International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 882-885 www.ijaem.net ISSN: 2395-5252

Risk Forecast of Chronic Kidney Disease Using ANN

Rashmi N[1], Ankitha R Desai[2], Mukundha U V[3], Rachna G S[4*]
[1] *Asst.Prof, Department of Computer Science and Engineering, Dr Ambedkar Institute of Technology, Bengaluru*
[2,3,4,] *Student, Department of Computer Science and Engineering, Dr Ambedkar Institute of Technology, Bengaluru*

Submitted: 10-07-2022 Revised: 17-07-2022 Accepted: 21-07-2022

ABSTRACT: One of the most serious illnesses nowadays is chronic kidney disease, for which a definitive diagnosis must be made as soon as possible. The use of machine learning in healthcare has increased. The doctors can identify the ailment early by applying machine learning classifier algorithms. The data on chronic renal disease were compiled using the UCI repository. The seven classifier techniques utilized in this study included the artificial neural network, C5.0, logistic regression, linear support vector machine with penalty L1 and L2, random tree, and logistic regression. The dataset also utilized the appropriate feature picking method. The outcomes of each classifier have been computed based on the following characteristics: Synthetic minority over-sampling with full features, least absolute shrinkage and selection operator regression, correlation-based feature picking, wrapper method feature picking, and least absolute shrinkage and selection operator regression are all examples of feature picking techniques. According to the findings, LSVM with penalty L2 provides the best accuracy. Again, the linear support vector machine provided the maximum accuracy of 98.46 percent in the synthetic minority over-sampling technique with the least absolute shrinkage and selection operator selected features. On the same dataset, a single deep neural network (ANN) was used as machine learning model and it attained the maximum accuracy of 99.6%.

Keywords: Chronic Kidney Disease, Artificial Neural Network, Machine Learning Classifiers.

I. INTRODUCTION

If you have chronic renal illness, your kidneys are unhealthy and aren't properly filtering your blood (CKD). Since the kidneys' primary job is to filter extra water and waste from the bloodstream so that urine may be generated, CKD is a symptom that the body's wastes have accumulated. Because the harm was done gradually over a lengthy period of time, this illness tends to be fatal. The fact that it's a common illness is flattering. CKD can result in a few health problems. CKD is caused by a variety of conditions, including diabetes, hypertension, and heart disease. Age and gender have an impact on CKD in addition to these major illnesses. If your kidneys are not working properly, you may suffer one or more symptoms like back pain, nausea, diarrhea, fever, nosebleeds, and vomiting..

The two primary inducements that cause CKD are diabetes and hypertension. Therefore, preventing these diseases also prevent the development of CKD. Typically, CKD symptoms do not appear until the kidneys have suffered severe damage. Studies show that while hospitalizations are rising 6.23% per year, the global mortality rate is staying the same. Cases of CKD are increasing quickly. To assess the stage of CKD, these diagnostic techniques are available.: (i) estimated glomerular filtration rate (eGFR) (ii) urine screening (iii) BP test.

II. METHODOLOGY

The project will be developed on Python coding. We collect the dataset from the specific website and train them using machine learning algorithm (ANN). The website will be created where we can give the inputs of the patient.

Machine Learning Models:

In this research, we have come about a prototype to forecast the risk of CKD disease in patients. There were several ways for selecting important features, including Wrapper, Filter, and Embedded. Artificial neural networks, logistic regression, linear support vector machines, KNN, and random trees were among the machine learning classifiers used to train the model. The validation and performance matrices for each classifier were calculated. The research process consists of five

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 882-885 www.ijaem.net ISSN: 2395-5252

stages: (i) dataset preprocessing (ii) feature picking (iii) classifier implementation (iv) SMOTE and (v) analyzing the classifier's execution. In order to compare their results to those of those models, deep neural networks were deployed in combination with machine learning techniques. For this, an artificial neural network classifier was employed.

Wrapper Method:

In the wrapper method, a particular machine learning method is used to select the subset of characteristics. It uses the greedy search strategy to discover a potential subset of features. The technique can be carried out using the forward selection, backward elimination, and recursive elimination algorithms. We applied the forward feature picking method during the research. The forward feature picking method is used to iteratively pick the feature. This method begins with a null model and gradually increments attributes after each iteration. The attribute is retained in the model as long as it stops improving the model performance.

Artificial Neural Network:

Artificial intelligence includes artificial neural networks. This kind of machine learning is supervised. Similar to the human brain, it has the same structure. ANNs have neurons, just like humans, and layers of the network connect ANN neurons to one another, similarly to human neurons. There, neurons are referred to as nodes. An issue that has eluded human or statistical criteria can now be resolved by ANN. The input, hidden, and output layers make up an ANN. Input and weight are passed from the input layer to the hidden layer, which performs calculations and looks for hidden structures and patterns.

Linear Support Vector Machine (LSVM):

Linear support vector machine (LSVM) is a contemporary machine learning algorithm which is very speedy and is utilized for solving multiclass segregation problem for huge datasets and is based on a plain iterative method. It has given rise to the SVM framework in linear CPU time of the dataset. LSVM can be of use while working with multi-dimensional datasets both in scanty and populated format. It is convenient for using low-cost computer resources to solve machine learning challenges with huge datasets. SVM of the supervised classifier algorithm type. For resolving the segregation/classification issue, it employs the kernel trick. The best edge among the potential outputs is identified using these changes. SVM may be applied to nonlinear kernels like RBF. For the linear kernel, LSVM suits perfectly. LSVM classifier suffices all linear problems.

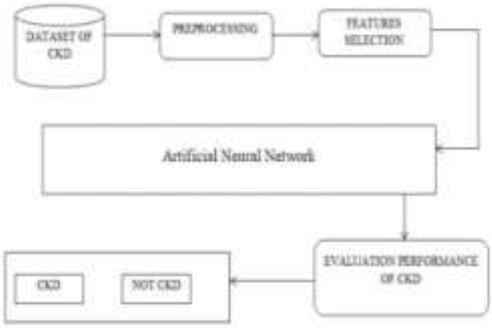

Fig. 1. System design of proposed method.

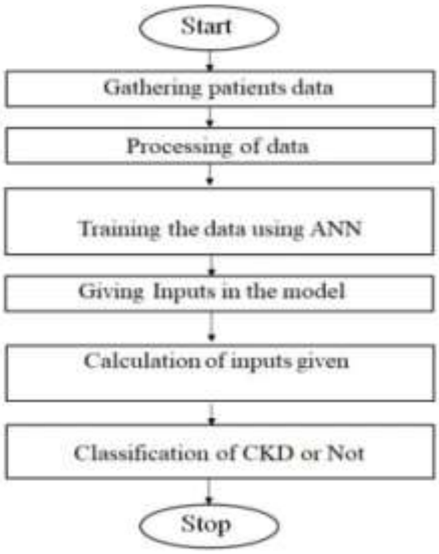

Fig. 2. Dataflow diagram.

III. IMPLEMENTATION

Python is a high-level, conjoint and object-oriented scripted language in which instructions are executed without being converted into machine-level language. It is written to be thoroughly intelligible. It has English keywords occurring routinely as opposed to other languages that use punctuations and it has lesser syntactical build ups than other languages. Since it provides extensive libraries and is a high-level language, we have implemented the project in Python.

Module Implementation:

Homepage Module- This module includes the website homepage with CKD details.

Predictor Module- Here we will be giving inputs of the patient details from the dataset.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 882-885 www.ijaem.net ISSN: 2395-5252

Result Module- This module will display the predicted result.

IV. EXPERIMENTAL RESULTS

Table 1**:** Comparing accuracies of different types of classifiers.

Decision Tree Classifier

Accuracy of Decision Tree Classifier(Test Size 0.3): 96.67%

	precision	recall	f1-score	support
ckd	0.97	0.97	0.97	74
notckd	0.96	0.96	0.96	46
accuracy			0.97	120
macro avg	0.96	0.96	0.96	120
weighted avg	0.97	0.97	0.97	120

Random Forest Classifier

Accuracy of Random Forest Classifier(Test Size 0.3): 99.17%

	precision	recall	f1-score	support
ckd	0.99	1.00	0.99	74
notckd	1.00	0.98	0.99	46
accuracy			0.99	120
macro avg	0.99	0.99	0.99	120
weighted avg	0.99	0.99	0.99	120

MLP Classifier

Accuracy of MLP Classifier(Test Size 0.3): 90.83%

	precision	recall	f1-score	support
ckd	0.91	0.95	0.93	74
notckd	0.91	0.85	0.88	46
accuracy			0.91	120
macro avg	0.91	0.90	0.90	120
weighted avg	0.91	0.91	0.91	120

========================== Test Size 20.0% ==========================

Logistic Regression

Accuracy of Logistic Regression(Test Size 0.2): 93.75%

	precision	recall	f1-score	support
ckd	0.95	0.96	0.95	55
notckd	0.92	0.88	0.90	25
accuracy			0.94	80
macro avg	0.93	0.92	0.93	80
weighted avg	0.94	0.94	0.94	80

Naive Bayes Classifier

Accuracy of Naive Bayes Classifier(Test Size 0.1): 92.5%

	precision	recall	f1-score	support
ckd	1.00	0.90	0.95	30
notckd	0.77	1.00	0.87	10
accuracy			0.93	40
macro avg	0.88	0.95	0.91	40
weighted avg	0.94	0.93	0.93	40

Decision Tree Classifier

Accuracy of Decision Tree Classifier(Test Size 0.1): 100.0%

	precision	recall	f1-score	support
ckd	1.00	1.00	1.00	30
notckd	1.00	1.00	1.00	10
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40

Random Forest Classifier

Random Forest Classifier

Accuracy of Random Forest Classifier(Test Size 0.1): 100.0%

	precision	recall	f1-score	support
ckd	1.00	1.00	1.00	30
notckd	1.00	1.00	1.00	10
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40

MLP Classifier

Accuracy of MLP Classifier(Test Size 0.1): 82.5%

	precision	recall	f1-score	support
ckd	0.96	0.80	0.87	30
notckd	0.60	0.90	0.72	10
accuracy			0.82	40
macro avg	0.78	0.85	0.80	40
weighted avg	0.87	0.82	0.83	40

In the given Fig.3, by running "python app.py" in anaconda prompt we get the URL for the website.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 882-885 www.ijaem.net ISSN: 2395-5252

Fig. 3.

V. CONCLUSION

The suggested CKD diagnostic methodology is practical in terms of data imputation and sample diagnosis. After supervised imputation of the missing values in the data set using ANN imputation, the integrated model could achieve a sufficient degree of accuracy. Therefore, we believe that adopting this technique to accurately diagnose CKD would be beneficial. Additionally, this concept may be relevant to the clinical data from other disorders utilized in actual medical diagnosis. However, due to the limitations imposed by the requirements, only 400 samples of the available data were employed in the model-building process. As a result, the generalizability of the model can be limited. Additionally, because there are only two kinds of data samples in the data set (CKD and not-CKD), the model is unable to determine the severity of CKD.

VI. FUTURE WORK:

One of the major future enhancements would be for the model to be integrated with python 3 and database. The second major further enhancement would be to move the model of machine learning to the stage of production. To train the model and increase its accuracy the model has to be scaled up. This project can be further developed for a greater number of diseases and provide the gateway to reach out to medical institutions. A user-friendly website can also be developed using cloud space which will benefit more and more farmers for sharing their experiences.

REFERENCE

[1] Q.-L. Zhang and D. Rothenbacher, ``Prevalence of chronic kidney disease in population-based studies: Systematic review," BMC Public Health, vol. 8, no. 1, p. 117, Dec. 2008.

[2] W. M. McClellan, D. G. Warnock, S. Judd, P. Muntner, R. Kewalramani, M. Cushman, L. A. McClure, B. B. Newsome, and G. Howard, ``Albuminuria and racial disparities in the risk for ESRD," J. Amer. Soc. Nephrol., vol. 22, no. 9, pp. 1721_1728, Aug. 2011.

[3] M. K. Haroun, ``Risk factors for chronic kidney disease: A prospective study of 23,534 men and women in Washington County, Maryland," J. Amer. Soc. Nephrol., vol. 14, no. 11, pp. 2934_2941, Nov. 2003.

[4] W. D. Souza, L. C. D. Abreu, L. G. D. SilvaI, and I. M. P. Bezerra, ``Incidence of chronic kidney disease hospitalisations and mortality in Espírito Santo between 1996 to 2017," Wisit Cheungpasitporn, Univ. Mississippi Medical Center, Rochester, MN, USA, Tech. Rep., 2019, doi: 10.1371/journal.pone.0224889.

[5] W. Mula-Abed, K. A. Rasadi, and D. Al-Riyami, ``Estimated glomerular _ltration rate (eGFR): A serum creatinine-based test for the detection of chronic kidney disease and its impact on clinical practice," Oman Med. J., vol. 27, no. 4, pp. 339_340, 2012.

[6] A. S. Levey, D. Cattran, A. Friedman, W. G. Miller, J. Sedor, K. Tuttle, B. Kasiske, and T. Hostetter, ``Proteinuria as a surrogate outcome in CKD: Report of a scienti_c workshop sponsored by the national kidney foundation and the US food and drug administration," Amer. J. Kidney Diseases, vol. 54, no. 2, pp. 205_226, Aug. 2009.

[7] S. Gerogianni, ``Concerns of patients on dialysis: A research study," Health Sci. J., vol. 8, no. 4, pp. 423_437, 2014.

[8] J. R. Chapman, ``What are the key challenges we face in kidney transplantation today?" Transplantation Res., vol. 2, no. S1, pp. 1_7, Nov. 2013.

[9] T. Xiuyi and G. Yuxia, ``Research on application of machine learning in data mining," in Proc. IOP Conf., Mater. Sci. Eng., 2018, doi: 10.1088/1757-899X/392/6/06220.

[10] B. Zupan, A. J. Halter, and M. Bohanec, ``Qualitative model approach to computer assisted reasoning in physiology," in Proc. Intell. Data Anal. Med. Pharmacol. (IDAMAP), Brighton, U.K., 2018, pp. 1_7. ddd