Chronic Kidney Disease Prediction Using Boruta Feature Selection Technique

[1]Muhammad Umar Yusif and [2]Zahraddeen Sufyan

[1&2] Department of Computer Science
Federal University Dutse

Corresponding Author: bakano2015@gmail.com

Abstract
Chronic kidney disease (CKD) is a life-threatening medical condition that affects millions of people globally. This occurs when both kidneys are damaged and unable to filter the blood properly. The kidneys filter the waste fluids in our blood and excrete them through urine. The aim of this paper is to investigate how the accuracy of CKD prediction will improve with minimal number of features using machine learning algorithms. Dataset from Yobe Teaching Hospital containing 400 records with 11features has been explored. Boruta was used in selecting six (6) features, classification algorithm used in machine learning was conducted using Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB). Support Vector Machine (SVM) classifier are shown the result in highest Accuracy of 98.75%, Precision 100%, F1 Score 99%, Cohen Kappa 97% and Classification Error of 0.0125%.

Keywords: Chronic Kidney Disease, Boruta, Machine Learning, Classification Error

INTRODUCTION

Chronic kidney disease (CKD) occurs when both kidneys are damaged and unable to filter the blood properly and control the levels of salt and minerals in the blood, which include calcium, phosphorus, sodium, and potassium(Priya *et al*., 2020). CKD is associated with several features such as diabetes, age, hypertension, obesity, and primary renal disorders (Nandhini & Aravinth, 2021).

The following are two well-known tests for chronic kidney disease:

a) eGFR (estimated glomerular filtration rate) value: This metric measures how well the kidneys filter blood.; if the eGFR value is greater than 90, The kidneys are healthy if the eGFR value is greater than 90. CKD is present(Mula-Abed, 2012).
b) Urine Test: The doctor may order a urine test because the kidneys produce urine, and if the urine contains blood or protein, it means that the kidneys aren't working properly(Levey *et al*., 2009).

CKD is a serious medical condition that affects millions of people worldwide. According to recent studies, the prevalence of chronic

kidney disease in Sub-Saharan Africa was 16% in west Africa, the continent's highest rate (Abd Elhafeez *et al*., 2018). CKD is responsible for about 8% to 10% of all hospital admissions in Nigeria (Ulasi & Ijoma, 2010). Additionally, there are no national of the population, diagnostic techniques are required to control the disease's prevalence.

Machine learning, on the other hand, is a key component of artificial intelligence (AI) when it comes to making predictions from a variety of data sets using the classification method. Exploration of a large dataset of diseases that can be fed into these algorithms to develop a machine learning model is what machine learning is all about (Nishat *et al*., 2018).

LITERATURE REVIEWS

Haq *et al*., (2020) proposed a machine learning model to predict diabetes disease at an early stage. It was also concluded that machine learning can play an important role in healthcare.

Desai, M.(2019) in their paper, used the Boruta algorithm to examine the factors that increase the likelihood of a patient having CKD. The findings revealed that hypertension, blood pressure, uric acid, albumin, age, and serum creatinine are some of the leading causes of CKD. Only seven of the total 24 attributes were confirmed.

In a study by Akmal *et al*., (2020) J48 and Random Forest were used to detect mobile malware, and Boruta feature selection was used to select important features; the dataset contained 215 features, were reduced to 124.

data on the prevalence of chronic kidney disease, and only a few communities have been studied, Five were discovered in the south and two in the north (Chukwuonye *et al*., 2018).

Because of the rapid growth

In another study by Pal & Parija, (2021) Random Forest algorithm was used to create a heart prediction Model using the kaggle dataset. The model had an accuracy of 86%, a sensitivity of 90.6%, and a Specificity of 82.7%.

De Almeida *et al*., (2020) created a model for detecting kidney failure using Decision Tree, Random Forest, and Support Vector Machine (SVM). As well as SVM with linear, polynomial, sigmoid, and RBF with MIMIC-II database. The conclusion was that random forest and decision tree had the best prediction accuracy of 80% and 87% respectively.

Nishat *et al*., (2018) used seven machine learning algorithms in the UCI dataset repository, with 400 instances and 25 attributes. The model was evaluated using various metrics such as accuracy, precision, sensitivity, specificity, F1-score, and ROC. Random Forest produced the highest accuracy of 99.75%.

Meghana *et al*.,(2021) used three models in the dataset: Random Forest, Support Vector Machine and Hybrid Neural Network. The dataset has 400 rows and 26 columns including the id column. The accuracy of the hybrid neural network is the highest at 97.82%. Indicating that the elderly

people with high blood pressure are at a higher risk of developing CKD.
Anantha Padmanaban and Parthiban, (2016) also used machine learning classifier algorithms to develop a system for early detection of CKD in diabetic patients. The data collected from diabetes research center in Chennai and run it through Nave Bayes and Decision Tree. The accuracy was determined using the Weka tool, and the Nave Bayes classifier achieved the highest accuracy of 91%.

Alshebly and Ahmed, (2019) applied different machine learning algorithms, including Artificial Neural Network and Logistic Regression (LR), to a problem in the domain of medical diagnosis and analyzed their efficiency of prediction on 153 cases and 11 attributes of CKD patients. The Ann's classifier performed better than the LR mode, with an accuracy of 84.44%, sensitivity 84.21%, specificity 84.61% and Area under the Curve (AUC) is 84.61%. The study discovered that creatinine and urea are the two most important factors that have a direct impact on patients with chronic kidney disease.

Shaji *et al*., (2020) used Ant Colony Optimization to apply feature selection to a chronic kidney disease dataset (ACO), a wrapper method for the heuristic algorithm for feature selection. 24 attributes and 400 records in the dataset. Following the application of the feature selection algorithm, 12 features remained for the construction of the model, and the SVM classifier was used. The model achieved 96% accuracy, 97% f1-score, 94% recall, and 100% precision.
Mtsweni *et al*., (2020) used a Neural Network Model and 10-fold cross validation to propose a model for classifying patients with CKD. Support Vector Machine, K-nearest neighbors, Decision Tree, and Gradient Boost are the classifiers used. The experiment revealed that the performance of the Neural Network with 10-fold cross validation was 98.25%, the F1-score was 98%, the Cohen Kappa was 96% and the MSE was 0.0175.

Iliyas *et al*., (2020) used a Deep Neural Network model to predict the presence and absence of CKD in 400 patients from Bade General Hospital in Yobe State, Nigeria. The model achieved accuracy 98%, precision100%, recall 96%, F1-score 98%, Cohen kappa 96.6%, ROC 100%, sensitivity 96%, and specificity 100%.

METHODOLOGY

The methodology carried out in our approach, we preprocessed the data and select the most important features. The model is trained and tested with RF, SVM and NB. The performance of algorithms is compared using different metrics.

ISSN: 2955 – 0580

Figure 1: Flow Chart of the proposed work

Data Collection

The dataset used in this research work is a chronic kidney dataset collected from Yobe Teaching Hospital for training and testing. The Python programming language with Anaconda IDE (Integrated Development Environment) is used, Scikit Learn library used to process the chronic kidney disease dataset. 400 records with 11 features. But in this work, only 6 features were selected with the help of Boruta Feature selection.

Table 1: Description of Dataset

S/N	Representation	Features	Type	Number of missing values	Percentage of missing values
1	Sex	Gender	Discrete integer values	0	0.00%
2	Age	Age	Discrete integer values	0	0.00%
3	Sod	Sodium	Numerical Values	0	0.00%
4	Pot	Pottsium	Numerical Values	0	0.00%
5	Chl	Chloride	Numerical Values	0	0.00%
6	Bica	Bicarbonate	Numerical Values	0	0.00%
7	Urea	Urea	Numerical Values	0	0.00%
8	Cre	Creatinine	Numerical values	0	0.00%
9	UA	Urea Acid	Numerical Values	356	89.00%
10	Alb	Albumin	Nominal Values	383	95.75%
11	Class	CKD or Not CKD	Nominal	0	0.00%

Iliyas *et al*., (2020)

Data Preprocessing

Data preprocessing is a basic step in training every machine learning classification algorithm and is a strategy for converting raw data into a clean dataset. Missing values are normally filled with the mean, median, or mode values of the respective features. This study adopts the use of median on numerical missing values of the entire attribute columns and mode for categorical missing values, which was shown in Table 1.The categorical values were converted into binary numbers 1 and 0 using a label encoder for model training and testing.

Feature Selection

The process of selecting a subset of the most relevant features in a dataset to describe the target variable and it becomes prominent, especially in the data sets with many variables and features. Also, the choice of features which could be based on presence or lack interaction among features and classification algorithms. Classifications algorithms with feature

selection provide better performance and reduce model execution time (Chen *et al*., 2020).

The technique searches for important features from the top down by comparing the significance of the original features to the randomly attainable features, evaluating the value of permuted duplicates (shadows), and gradually eliminating unimportant features to balance the test. Confirmed features are those that outperform shadows significantly. On each iteration generates a shadow, and the algorithm terminates when only confirmed features remain.

Boruta Algorithm

Boruta is a feature selection algorithm that functions as a wrapper for a random forest classification algorithm. Boruta algorithms employ a top-down search strategy for significant features by contrasting original attributes and shadow features reachable at random, evaluated using their permuted duplicates, and dynamically removing irrelevant features to stabilize that test. It is a popular method for variable selection in all types of datasets because it works well on both classification and regression problems and takes into account multivariable relationships.(Akmal *et al*., 2020)

Classification Algorithms

Is an important feature of supervised learning, where the classifiers learns from the training dataset and apply on the testing dataset for finding the target feature.

Support Vector Machine

Support vector machine (SVM) is a supervised learning model that examines data for classification and regression analysis.

Also, based on statistical learning, is the most robust algorithm (Shaji *et al*., 2020).The training data point is mapped in an n-dimensional space by the Support Vector Machine (where n is the number of features in the dataset). The goal is to find the best hyperplane for classifying the data points into their appropriate categories. The hyperplane is the perpendicular bisector of the shortest line connecting the two classes. The decision boundary that separates the two classes is the hyperplane, whose dimension is determined by the number of input features in the dataset. Support vectors are training instances that are closest to the hyperplane.

$$f(x) = \beta_o + \sum_{i \in S} \alpha i K(xi, xi) \qquad \textbf{(1)}$$

β = Bias

S = observations

α = model parameter learned

Naïve Bayes

The Nave Bayes algorithm is one of the most efficient and effective inductive learning algorithms for machine learning and data mining. The Bayes rule of conditional probability is involved. It assumes that a class's parameters are independent of one another; each parameter contributes independently to the prediction of the final result. This assumption can sometimes have a negative impact on the model's accuracy. It's simple to use and, in some cases, outperforms more complicated algorithms. It's a robust

and probabilistic representation that can handle null values.

$$P\ (E \mid H) = (E1 \mid H) * (E2 \mid H) * (E3 \mid H) * \ldots * (En \mid H) * (H) \ / \ (E) \quad \textbf{(2)}$$

H stands for hypothesis (an output attribute), and E stands for evidence (set of input attributes).

Random Forest

The random forest model is a non-parametric model that can be used for classification. Is one of the most popular ensemble methods and belongs to the bagging method category. Random forest is an ensemble classifier that employs decision trees to improve prediction results. Many trees are being formed while the bootstrap technique is being applied to each tree from the set of training data. In a classification problem, input is given to each tree in the forest, each tree representing an individual class. Finally, the random forest chooses the class with the most votes.

Data Splitting

Data splitting is critical in machine learning and is widely used. The data was subdivided by the algorithm into subsets, training and testing. The training test is used to fit the model, while the testing set is used to evaluate it. The data in this work is divided into two parts: 80% for training and 20% for testing.

Performance Evaluation

- **Accuracy**

Is a model's ability to detect the ratio of true predictions to total number of instances, and it can be defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP+TN+FP+FN}} \quad \textbf{(3)}$$

In the medical field, accuracy is not a sufficient metric for assessing model performance. As a result of not taking into account the incorrectly predicted cases, I proposed the use of other metrics such as Precision, F1-measrue, Kohen Kappa, and Classification Error.

- **Precision**

A metric for correctness that can be defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \textbf{(4)}$$

- **F1-Score**

This is calculated as the harmonic mean of recall and precision. It can also be written as

$$\text{F1-measure} = \frac{2*Precision + Recall}{\text{precision + Recall}} \quad \textbf{(5)}$$

(5)

- **Kohen Kappa**

Kohen Kappa is a Statistical method for determining inter-rate reliability for quantitative items in a classifier.

$$\text{Is define as} \quad \text{K} = \frac{\mathbf{P(o) - P(E)}}{\mathbf{1 - P(E)}}$$

(6)

Where Po indicate relative observed agreement among the raters

Where Pe, is the probability of agreement by chance.

- **Classification Error**

Classification error shows the incorrect rate of prediction results it computes from the confusion matrix. Is define as

$$\text{Error} = \frac{\text{FP+FN}}{\text{TP +TN+FP+FN}} \qquad (7)$$

Table 2: Confusion Matrix

	Positive(CKD)	Negative (NOTCKD)
Positive (CKD)	TP	FN
Negative (NOTCKD)	FP	TN

True Positive: a person with chronic kidney disease was correctly identified as having chronic kidney disease.

False Positive: a non-chronic kidney disease person was mistakenly identified as having chronic kidney disease.

True Negative: The person was correctly identified as having non-chronic kidney disease.

False Negative: A person with chronic kidney disease was mistakenly identified as having non-chronic kidney disease.

RESULT AND DISCUSSION

Result

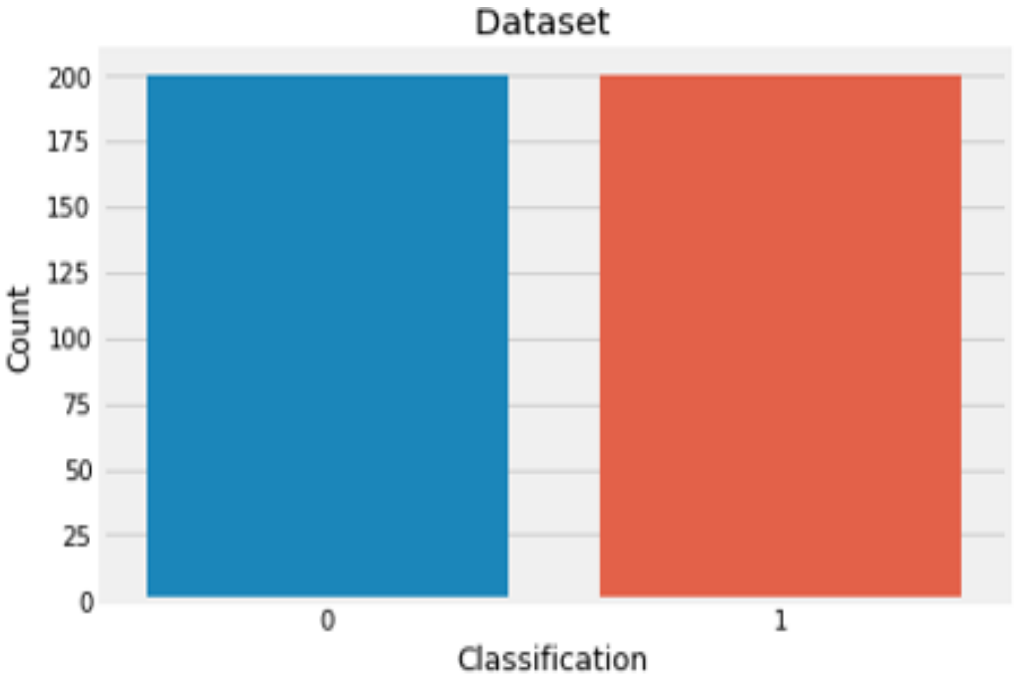

Figure 2: Dataset Distributions

Table 3: Ranking of significant features of Dataset

Serial Number	Feature	Ranking
1	Age	1
2	Sod	1
3	Pot	1
5	Bica	1
6	Urea	1
7	Cre	1
4	Chl	2
9	UA	3
10	Sex	4
8	Alb	4

Table 4: Confusion Matrix of Classifiers with Feature Selection on Dataset

S/N	Classifiers	Desired Result	Prediction	
			Not ckd	ckd
1	**Random Forest**	Not ckd	**34**	**2**
		ckd	**0**	**44**
2	**Support Vector Machine**	Not ckd	**34**	**0**
		ckd	**1**	**45**
3	**Naïve Bayes**	Not ckd	**32**	**4**
		ckd	**0**	**44**

Table 5: Prediction Accuracy and Time of the Classifiers on Dataset

S/N	Classifiers	Accuracy	Time (s)
1	**Random Forest**	97.5%	0.5s
2	**Support Vector Machine**	**98.75%**	**0.2s**
3	**Naïve Bayes**	95%	0.5s

Table 6: Result of the Classifiers on Dataset

S/N	Classifiers	Precision	F1-Measure	Cohen kappa	Classification Error
1	**Random Forest**	96%	98%	94%	0.025
2	**Support Vector Machine**	100%	99%	97%	0.0125
3	**Naïve Bayes**	92%	96%	89%	0.05

Table 7: Comparison of the proposed work

S/N	Proposed Classifiers	Number of features	Accuracy	Precision	F1-Score	Cohen Kappa	Classification Error	Time (s)
1	Random Forest	6	97.5%	96%	98%	94%	0.025	0.5
2	***Support Vector Machine***	***6***	***98.75%***	***100%***	***99%***	***97%***	***0.0125***	***0.2***
3	Naïve Bayes	6	95%	92%	96%	89%	0.05	0.5

Table 8: Comparison of Our Work with the Existing Work

S/N	Author	Model	Dataset used	Purpose	FS Approach	Number of Features	Metrics %
1	Reshma *et al.*,2020	Support vector machine (SVM) and Ant Colony Optimization (ACO)	UCI repository Dataset	Chronic kidney disease prediction	Yes	11	Accuracy 96%
2	Deepika (2020)	k-nearest neighbor and Naïve Bayes	UCI repository with 400 records and 25 features	Early prediction of chronic kidney disease prediction	No	24	Accuracy 97%
3	Ilyas *et al.*, (2020)	Deep Neural Network	General Teaching Hospital Gashua, Yobe State. With 400 records and 11 features	Prediction of chronic kidney disease	No	10	Accuracy 98% Precision 100% F1 Score 98% Cohen Kappa 96%

4	Priya *et al.*, (2021)	Neural Network	Kaggle dataset with 400 record and 25 features	Chronic kidney prediction	NO	13	Accuracy 95.36
5	Meghana *et al.*, (2021)	Hyrid Neural Network, support Vector machine (SVM) and random Forest	Kaggle Dataset, with 400 records and 18 features	Chronic kidney prediction	YES	18	Accuracy 97.52% Precision 90% F1-score 96%
6	Shawni(2020)	Neural Network , support vector machine, k-nearest neighbor, Decision Tree and GDboost	UCI repositoryData set with 400 records and 25 features	Chronic Kidney Disease Prediction	NO	24	Accuracy 98% with 10 k-fold cross validation F1-score 98% Cohen Kappa 96% Mean square error 0.0175
7	Abhijitb *et al.*, (2021)	Support vector machine and Naïve bayes	Kaggle Dataset with 400 records and 25 features	Chronic kidney disease prediction	NO	24	Accuracy 61.25% F1 measure 97% Precision 95% Recall 96% Time NB is 0.04, SVM is 0.4
	Our proposed work	***Support Vector Machine with Boruta Algorithm***	***General Hospital Yobe State***	***CHRONIC KIDNEY DISEASE PREDCTION USING BORUTA FEATURE SELECTION TECHNIQUE***	YES	***6***	***Accuracy 98.75% Precision 100% F1 Score 99% Cohen Kappa 97% Classification Error 0.0125***

Discussion

Figure 2 depicted the distributions of the chronic kidney disease dataset, which had the same size of 200 CKD and 200 non-CKD patients. Table 3 depicts how the Boruta algorithm selects the most significant features by ranking them with integer numbers, 1 means the feature is selected, otherwise is rejected, as well as a description of the features. Only six features are chosen.

Table 4 display the confusion matrix of the proposed model. The model

succeeded in classifying 34 samples correctly as ckd and 45 not ckd. The model fail to classify 1 sample correctly which shown the model is good for ckd prediction.

Table 5 display the accuracy and execution time, the proposed model support vector machine has the promising accuracy of 98.75% and 0.2s execution time, which is relative low compared to other models.

Table 6 display the result of the model using confusion matrix.

Table 7 represent the comparison of proposed work, where the proposed model support vector machine outperformed with the highest accuracy, precision,f1-score, cohen kappa and classification error.

Table 8 shows the comparison of proposed model with the existing work. The proposed model achieved the highest accuracy with minimal number of features using Boruta feature selection techniques. Where the article number [3] in the table above provide the accuracy of 98% with 11 features on the same dataset.

Conclusion

We divided the data into 80% training and 20% testing. The explanation on the previous pages demonstrated that the entire experiment was carried out and recorded. The most important feature has been selected using Boruta feature selection which is wrapper method, Random Forest, Naïve Bayes and Support vector Machine was applied for the dataset. Support Vector Machine provide the promising accuracy of 98.75 which is the highest , while using the minimal number of features which is 6. Performance of the model can be improved using other feature selection techniques in the future.

REFERENCES

[1] Abd Elhafeez, S., Bolignano, D., D'Arrigo, G., Dounousi, E., Tripepi, G., & Zoccali, C. Prevalence and burden of chronic kidney disease among the general population and high-risk groups in Africa: A systematic review. *BMJ Open*, *8*(1). https://doi.org/10.1136/bmjopen-2016-015069. 2018.

[2] Akmal, C., Yahaya, C., Firdaus, A., Mohamad, S., Ernawan, F., Faizal, M., & Razak, A. Automated Feature Selection using Boruta Algorithm to Detect Mobile Malware. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(5), 9029–9036. https://doi.org/10.30534/ijatcse/2020/307952020. 2020.

[3] Alshebly, O. Q., & Ahmed, R. M. Prediction and factors affecting of chronic kidney disease diagnosis using artificial neural networks model and logistic regression model. *Iraqi Journal of Statistical Science*, *June*. https://www.researchgate.net/profile/Omar_Al-Shebly/publication/336669325_Prediction_and_Factors_Affecting_of_Chronic_Kidney_Disease_Diagnosis_using_Artificial_Neural_Networks_Model_and_Logistic_Regression_Model/links/5dab0c70299bf111d4bed632/Prediction-a.

2019.

[4] Anantha Padmanaban, K. R., & Parthiban, G. Applying machine learning techniques for predicting the risk of chronic kidney disease. *Indian Journal of Science and Technology*, *9*(29). https://doi.org/10.17485/ijst/2016/v9i29/93880.2016.

[5] Chen, R. C., Dewi, C., Huang, S. W., & Caraka, R. E. Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00327-4.2020.

[6] Chukwuonye, I. I., Samuel Ogah, O., Ndukaife Anyabolu, E., Arinze Ohagwu, K., Collins Nwabuko, O., Onwuchekwa, U., Erinma Chukwuonye, M., Chukwuebuka Obi, E., & Oviasu, E. Prevalence of chronic kidney disease in Nigeria: Systematic review of population-based studies. *International Journal of Nephrology and Renovascular Disease*, *11*, 165–172. https://doi.org/10.2147/IJNRD.S162230.2018.

[7] De Almeida, K. L., Lessa, L., Peixoto, A., Gomes, R., Celestino, J., Kidney, J. C., De, K. L., Almeida, A., Peixoto, A. B. S., & Gomes, R. L. Kidney Failure Detection Using Machine Learning Techniques. *… on Advances in …*, 1–8. https://hal.archives-ouvertes.fr/hal-02495264.2020.

[8] Desai, M. Early detection and prevention of chronic kidney disease. *Proceedings - 2019 5th International Conference on Computing, Communication Control and Automation, ICCUBEA 2019*. https://doi.org/10.1109/ICCUBEA47591.2019.9128424.2019.

[9] Haq, A. U., Li, J. P., Khan, J., Memon, M. H., Nazir, S., Ahmad, S., Khan, G. A., & Ali, A. Intelligent machine learning approach for effective recognition of diabetes in e-healthcare using clinical data. *Sensors (Switzerland)*, *20*(9). https://doi.org/10.3390/s20092649.2020.

[10] Iliyas, I. I., Rambo Saidu, I., Dauda, A. B., & Tasiu, S. *Prediction of Chronic Kidney Disease Using Deep Neural Network*.2020.

[11] Levey, A. S., Cattran, D., Friedman, A., Miller, W. G., Sedor, J., Tuttle, K., Kasiske, B., & Hostetter, T. Proteinuria as a Surrogate Outcome in CKD: Report of a Scientific Workshop Sponsored by the National Kidney Foundation and the US Food and Drug Administration. *American Journal of Kidney Diseases*, *54*(2), 205–226. https://doi.org/10.1053/j.ajkd.2009.04.029.2009.

[12] Meghana L, H., Kuber, V. S., S, Y. B., L, V. T., & M, V. B. *Chronic Kidney Disease Prediction using Neural Network and ML Models*. www.ijert.org.2021.

[13] Mtsweni, E. S., Hörne, T., Poll, J. A. van der, Rosli, M., Tempero, E., Luxton-reilly, A., Sukhoo, A., Barnard, A., M. Eloff, M., A. Van Der Poll, J., Motah, M., Boyatzis, R. E., Kusumasari, T. F., Trilaksono, B. R., Nur Aisha, A., Fitria, -, Moustroufas, E., Stamelos, I., Angelis, L., … Khan, A. I. No 主観的健康感を中心とし

た在宅高齢者における 健康関連指標に関する共分散構造分析Title. *Engineering, Construction and Architectural Management*, *25*(1), 1–9. http://dx.doi.org/10.1016/j.jss.2014.12.010%0Ahttp://dx.doi.org/10.1016/j.sbspro.2013.03.034%0Ahttps://www.iiste.org/Journals/index.php/JPID/article/viewFile/19288/19711%0Ahttp://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.678.6911&rep=rep1&type=pdf%0Ahtt.2020.

[14] Mula-Abed, W. A. S. Estimated Glomerular Filtration Rate (eGFR): A serum creatinine-based test for the detection of chronic kidney disease and its impact on clinical practice. *Oman Medical Journal*, *27*(4), 339–340. https://doi.org/10.5001/omj.2012.87.2012

[15] Nandhini, G., & Aravinth, J. Chronic kidney disease prediction using machine learning techniques. *2021 6th International Conference on Recent Trends on Electronics, Information, Communication and Technology, RTEICT 2021*, *September*, 227–232. https://doi.org/10.1109/RTEICT52294.2021.9573971.2021.

[16] Nishat, M., Faisal, F., Dip, R., Nasrullah, S., Ahsan, R., Shikder, F., Asif, M., & Hoque, M. A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, 170671. https://doi.org/10.4108/eai.13-8-2021.170671.2018.

[17] Pal, M., & Parija, S. Prediction of Heart Diseases using Random Forest. *Journal of Physics: Conference Series*, *1817*(1). https://doi.org/10.1088/1742-6596/1817/1/012009.2021.

[18] Priya, S., S.Nirmal, K., G.Sibi, S., E., P., & Ashuthosh, K. P. *Chronic Kidney Disease Prediction Using Data Mining Algorithms*. 92–111. https://doi.org/10.4018/978-1-7998-2742-9.ch006.2020.

[20] Shaji, S., Ajina, S. R., & Priya R, V. S. (n.d.). *Chronic Kidney Disease Prediction using Machine Learning*. www.ijert.org

[21] Ulasi, I. I., & Ijoma, C. K. The enormity of chronic kidney disease in Nigeria: The situation in a teaching hospital in south-east Nigeria. *Journal of Tropical Medicine*, *2010*. https://doi.org/10.1155/2010/501957.2010.