



Systematic Survey on Chronic Kidney Disease Prediction Using Different Machine Learning Techniques

Maha Lakshmi Tirumalasetty¹, Ravi Sekhara Reddy Vuppuloori², Balaji Tata¹, Venkata Ganeswara Rao Maddipati³, Jaya Navaneethan⁴, Upendra Chowdary Kurra⁵, Koteswara Rao Kodepogu^{6*}, Lalitha Kumari Gaddala⁶, Surekha Yalamanchil⁶

¹ Department of ECE, P.V.P Siddhartha Institute of Technology, Vijayawada A.P 520007, India

² Department of ECE, Lakireddy Bali Reddy College of Engineering, Mylavaram A.P 521230, India

³ Department of ECE, Shri Vishnu Engineering College for Women, Bhimavaram 534201, India

⁴ Department of EIE, Faculty of Engineering and Technology, Annamalai University, Chidambaram 608002, India

⁵ Dept of ECE, R.V.R & J.C. College of Engineering, Guntur 522019, India

⁶ Department of CSE, PVP Siddhartha Institute of technology, Vijayawada A.P 520007, India

Corresponding Author Email: kkrao@pvpsiddhartha.ac.in

Copyright: ©2023 IETA. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>)

<https://doi.org/10.18280/ria.370629>

ABSTRACT

Received: 31 August 2023

Revised: 3 November 2023

Accepted: 19 November 2023

Available online: 27 December 2023

Keywords:

CKD, machine learning models, predictive analytics

Chronic kidney disease (CKD), a condition characterized by gradual loss of kidney function, poses a global health challenge, affecting a substantial population worldwide. Attributable to renal pathology or sustained kidney damage, CKD serves as a harbinger of morbidity and mortality. The diagnosis of this condition remains a formidable task, fraught with risks, high costs, and extensive duration, often leading to late detection in resource-constrained settings. Recent advancements have introduced an array of improved algorithms that have enhanced the efficiency of risk assessment for CKD. This survey provides a comprehensive examination of various machine learning (ML) algorithms employed in the prognostication of CKD. It is posited that the deployment of machine learning technologies could revolutionize diagnostic methodologies, transitioning to a machine-assisted strategy. The efficacy of diverse algorithms has been systematically evaluated using multiple criteria. Through this analysis, the most effective classifiers for predicting CKD have been identified, with the potential to significantly refine clinical practices. Ultimately, this study delineates the ML algorithms that hold promise for the future of CKD diagnosis and treatment, contributing to the advancement of medical informatics in nephrology.

1. INTRODUCTION

One form of kidney illness that causes a progressive decrease of kidney working wellness is chronic kidney disease (CKD) [1]. Due to various patient living circumstances, this phenomena might be noticed over the course of several months or years. According to current medical data, 10% of the population globally has CKD, which is also known as chronic renal failure. In 2005, there were almost 58 million fatalities globally. Any chronic disease has several phases, and the mortality depends on the stage it reached before being treated. Major risk factors for chronic kidney disease (CKDs) include a high incidence of individuals with diabetes, a history of hyper coronary mellitus, and a family history of renal failure [2]. Undiagnosed and untreated CKD can result in hypertension and, in extreme cases, renal failure if it is not addressed.

A condition known as kidney disease (CKD) occurs when the kidneys stop working properly and are unable to filter blood as well as they should. As a result, the body retains excess fluid and blood waste, which can cause a number of

health problems, including heart disease and stroke. Every one of the two kidneys, which look like beans, is about the size of a clenched hand. On one or the other side of the spine, one is arranged just beneath the rib confine. The kidneys channel somewhere in the range of 120 and 150 pints of blood each day, yielding one to two quarts of pee. The essential occupation of the kidneys is to dispense with squander materials and additional liquid from the body through pee. A number of incredibly intricate excretion and re-absorption mechanisms result in the production of urine. This mechanism is necessary for the body's chemical makeup to remain stable. The salt, potassium, and acid levels in the body are tightly regulated by the kidneys. They also generate hormones that influence the efficiency with which other organs operate. For instance, one chemical created by the kidneys advances the development of red platelets, manages blood pressure, and governs the metabolism of calcium.

1.1 Kidney infection types

Acute Perpetual Renal Failure: Impeded plasma stream to the organs has the potential to cause critical perennial kidney

failure. Without enough haemoglobin mobility, the renal organs are unable to remove waste from the serum.

Acute Intrinsic Kidney Failure: It results from an acute organ damage, such as a natural disaster's aftereffects. **Chronic perpetual renal disease:** If the kidneys receive insufficient blood flow over an extended period of time, they start to constrict and struggle to filter blood.

Chronic Intrinsic Kidney Dysfunction: This disease is brought about by a characteristic kidney condition that damages the kidneys over the long term. The main wound in the kidneys leads to an inherent kidney infection.

Imaging scans, blood testing, clinical data, and eventually a biopsy are often used in the diagnosis of chronic kidney disease (CKD). The run of the mill indicative methodology is a biopsy, yet it has a few disadvantages, including being costly, tedious, nosy, and some of the time deadly. For example, the patient may get an infection, worry about having surgery, or have the wrong diagnosis if a biopsy is done. Sonograms, renal MRIs, and mammograms are among the imaging methods that have traditionally been utilised to identify the condition [3]. Nevertheless, there are a lot of drawbacks to employing them, such as radiation exposure risks. Imaging is dangerous and provides little information for CKD diagnosis.

Computer science's machine learning field allows machines to learn without explicit programming [4]. Artificial intelligence (AI) algorithms create a model that depends on example data in order to generate predictions or conclusions without being explicitly modified to do so. This process is known as "preparing information." Using computational methodologies, AI has demonstrated fruitful in giving answers for beginning phase finding across various clinical disciplines. Decision-makers use these techniques to uncover hidden patterns in data [5]. This study uses the to forecast chronic kidney disease, DT, RF, LR, DL, SVM, CNN, NB, and other ML techniques.

2. LITERATURE SURVEY

Wibawa et al. [6] created a machine-learning algorithm prototype that employed characteristic choice and aggregate training to provide a rapid assessment of chronic renal illness. Correlation-based characteristic selection was used to gather features, and adaptive boosting was used for aggregate training to gain a deeper understanding of the kind of chronic renal disease. Classifiers such the Bayes, SVM, and KNN algorithms were employed.

Using DT classifiers, LR, SVM, and KNN, Charleonnan et al. [7] investigated four machine learning methods to predict CKD. The performance of many models has been evaluated to figure out which classifier is most precise in foreseeing chronic renal illness. The CKD dataset that Apollo Hospitals gathered from Indian patients serves as the foundation for these models. According to their test findings, the SVM classifier had the best accuracy, at 98.3%. After training the dataset, SVM also has the maximum sensitivity.

In their investigation on chronic renal illness, Salekin and Stankovic [8] used classifiers from KNN, RF, and BP to provide an accurate response. By utilizing a component determination technique and a characteristic compression evaluation to identify certain traits, they were able to successfully uncover chronic renal illness. A novel chronic renal illness dataset was presented by Rubini and Eswaran [9] using three classifiers: logistic regression, multilayer

perceptrons, and radial basis function networks. MLP (99.75%), RBF (97.5%), and LR (97.5%) have the highest rates of accuracy.

When using the proton Nuclear Magnetic-Resonance (1H) data to predict grades for chronic kidney illness, Luck et al. [10] made sense of how for utilize an information mining framework that thoroughly searches in the distinctive features. They looked at acquired prediction designs such as a standard model, a prototype incorporating includes both local and global, and a connection between typical univariate statistical feature gathering methods and an L2LR.

2.1 A fresh method for forecasting CKD using ML algorithms

Gudeti et al. [11] states that the CKD dataset from the UCI AI Archive has 25 highlights, of which 11 are mathematical and 14 are notional. The dataset's 400 complete examples are utilized to prepare and display AI calculations. It is necessary to assess, forecast, and treat a specific disease using this data in a useful way. A categorization model delivers a solution based on values that have been determined. There is a propensity for categorization type to assume few or many inputs to predict values of their outcomes. The samples are tested on one group of the dataset, and they are trained on the other group. The proportions of test and training data are 30% and 70%, respectively. To categorise the sickness at puerility, three ML techniques-LR, SVM, and KNN are used. Every algorithm's molarity is examined. Our proposed model, as shown in Figure 1, combines the SVM, LR, and KNN.

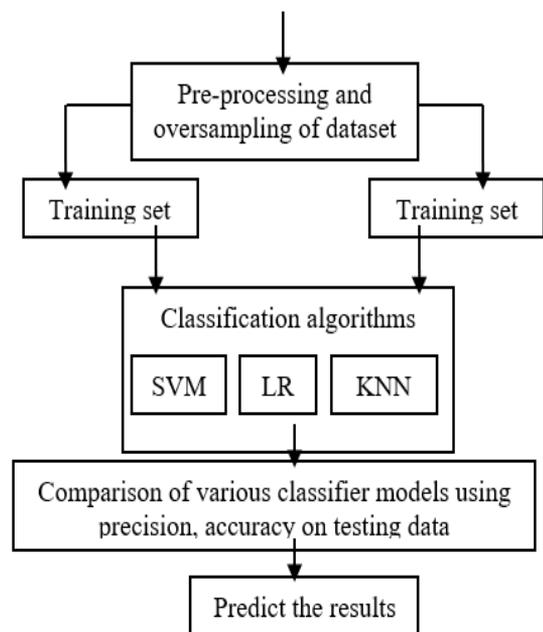


Figure 1. Proposed model using various classification algorithms

2.2 Analysis of the performance of ML classifier for predicting CKD

According to Gupta et al. [12], the research's proposed route is shown in Figure 2. We preprocess the raw dataset after obtaining it from the UCI ML repository in order to remove any irregularities, such as null values. The work of component

choice models affects each model's exhibition. To identify the features that best describe the dataset, feature selection.

techniques including univariate selection and correlation matrices are used. We use the data to shrink the amount of our attributes before testing three different categorization techniques. In the categorising process, classification, we divide a bunch of data into multiple categories. In this case, we classify using a binary system with two classes.

The result of a CKD test would therefore be either positive or negative. These algorithms include Logistic Regression, Random Tree, and Decision Trees. Next, we compare the algorithms using several performance metrics.

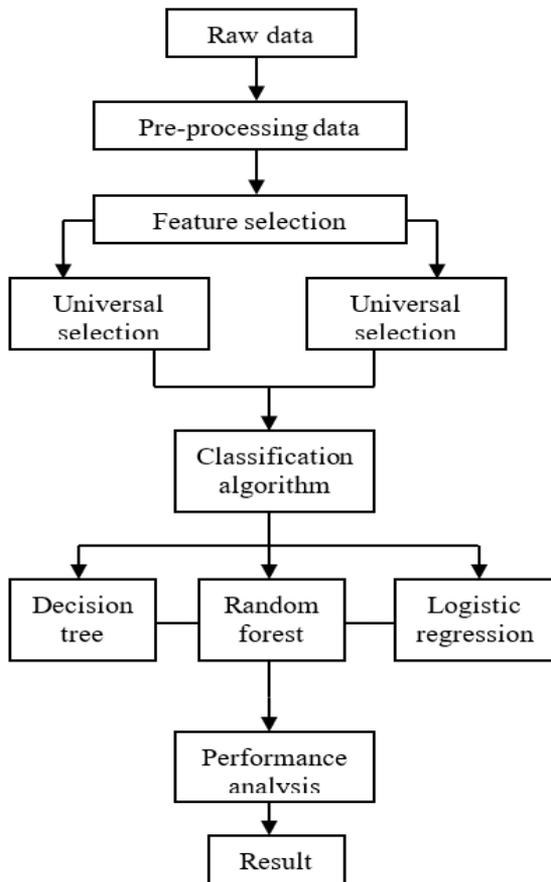


Figure 2. Proposed method

2.3 NN based model for predicting CKD

Suresh et al. [13] it is more crucial to diagnose chronic kidney disease at the right moment in order to halt its progression. An ANN ML model has been constructed as part of the proposed research activity. The ideal machine learning model required a number of stages to create. Removing noisy data and replacing missing values are the two key procedures that make up data pre-processing. Explicit gravity, egg whites, serum creatinine, hemoglobin, stuffed cell volume, and a class name were among the characteristics that could be retrieved from the dataset.

Data for tests and training are kept apart from health records. A column containing a class variable is remembered for the learning period of a wellbeing record, and the algorithm is trained on this data before being applied to all subsequent records. The health records are mixed before learning, and eighty % of them are used for training and twenty % for assessment.

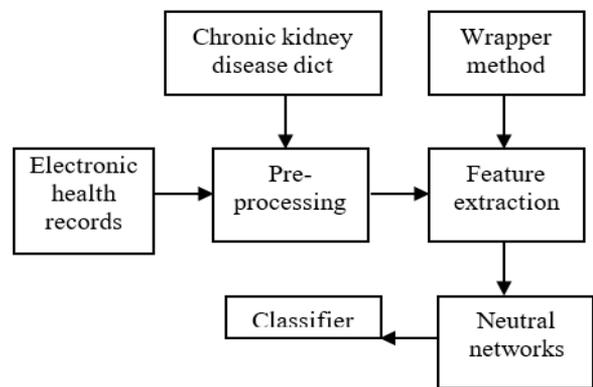


Figure 3. Identification of chronic kidney disease

The suggested model initially adds two layers, the first contains 256 neurons and the "ReLU" enactment capability with an initializer for the loads in light of an ordinary conveyance. Since that layer is the top layer, it is required to provide the quantity of highlights contained in the informational collection. The neuron in the subsequent layer, which is the last layer, will activate using the 'hard sigmoid' function.

Before attempting to increase the technique's effectiveness with an optimizer, build the model and give it the "binary cross entropy," a binary classification loss function that measures how well the strategy performed during training, as seen in Figure 3.

The adjustment is made by measuring the gradient after Function f provides the loss function to optimise, which aids in categorising the chronic renal disease.

2.4 Prediction of CKD using AHDCNN on the IoMT platform

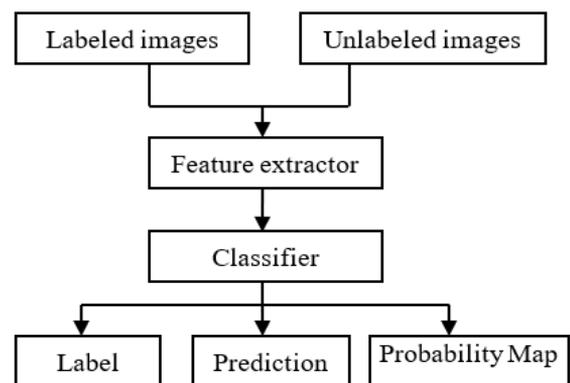


Figure 4. The proposed ahdcnn method

The AHDCNN has been proposed for the prescient analysis of constant renal illness, as per Chen et al. [14]. The datasets used in this work were obtained from the datasets at <http://www.mediafire.com>. Using AHDCNN, the huge scope picture of renal infection has been prepared to accomplish the best thing discovery execution. Three steps make up each convolutional layer: group normalisation, spatial max pooling. Gating for ReLU and linear convolution. For each picture that CNN received as an input, the output from each layer was taken to create the characteristics of the picture hierarchy. In contrast to an FCN, AHDCNN can accurately segment the integrated system using smoothing and prior information. Additionally, as opposed to involving the model as a post-

handling device, the model is consolidated during the preparation stage to change the CNN in a new way. During preparation, it offers the use of unlabelled data in a semi-monitored setting. The suggested AHDNN technique architecture is represented in Figure 4.

2.5 CKD predictive analytics employing machine learning methods

Figure 5 depicts the creation of the prediction models, according to Charleonnann et al. [15]. In this study, four distinct ML calculations are utilized to foresee the gamble of persistent renal sickness. The proposed strategy analyzes order execution of KNN, SVM, DT, and LR. The Indians CKD dataset is utilised for the proposed model of chronic kidney disease prediction.

We initially translate the nominal features of the training data into binary attributes. To eliminate how much characteristics and training time, the second phase includes selecting a subset of features using the Best First features selection technique.

To explore the space of feature subsets, The Best Initially utilizes voracious slope climbing and a backtracking calculation. The classifier model is prepared to create a prescient model that forecasts future data in the third step. The testing results from the last phase are used to forecast the chronic renal class.

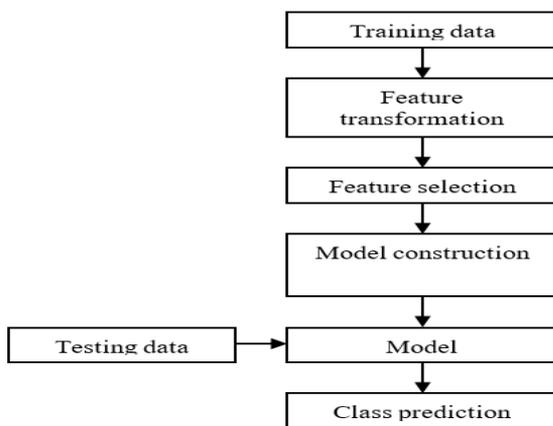


Figure 5. Creating the model to predict the case of CKD

3. RESULT ANALYSIS

3.1 An innovative method for CKD prediction using ML algorithms

The UCI AI Store's CKD dataset, which has 25 qualities — 11 of which are mathematical and 14 of which are ostensible is used in this work. A number of assessment criteria were used to each classifier's results, and 10-fold cross-validation was employed to look for overfitting. One technique that has proven effective in improving the model's parameters is nested crossvalidation. Python 3.3 will be used to administer the tests via the Jupyter Notebook web application. Sciket-learning is a Python framework for open machine learning systems, and some of its modules were used. This inquiry considers AUC, sensitivity, specificity, accuracy using F1-measurement, and other assessment criteria. A unique set of outcomes is produced by each model, contingent upon the values of its

parameters. Our results show that SVM predicts CKD more correctly (99.25%) than LR and KNN within the limited constraints of this medical scenario. The detailed CKD prediction with respect to the algorithm can be seen in Table 1.

Table 1. CKD prediction using ML

Algorithm	Accuracy
SVM	99.25
Logistic Regression	92.10
K-Nearest Neighbors	90.10

3.2 Performance analysis of ML classifier for predicting CKD

We utilized a dataset that was accessible from UCI's AI library. Records for 400 patients are included in this collection. Our algorithms, which comprise DT, LR, and RF, will be evaluated using three different evaluation metrics: accuracy, precision, and recall (RF). TP in these functions refers to instances where the algorithms correctly anticipated a positive outcome. As negative cases were expected, TN. FP stands for anticipated positive but actual negative instances. FN represents cases that were anticipated to be negative yet turned out to be positive [12].

The accuracy of the suggested algorithm-DT, RF, and LR- was 97.48 percent, 95.16 percent, and 98.24 percent, respectively. Recall of 96.61, 97.29, and 100, and precision of 100, 96.12, and 97.82. Utilizing the advantages of each feature selection approach, two feature selection strategies are merged. LR has the best accuracy and recall when compared, nevertheless. The detailed ML classifier results can be seen in Table 2.

Table 2. ML classifier for predicting CKD

Algorithm	Accuracy	Precision	Recall
DT	97.48	100	96.61
RF	95.16	96.12	97.29
LR	98.24	97.82	100

3.3 NN based model for predicting CKD

A disarray framework is utilized to survey the profound brain organization's presentation on an assortment of test records. The qualities in the disarray grid are right. The numbers are: True Negative (+)=31, False Positive (-)=0, False Positive (+)=1, and False Negative (-)=2. Table 3 is an illustration of a confusion matrix. Our results showed that the Neutral network algorithm predicted chronic kidney disease with 96% accuracy.

Table 3. Confusion matrix

Test Result	Chronic Kidney Disease	
	Predicted no	Predicted yes
Actual no	31	0
Actual yes	2	25

3.4 Prediction of CKD using AHDNN on the IOMT

The analytical approach for diabetes and renal disease in healthcare is presented in this research. The expected outcomes show that incorporating a characterization model into the recommended framework to perceive potential

dangers during the primary phases of treating specific diseases is doable. One example of a Web of Things arrangement in the medical care industry is the observing of imperative signs by distant sensors and the constant and extremely durable data move to the pertinent master through cloud administrations. The Web of Things depends on distributed computing and fills in as an organization for the get-together of information from sensors instead of the designed machine-to-machine (M2M) framework with its few link associations, which occupy valuable room.

The stacked layers must fit directly into the residual maps for the residual network to function, not into the target frame. The experimental results might lead to far more accurate modelling and scaling up of residual networks. The correct categorization of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) is established by accuracy. The suggested AHDCNN's F1-score and accuracy are both 97.3. The Dice Index has compared the automatic segmentation results, including the detection step, with the actual data. In 80% of the kidneys, the proper identification and segmentation have been carried out (Dice>0.90). The method failed in just 6% of cases (Dice 0.65). The entire installation process takes about 10 seconds. At every categorization level, the ROC curve (ROC-curve) illustrates the efficacy of the proposed model. The ratio of true positives to false positives (TPR vs. FPR) is shown in this graph [15].

3.5 Predictive analytics for CKD using ML

The Indians Persistent Kidney Sickness (CKD) dataset is utilized in this CKD forecast examination. This experiment uses four distinct categorization techniques: decision trees, LR, SVM with Gaussian kernels, and KNN. The Weka data mining tool and Matlab are used to build the experiments. Sensitivity, specificity, and accuracy are used to gauge how well the suggested approach performs. In the experiments, the outcomes of four machine learning approaches are contrasted. Every machine learning algorithm is trained and assessed using the recommended approach. The 5-crease cross approval approach is utilized in this analysis to prepare and assess the AI models. According to the experimental results, the SVM classifier has an accuracy rate of 98.3%, which is greater than the average accuracy rates of 96.55%, 94.8%, and 98.1% for the other classifiers.

The survey mentioned above revealed the following findings, which are shown in Table 4.

Table 4. Comparison of different methods based prediction of chronic kidney disease

S. No.	Used Methods	Best Results at	Dataset	Obtained Accuracy (%)
1	KNN, LR, SVM	SVM	UCI ML repository dataset	99.25
2	DT, RF, LR	LR	UCI ML repository dataset	99.24
3	NN	NN	Health records	96
4	AHDCNN	AHDCNN	Collected from sensors on IMoT	97.3
5	SVM	SVM	CKD dataset	98.3

4. CONCLUSION

Improved diagnosis is necessary for some health issues,

such as chronic kidney disease. Many ML techniques for CKD prediction are reviewed in this publication. This exploration study utilizes an assortment of ML calculations, such as AHDCNN, KNN, SVM, DT, NN, RF, and LR. The models built using CKD patient data are trained and certified using the input parameters mentioned above. This survey has the advantage that the prediction approach is significantly faster, enabling physicians to identify a bigger patient group more rapidly and treat CKD patients as soon as possible..

REFERENCES

- [1] Arafat, F., Khan, T., Bapon, A.D., Khan, M.I., Noori, S.R.H. (2021). A deep learning approach to predict chronic kidney disease in human. In 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, pp. 1010-1015. <https://doi.org/10.1109/IEMCON53756.2021.9623101>
- [2] Akter, S., Habib, A., Islam, M.A., Hossen, M.S., Fahim, W.A., Sarkar, P.R., Ahmed, M. (2021). Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease. *IEEE Access*, 9: 165184-165206. <https://doi.org/10.1109/ACCESS.2021.3129491>
- [3] El-Melegy, M., Abd El-karim, R., El-Baz, A., Abou El-Ghar, M. (2018). Fuzzy membership-driven level set for automatic kidney segmentation from dce-mri. In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Rio de Janeiro, Brazil, pp. 1-8. <https://doi.org/10.1109/FUZZ-IEEE.2018.8491552>
- [4] Amirgaliyev, Y., Shamiluulu, S., Serek, A. (2018). Analysis of chronic kidney disease dataset by applying machine learning methods. In 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, pp. 1-4. <https://doi.org/10.1109/ICAICT.2018.8747140>
- [5] Aljaaf, A.J., Al-Jumeily, D., Haglan, H.M., Alloghani, M., Baker, T., Hussain, A.J., Mustafina, J. (2018). Early prediction of chronic kidney disease using machine learning supported by predictive analytics. In 2018 IEEE Congress on Evolutionary Computation (CEC), Janeiro, Brazil, pp. 1-9. <https://doi.org/10.1109/CEC.2018.8477876>
- [6] Wibawa, M.S., Maysanjaya, I.M.D., Putra, I.M.A.W. (2017). Boosted classifier and features selection for enhancing chronic kidney disease diagnose. In 2017 5th International Conference on Cyber and IT Service Management (CITSM), Bali, Indonesia, pp. 1-6. <https://doi.org/10.1109/CITSM.2017.8089245>
- [7] Charleonnann, A., Fufaung, T., Niyomwong, T., Chokchueypattanakit, W., Suwannawach, S., Ninchawee, N. (2016). Predictive analytics for chronic kidney disease using machine learning techniques. In 2016 Management and Innovation Technology International Conference (MITicon), Bangkok, Thailand, pp. MIT-80. <https://doi.org/10.1109/MITICON.2016.8025242>
- [8] Salekin, A., Stankovic, J. (2016). Detection of chronic kidney disease and selecting important predictive attributes. In 2016 IEEE International Conference on Healthcare Informatics (ICHI), Chicago, IL, USA, pp. 262-270. <https://doi.org/10.1109/ICHI.2016.36>

- [9] Rubini, L.J., Eswaran, P. (2015). Generating comparative analysis of early stage prediction of chronic kidney disease. *International Journal of Modern Engineering Research (IJMER)*, 5(7): 49-55.
- [10] Luck, M.M., Yartseva, A., Bertho, G., Thervet, E., Beaune, P., Pallet, N., Damon, C. (2015). Metabolic profiling of ¹H NMR spectra in chronic kidney disease with local predictive modeling. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, pp. 176-181. <https://doi.org/10.1109/ICMLA.2015.155>
- [11] Gudeti, B., Mishra, S., Malik, S., Fernandez, T.F., Tyagi, A.K., Kumari, S. (2020). A novel approach to predict chronic kidney disease using machine learning algorithms. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, pp. 1630-1635. <https://doi.org/10.1109/ICECA49313.2020.9297392>
- [12] Gupta, R., Koli, N., Mahor, N., Tejashri, N. (2020). Performance analysis of machine learning classifier for predicting chronic kidney disease. In 2020 International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-4. <https://doi.org/10.1109/INCET49848.2020.9154147>
- [13] Suresh, C., Pani, B.C., Swatisri, C., Priya, R., Rohith, R. (2020). A neural network-based model for predicting chronic kidney diseases. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 157-162. <https://doi.org/10.1109/ICIRCA48905.2020.9183318>
- [14] Chen, G., Ding, C., Li, Y., Hu, X., Li, X., Ren, L., Ding, X., Tian, P., Xue, W. (2020). Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform. *IEEE Access*, 8: 100497-100508. <https://doi.org/10.1109/ACCESS.2020.2995310>
- [15] Debal, D.A., Sitote, T.M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*. <https://doi.org/10.1186/s40537-022-00657-5>