International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.025

Machine Learning Techniques for Real-Time Chronic Kidney Disease Tracking

Dr. P. Senthil Raja
Department of Computer Science and Engineering
K.S.Rangasamy College of Technology
Tiruchengode, India
senthilraja@ksrct.ac.in

Gokulraj P
Department of Computer Science and Engineering
K.S.Rangasamy College of Technology
Tiruchengode, India
gokulrajppm@gmail.com

Sarulatha D
Department of Computer Science and Engineering
K.S.Rangasamy College of Technology
Tiruchengode, India
sarulathadhavamani@gmail.com

Sriram K
Department of Computer Science and Engineering
K.S.Rangasamy College of Technology
Tiruchengode, India
sriramsachin696@gmail.com

***Abstract* – The majority of non-communicable disease-related morbidity and death is caused by chronic kidneydisease (CKD), which affects 10% to 15% of the global population. The impact of patient health issues likehypertension, poor nutritional health, mineral bonedisorder, anaemia, acid base abnormalities, andneurological complications is thought to be significantlyreduced by early and accurate detection of CKD stagesand prompt intervention with appropriate medications. Incertain studies, machine learning approaches have beenutilised to detect CKD early. In this assessment, CKD isidentified using the logistic regression technique. To determine which algorithm offers the most accuracy, many proposed algorithms are compared. AI repository containing a great deal of missing traits.**

***Keywords* – CKD, Machine learning, SVM, RF, ANN, K-NN, Logistic.**

I. Introduction

Their results in CKD diagnosis have been promising. The models indicated above are based on the disease criteria in the data and fill in missing values using the mean imputation technique. Due to the lack of knowledge regarding the diagnostic implications of a sample, their method cannot be applied in this situation. For a variety of reasons, patients may actually skip many assessments before diagnosis. Aside from that, average imputation statistics, which are employed to fill in gaps in categorical data, may drastically deviate from the actual values. Despite the fact that components with just two categories can have their categories set to 0 and 1, the average value of the variables may lieanywhere between 0 and 1. The models suggested, which are based on attribute selection technology, had their computational costs lowered through feature selection.

A. Chronic Kidney Diseases

A gradual deterioration in renal function over months or years is a defining feature of the kidney failure type known as chronic kidney disease (CKD). Initial symptoms are absent, however later warning signs include confusion and fatigue, nausea, and leg edema. Among the complications are heart failure, elevated blood pressure, bone damage, and anaemia. Polycystic kidney disease, glomerulonephritis, diabetes, and high blood pressure all contribute to chronic kidney disease. One of the risks is a history of chronic renal illness. The diagnosis is established through urine testing for albumin concentration and blood tests to measure glomerular filtration rate. The root of the issue could be identified via an ultrasound or a kidney sample. The utilisation of several severity- based staging schemes. At-risk people should be screened, according to the advice. As a first line of treatment, patients may use medications to decrease cholesterol, blood pressure and blood sugar. Angiotensin converting enzyme inhibitors (ACEIs) or angiotensin II receptor antagonists (ARBs) are frequently used as first-line blood pressure drugs since they lowered the risk of heart disease and delayed the course of renal impairment.

B. Machine Learning

The study of changing computer algorithms over time is known as machine learning (ML). It falls within the heading of artificial intelligence. Even when they are not explicitly trained to do so, machine learning algorithms create a model utilising test data, commonly referred to as "training data," to make judgements or predictions. When it would be difficult or impossible to create standard algorithms, machine learning approaches are typically applied. Examples include email classification and computer vision. But not every machine learning technique involves statistical learning. Machine learning does not always rely on statistical learning, though. Computational statistics, a subset of machine learning, is an area of AI technology that concentrates on computer-aided prediction. Mathematical optimization research advances machine learning by providing methodologies, application fields and ideas. Similar research is done in the discipline of data mining, which concentrates on unsupervised learning for exploratory data analysis. The practise of teaching computers to perform tasks without explicit programming is known as machine learning. In order to accomplish particular jobs, computers acquire knowledge from available data.

DOI: rp-9788770040723.025

II. Literature Review

A. ANN and SVM for the prediction of chronic kidney disease: A comparative study.

The authors of this study, Njoud Abdullah Almansour, HajraFahim Syed, and others, set out to use system learning methodologies to detect CKD at an early stage in order to help prevent it (CKD). Artificial Neural Networks and Support Vector Machines are employed as teaching tools in this test. In order to conduct experiments, every missing value from the dataset was changed by recommending the relevant attributes. The Artificial Neural Network and Support Vector Machine algorithms' ideal parameters have been established after significant parameter tuning and a great deal of testing. The two recommended tactics' final versions were built based on the attributes and qualities that were most well-liked.

B. Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning

These studies by Sergio M. Martnez- Monterrubio, Gabriel R. Vasquez-Morales, et al.Following training, In the test statistics set, the model achieves a 95-accuracy rate, enabling its use in disease diagnosis. According to recent research on explainable AI, dual systems should be used, where any other white-field technique that produces results that are broadly consistent with the expected values is added to a black-field system-learning technique. Because to its ability to identify illustrative explanations, case-based reasoning (CBR) has shown to be a very effective supplemental method. It may augment a neural network's prediction with explanations provided by examples. This article puts the NN-CBR dual device that generates CKD prediction logic to the test and evaluates it. This study determined that 3,494,516 Colombians, or 7% of the population, were at risk for developing CKD.

C. A Machine Learning Methodology for Diagnosing Chronic Kidney Disease.

This image reflects the views of ErlendHodneland, EirikKeilegavlen, and others. Renal function gradually deteriorates over time, which is a sign of chronic kidney disease, a dangerous medical condition. In patients with chronic renal disease, the most recent research highlights the effectiveness of photo registration techniques for detecting pathological aberrations. Methods: Nine patients with suspected chronic real illness and ten healthy volunteers underwent dynamic T2 weighted imaging. Results: From biopsy evaluations, it is found that that deformation, normalized extent changes, and strain gradients were all substantially associated with arteriosclerosis. The findings also imply that the sensitivity required to accurately detect minute variations in tissue stiffness is lacking in the photo registration methods now in use. The use of image registration with dynamic time collecting as a tool for invasive measures of arteriosclerosis should be further investigated.

D. Amharic based Knowledge-Based System for Diagnosis and Treatment of Chronic Kidney Disease using Machine Learning.

In their research on developing a self- learning system and experience for identifying and treating the initial three stages of chronic kidney disease, Mohammed and Beshah used machine learning. This study used a tiny amount of data, and they developed a prototype that allows patients to query KBS to learn more about how advice was given. They used a decision tree to generate the rules. It has been stated that the prototype's overall performance is 91% accurate.

III. Methodology

We built a model to predict CKD sickness for the research subjects. Both the model's overall performance and the effectiveness of particular elements were evaluated. Machine learning classifiers like Logistic Regression, Artificial Neural Network SVM, K-Nearest Neighbor and Random Forest were used to train the model. Each classifier's performance and validation matrices were calculated independently. The following stages were included in this study's procedure: (I) preparing the dataset, (II) choosing the features, (III) applying the classifier, (IV) Using SMOTE, and (V) analysing the classifier performance. A deep neural network and machine learning models were used to compare the results of the two different types of neural networks. This was done using an artificial neural network classifier. Statistical testing, namely the McNemar's test, was utilised in this work to determine the significance of two models.

A. Logistic Regression

By employing supervised learning, the classification technique known as logistic regression forecasts the likelihood of a target variable. Due to the target or dependent variable's dichotomous character, there are only two viable classes.

B. SVM

Using a linear model called the Support Vector Machine, or SVM, classification and regression issues can be resolved. It has several advantageous uses and can solve both nonlinear and linear issues. The SVM method splits data into categories by creating a line or a hyperplane.

C. Random Forest

Both classification and regression can be performed using a powerful machine learning technique called Random Forest. Because ensemble learning is the approach employed, many tiny decision trees, known as estimators, which each generate a different set of predictions, make up a random forest model. To produce more accurate predictions, the random forest model combines estimates from the estimators.

D. K-Nearest Neighbor

Algorithm for Simple Machine Learning the Supervised Learning approach is used in K-Nearest Neighbour. This method assumes that occurrences and existing cases will be comparable, and it assigns new cases to the category that are most similar to previous ones.

E. Artificial Neural Network

Computers are taught to understand data similarly to the human brain using neural networks, which is an example of artificial intelligence. Deep learning employs network of

DOI: rp-9788770040723.025

nodes or neurons to replicate the layered structure of the human brain.

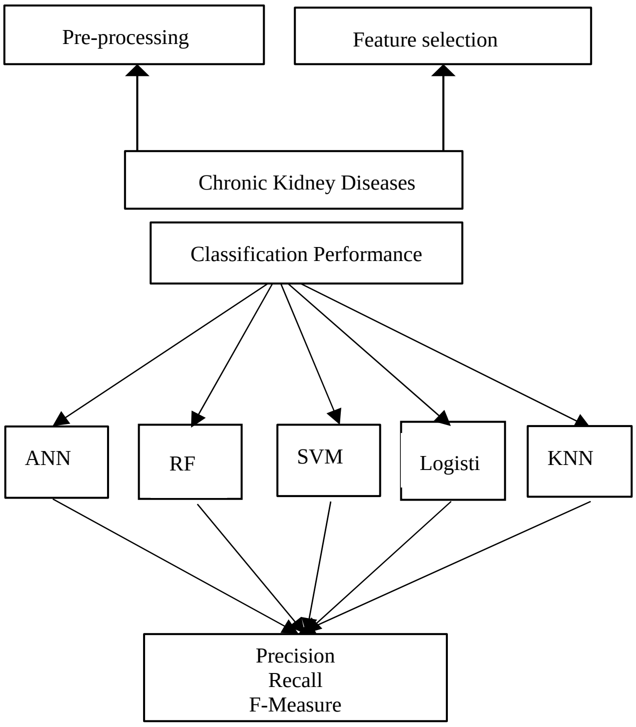

Fig. 1 – Flow Diagram for CKD prediction

IV. Data Description

The dataset from the UCI repository's public chronic kidney disease (CKD) repository was used in this investigation. This dataset includes 400 samples from two distinct classes. 11 of the 25 qualities are numerical, 13 are notional, while one is a category attribute. The dataset includes variables that can be used to anticipate the onset of CKD.

V. Result and Discussion

We feel that this methodology could beused in more complicated circumstances. When analysing more complex data, several algorithms are explored to generate models. Following misjudgement analysis, superior methods that result in diverse errors of judgement are retrieved as component models. Then, by creating an integrated model, the classifier's performance is enhanced.

Because the CKD data set comprises mixed variables, mixed data similarity analysis tools such as the general similarity coefficient could be used to determine sample similarity (numerical and categorical). We utilised EUCLIDEAN distance to determine sample similarity. Therefore, we did not use the techniques for figuring out sample similarity.

VI. Conclusion

The recommended CKD diagnostic methodology is workable with regard to data imputed and sample diagnosis. The model that is integrated was fairly accurate after unsupervised logistic imputation was used to impute empty values from the data set. In this evaluation, to identify CKD, we advocate utilising a Logistic Regression System. As a result, it will be useful to employ this technology in real-world settings to detect CKD. To develop a precise medical diagnosis, this technology might also be integrated with clinical information from other conditions. Unfortunately, the amount of data samples that can be utilised to build the model is rather little, at just 400, due to the restrictions of the parameters.

The generalizability of the model may be constrained as a result. The prototype is also unable to determine the severity of CKD because there are only two different types of data samples included in the data gathering (ckd and notckd). As more compr[illegible]cquired to train the model, its [illegible]e will improve in the future. This will enable the model to determinedisease severity. We believe our model will become better as data size and quality grow.

References

[1] M. M. Hossain et al., “Mechanical anisotropy assessment in kidney cortex using ARFI peak displacement: Preclinical validation and pilot in vivo clinical results in kidney allografts,” IEEE Trans. Ultrason. Ferr., vol. 66, no. 3, pp. 551- 562, Mar. 2019.

[2] Hodneland et al., “In vivo detection of chronic kidney disease using tissue deformation fields from dynamic MR imaging,” IEEE Trans. BioMed. Eng., vol. 66, no. 6, pp. 1779-1790, Jun. 2019.

[3] R. Vasquez-Morales et al., “Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning,” IEEE Access, vol. 7, pp. 152900-152910, Oct. 2019.

[4] N. Almansour et al., “Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study,” Comput. Biol. Med., vol. 109, pp. 101-111, Jun. 2019

[5] G. Chen et al., "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 100497-100508, 2020, doi: 10.1109/ACCESS.2020.2995310.

[6] L. Du et al., “A machine learning based approach to identify protected health information in Chinese clinical text,” Int. J.Med. Inform., vol. 116, pp. 24-32, Aug. 2018.

[7] R. Abbas et al., “Classification of foetal distress and hypoxia using machine learning approaches,” in Proc. Int. Conf. Intelligent Computing, pp. 767-776, Jul. 2018.

[8] L. Antony et al., "A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction," in IEEE Access, vol. 9, pp. 126481-126501,2021,doi: 10.1109/ACCESS.2021.3109168.

[9] Q. Zou et al., “Predicting diabetes mellitus with machine learning techniques,” Front. Genet., vol. 9, Nov. 2018.

[10] Z. Gao et al., “Diagnosis of diabetic retinopathy using deep neural networks,” IEEE Access, vol. 7, pp. 3360-3370, Dec. 2018.

[11] Pazhani. A, A. J., Gunasekaran, P., Shanmuganathan, V., Lim, S., Madasamy, K., Manoharan, R., &Verma, A. (2022). Peer–Peer Communication Using Novel Slice Handover Algorithm for 5G Wireless Networks. Journal of Sensor and Actuator Networks, 11(4), 82.

[12] B. Khan, R. Naseem, F. Muhammad, G. Abbas and S. Kim, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy," in IEEE Access, vol. 8, pp. 55012-55022, 2020, doi: 10.1109/ACCESS.2020.2981689.

[13] G. R. Vásquez-Morales, S. M. Martínez- Monterrubio, P. Moreno-Ger and J. A. Recio- García, "Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning," in IEEE Access, vol. 7, pp. 152900-152910,2019,doi: 10.1109/ACCESS.2019.2948430.

[14] J. Aljaaf et al., “Early prediction of chronic kidney disease using machine learning supported by predictive analytics,” in Proc. IEEE Congr. Evolutionary Computation, Jul. 2018.

[15] J. Parab, M. Sequeira, M. Lanjewar, C. Pinto and G. Naik, "Backpropagation Neural Network-Based Machine Learning Model for Prediction of Blood Urea and Glucose in CKD Patients," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 9, pp. 1-8, 2021, Art no. 4900608, doi: 10.1109/JTEHM.2021.3079714.

[16] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications. Computers and Electrical Engineering, 102, 108130.

[17] H. Polat, H.D. Mehr, A. Cetin, “Diagnosis of chronic kidney disease based on support vector machine by feature selection methods,” J. Med. Syst., vol. 41, no. 4, Apr. 2017.

[18] Papademetriou et al., “Chronic kidney disease, basal insulin glargine,

DOI: rp-9788770040723.025
and health outcomes in people with dysglycemia: The origin study," Am. J. Med., vol. 130, no. 12, Dec. 2017

[19] Y. Chen et al., "Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B," Comput. Biol. Med., vol. 89, pp. 18-23, Oct. 2017

[20] S. Kaur et al., "Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives," in IEEE Access, vol. 8, pp. 228049-228069, 2020, doi: 10.1109/ACCESS.2020.3042273.