure 19. CKD stages for FNC algorithm

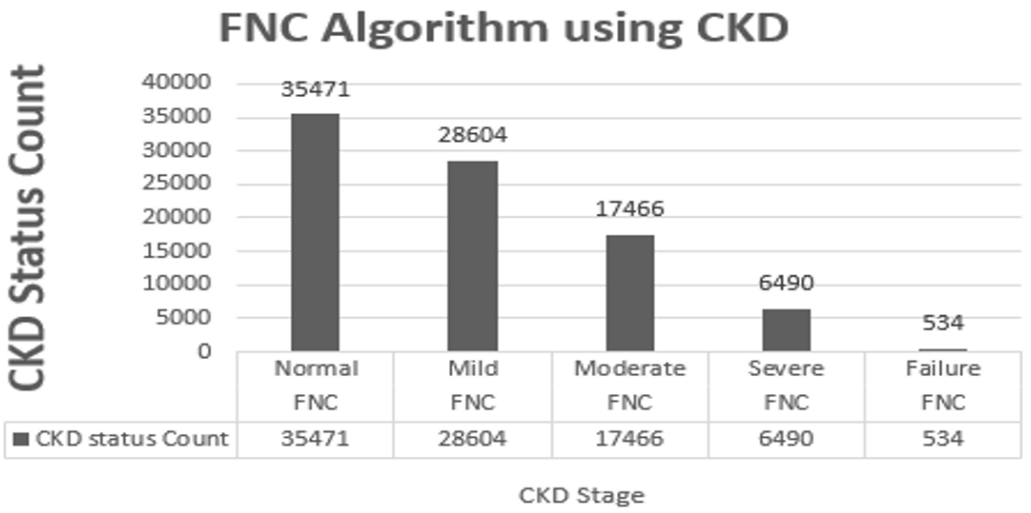


Figure 20. CKD stages for DNN algorithm

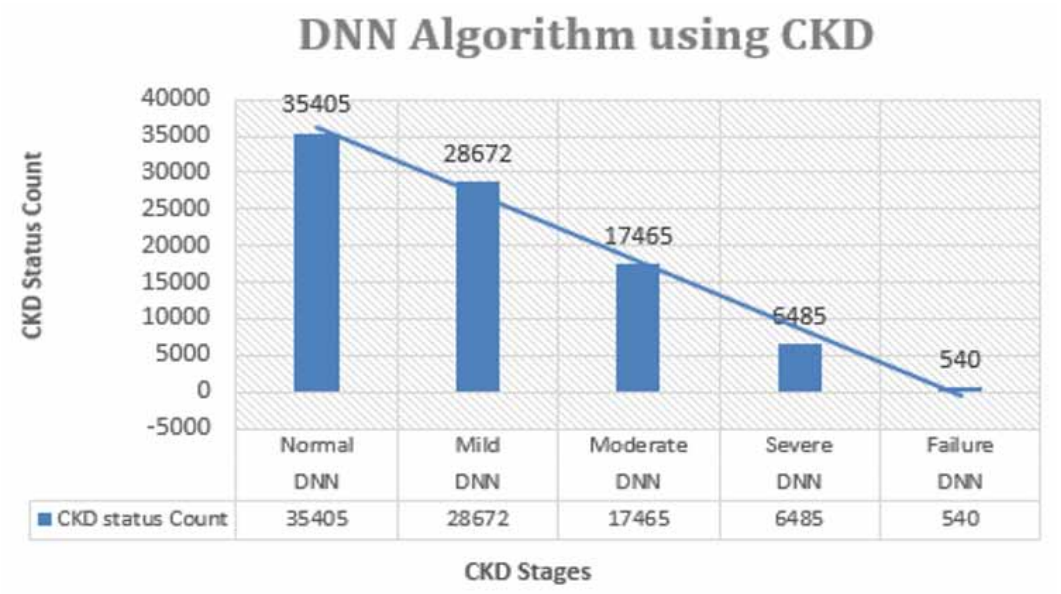
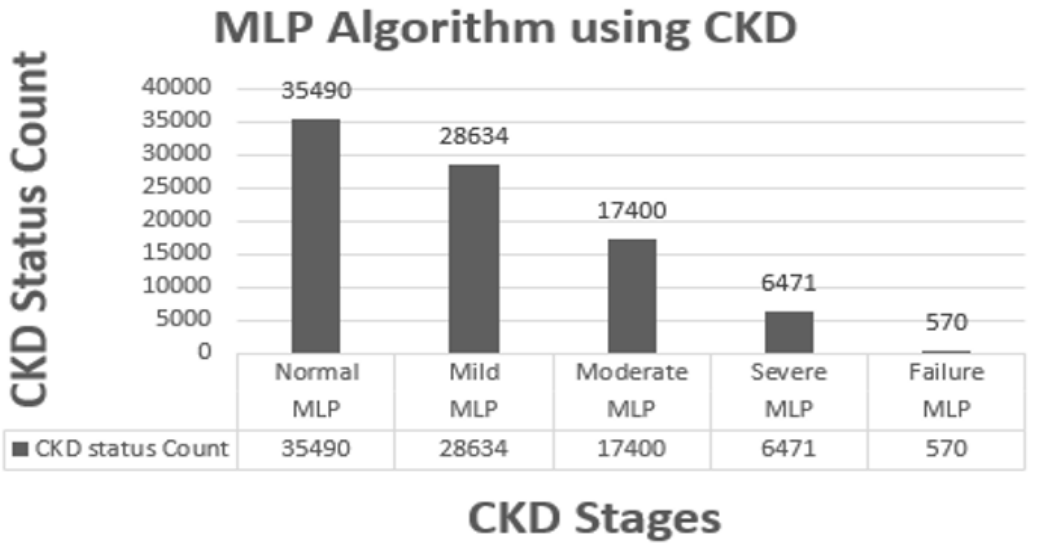


Figure 21. CKD stages for MLP algorithm



CONCLUSION

In final result, the chronic kidney disease study to approach emerging endorsements for machine learning techniques in healthcare has become a real world starting to emerge for obtaining accurate medical diagnosis results, using the machine learning techniques involved in collaborative healthcare is developing interest in a hopeful field for successful outcomes while reducing costs. This research study introduces a novel deep neural network architecture for chronic kidney disease data classification

Table 10. Literature review-based CKD benchmark dataset

Author	Approach	Description	Result
S.A.Ebiaredoh-Mienye, et al., 2020	Integrated unsupervised learnings SAE, SSR and semi supervised	- CKD -Cervical cancer and - Heart Disease	Accuracy =98%
A. Khamparia, et al., 2020	Unsupervised stacked autoencoder model using multimedia data with supervised SoftMax classifier-semi supervised deep learning framework for CKD classification	UCI dataset with 400 Data (25 attributes) The top ten most important characteristics were chosen. A feature selection was used with a stacked auto encoder.	Max Accuracy =98.89%
M. Elhoseny et al., 2019	Support Vector Machine and RBF	361 data points from a UCI dataset of 400 patients were utilized to identify five stages of CKD severity. -PNN had the best performance. -There was no feature selection process employed -Every single one of the 25 features were used 14 category and 1 numerical	Max Accuracy =96.7%
S. Gopika et al., 2017	- K-Medoids, K-Means and Fuzzy C –Means	Grouping of distinct phases of CKD based on severity -Excellent performance Fuzzy C-Means are a type of fuzzy C-Means.	Max Accuracy Fuzzy C-Means =89%
A. Ogunleye and Q.-G. Wang	XG-Boosting	No features selection algorithm was utilized All 25 features were gathered from a UCI dataset of 400 patient data with 250 CKD and 150 CKD-free cases.	Accuracy =98.7%
H. Polat et al., 2017	Support Vector Machine Classifier	The feature selection algorithm was used to wrapper method and filter method finally Best method is filter method	Accuracy =98.5%
Z. Wang et al., 2018	Association classification techniques implementing algorithms: IBk, Zero R, One R, NB, J48	UCI dataset 400 cases with 250 having CKD and 150 not having CKD Best performance by: IBk There was no feature selection algorithm applied and all features were utilized.	Accuracy =98%
Z. Rustam, et al., 2019	RF and SVM	Simulated 48 samples 36 training and 12 testing using the Gene Expression Omnibus Database Very tiny dataset	Accuracy =83.4%
Linta Antony et al., 2021	Algorithm are I-forest, K-means, DB-scan and Auto- encoder	Feature selection methods including SHAP used On the patient's medical record, there are a number of unsupervised algorithms. K-means clustering achieved the best results.	Accuracy =98.3%
Proposed Method	Algorithm Used: Supervised – C4.5, C5.0, ANN, KNN and Random Forest Unsupervised - stacked auto encoder model utilizing multimedia data with supervised SoftMax classifier-semi supervised, K-means, K-medoids, Fuzzy C-Means - Hybrid Model	Different feature selection methods including Random forest with XG-boosting used -Supervised Isolation Random Forest -Unsupervised Stacked Auto encoder -Supervised And Unsupervised algorithm on the patient electronic medical record -Real time dataset around 1030000 -Best hybrid model performance achieved by Stacked Ensample Deep Neural Network classifier outperforms the compared methods for the diagnosis of CKD	Supervised Accuracy =99.89% Unsupervised Accuracy =99.87% Hybrid Model= New Algorithm Accuracy =100% Random forest MapReduce with XG-Boosting Algorithm

utilizing a deep stacked autoencoder. The model is constructed employing stacked autoencoders, which combine two cascaded-style autoencoders with a single SoftMax classifier. The dataset of CKD patients was acquired from UCI's machine learning database and added to the suggested model. Based on the findings and observations, it can be concluded that the proposed deep stacked autoencoder model for the classification of chronic kidney disease would be an excellent tool for diagnosing CKD with improved precision, accuracy, specificity, and recall, all of which are critical in the medical field. In comparison to machine learning, the proposed deep stacked autoencoder model for the classification of chronic kidney disease would be a great tool for diagnosing CKD with improved precision, accuracy, specificity, and recall.

REFERENCES

Adam, T., Hashim, U., & Sani, U. S. (2012, February). Designing an Artificial Neural Network model for the prediction of kidney problems symptom through patient's metal behavior for pre-clinical medical diagnostic. In *2012 International Conference on Biomedical Engineering (ICoBE)* (pp. 233-236). doi:10.1109/ICoBE.2012.6179011

Aditya, K., Singh, A., Anand, D., Gupta, D., Ashish, K., Arun, K. N., & Tan, J. (2020). A novel deep learning-based multi-model ensemble method for the prediction of neuromuscular disorders. *Neural Computing & Applications*, *32*(15), 11083–11095. doi:10.1007/s00521-018-3896-0

Ahmed, S., Kabir, M. T., Mahmood, N. T., & Rahman, R. M. (2014, December). Diagnosis of kidney disease using fuzzy expert system. In *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)* (pp. 1-8). doi:10.1109/SKIMA.2014.7083522

Al-Chalabi, S., Alderson, H., Garratt, N., Green, D., Kalra, P. A., Ritchie, J., Santhirasekaran, S., Poulikakos, D., & Sinha, S. (2023). Improving outpatient clinic experience: The future of chronic kidney disease care and associated multimorbidity. *BMJ Open Quality*, *12*(3), e002188. doi:10.1136/bmjoq-2022-002188 PMID:37532458

Antony, L., Azam, S., Ignatious, E., Quadir, R., Beeravolu, A. R., Jonkman, M., & De Boer, F. (2021). A comprehensive unsupervised framework for chronic kidney disease prediction. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 126481–126501. doi:10.1109/ACCESS.2021.3109168

Bhaskar, N., Tupe-Waghmare, P., Nikam, S. S., & Khedkar, R. (2023). Computer-aided automated detection of kidney disease using supervised learning technique. *International Journal of Electrical & Computer Engineering,* 13(5).

Chetty, N., Vaisla, K. S., & Sudarsan, S. D. (2015, December). *Role of attributes selection in classification of Chronic Kidney Disease patients. In 2015 international conference on computing, communication and security*. ICCCS.

Ebiaredoh-Mienye, S. A., Esenogho, E., & Swart, T. G. (2020). Integrating enhanced sparse autoencoder-based artificial neural network technique and softmax regression for medical diagnosis. *Electronics (Basel)*, *9*(11), 1963. doi:10.3390/electronics9111963

Elhoseny, M., Shankar, K., & Uthayakumar, J. (2019). Intelligent diagnostic prediction and classification system for chronic kidney disease. *Scientific Reports*, *9*(1), 9583. doi:10.1038/s41598-019-46074-2 PMID:31270387

Gopika, S., & Vanitha, M. (2017). Machine learning approach of chronic kidney disease prediction using clustering. *Machine Learning*, *6*(7), 1–9.

Gupta, S., Garg, N., Sinha, D., Yadav, B., Gupta, B., & Miah, S. (2022). The Emerging Role of Implementing Machine Learning in Food Recommendation for Chronic Kidney Diseases Using Correlation Analysis. *Journal of Food Quality*, *2022*, 2022. doi:10.1155/2022/7176261

Kannadasan, K., Edla, D. R., & Kuppili, V. (2019). Type 2 diabetes data classification using stacked autoencoders in deep neural networks. *Clinical Epidemiology and Global Health*, *7*(4), 530–535. doi:10.1016/j.cegh.2018.12.004

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*(47-48), 35425–35440. doi:10.1007/s11042-019-07839-z

Lakshmanaprabu, S. K., Mohanty, S. N., Krishnamoorthy, S., Uthayakumar, J., & Shankar, K. (2019). Online clinical decision support system using optimal deep neural networks. *Applied Soft Computing*, *81*, 105487. doi:10.1016/j.asoc.2019.105487

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26. doi:10.1016/j.neucom.2016.12.038

Ogunleye, A., & Wang, Q. G. (2019). XGBoost model for chronic kidney disease diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *17*(6), 2131–2140. doi:10.1109/TCBB.2019.2911071 PMID:30998478

Pandey, B., Bhanodia, P. K., Khamparia, A., & Pandey, D. K. (2019). A comprehensive survey of edge prediction in social networks: Techniques, parameters and challenges. *Expert Systems with Applications*, *124*, 164–181. doi:10.1016/j.eswa.2019.01.040

Polat, H., Danaei Mehr, H., & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. *Journal of Medical Systems*, *41*(4), 1–11. doi:10.1007/s10916-017-0703-x PMID:28243816

Pujari, R. M., & Hajare, V. D. (2014, August). Analysis of ultrasound images for identification of Chronic Kidney Disease stages. In *2014 First International Conference on Networks & Soft Computing (ICNSC2014)* (pp. 380-383). doi:10.1109/CNSC.2014.6906704

Reddy, C. K., & Aggarwal, C. C. (Eds.). (2015). *Healthcare data analytics* (Vol. 36). CRC Press. doi:10.1201/b18588

Rosso, R., Munaro, G., Salvetti, O., Colantonio, S., & Ciancitto, F. (2010, August). CHRONIOUS: an open, ubiquitous and adaptive chronic disease management platform for chronic obstructive pulmonary disease (COPD), chronic kidney disease (CKD) and renal insufficiency. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 6850-6853). doi:10.1109/IEMBS.2010.5626451

Rustam, Z., Sudarsono, E., & Sarwinda, D. (2019, June). Random-forest (RF) and support vector machine (SVM) implementation for analysis of gene expression data in chronic kidney disease (CKD). *IOP Conference Series. Materials Science and Engineering*, *546*(5), 052066. doi:10.1088/1757-899X/546/5/052066

Sivasankar, E., Pradeep, R., & Sivanandham, S. (2019). Identification of important biomarkers for detection of chronic kidney disease using feature selection and classification algorithms. *International Journal of Medical Engineering and Informatics*, *11*(4), 368–385. doi:10.1504/IJMEI.2019.104981

Wang, Z., Chung, J. W., Jiang, X., Cui, Y., Wang, M., & Zheng, A. (2018). Machine learning-based prediction system for chronic kidney disease using associative classification technique. *International Journal of Engineering & Technology,* 7(4.36), 1161.