Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7922964

EFFECTIVE DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING TECHNIQUES

Mr. Anandkumar A. Sutariya[1] and Dr. Dushyantsinh B. Rathod[2]

[1] PhD Scholar, Computer Engineering, Sankalchand Patel College of Engineering, Sankalchand Patel University, Visnagar, Gujarat, India
asutariya82@gmail.com
[2] Professor & HoD, Computer Engineering, Ahmedabad Institute of Technology, Ahmedabad, Gujarat, India
dushyantsinh.rathod@gmail.com

Abstract. Because there are no symptoms, it might be challenging to detect chronic kidney disease (CKD) in its early stages. The development and validation of a predictive model for the prognosis of chronic renal disease is the aim of the proposed study. In order to diagnose and categorize diseases, machine learning algorithms are frequently utilized in medicine. Medical records are frequently inaccurate. Using a dataset on chronic kidney disease from the UCI Machine Learning Repository, we applied four machine learning classifiers for analysis: Logistic Regression (LR), Decision Tree (DT), Histogram Boosting Gradient (SHGB), and Support Vector Machine (SVM), using a total of 25 features. The machine learning classifiers were trained using the clusters of the dataset for chronic renal disease. The Kidney Disease Collection is then compiled using non-linear features and categories. The SHGB produces the best results, with an accuracy of 91%.

Keywords: *Chronic Kidney Disease, Support Vector Machine, Histogram boosting Gradient, Logistic Regression and Decision Tree*

1 Introduction

Researchers in engineering and medicine are working to create machine learning models and algorithms that can detect chronic kidney disease at an early stage. The issue is that the size and complexity of the data produced by the healthcare sector make data analysis challenging. However, by applying data mining technologies, we can transform this data into a format that can then be used by machine learning algorithms.
The severity of kidney disease can be determined using a combination of the estimated glomerular filtration rate (GFR), age, diet, pre-existing medical conditions, and albuminuria, but more precise knowledge of the risk to the kidney is needed for making clinical selections about diagnosis, treatment, and referral [1].

The development and validation of predictive models for chronic renal disease is the goal of this model. The major objective is to determine whether renal failure necessitates kidney dialysis or a kidney transplant in the first place [2].

These models also instruct the patient on how to lead a healthy lifestyle and assist the doctor in determining the likelihood and seriousness of the condition as well as the best course of action for the treatment going forward. Using ANN and mining techniques, it may be feasible to spot trends in data collection, and the incidence of specific diseases that could be harmful in the future may be anticipated [3].

The suggested model's goal is to forecast whether the patient will experience or grow chronic kidney failure in the future if their current lifestyle is maintained. The doctor can use this information to establish whether the kidney disease is using eGFR (glomerular filtration rate), which aids in therapy planning. The estimated glomerular filtration rate (eGFR) gauges kidney health and determines the severity of renal disease [4].

The kidney's primary job is to filter the body's blood. Renal disease is a silent killer because renal failure can occur without any warning signs or symptoms. The definition of chronic renal disease is a deterioration in kidney function over months or years. Diabetes and high blood pressure are common contributors to kidney damage. Globally, chronic kidney disease is a serious health issue that affects many people. People who can't afford therapy for chronic renal disease may suffer catastrophic effects if they don't receive it. The most accurate test to assess kidney function and the severity of chronic kidney disease is the glomerular filtration rate (GFR). It can be calculated using the blood creatinine level, as well as other factors including age, gender, and other details. Most of the time, becoming sick sooner is preferable. As a result, significant sickness can be avoided [5].

In particular, models that are suggested to predict renal disease involve data mining techniques. More data beyond the current model for chronic renal disease can be added to the database. That is, the prediction's accuracy can be improved by adding more data collected from people with chronic renal disease (albeit the data must be accurate). Additionally, with the aid of specialists, research can be conducted to discover additional features that contribute to chronic kidney disease, and these features can then be added as attributes to the fabric paper to improve the prediction's accuracy. The following is the design process: The data are first categorized as CKD or NOT_CKD using the best algorithm, and then the classification is finalized. If the categorization is chronic renal disease, the eGFR value will be calculated using the CKD_EPI equation. With this eGFR measurement, we can determine the patient's current condition.

2 Literature Survey

Kidney disease is also known as nephropathy or kidney damage. People who have renal disease experience kidney failure, which, if untreated, can result in kidney failure. According to the National Renal Foundation, 10% of the world's population has chronic renal disease, and due to insufficient medical care, millions of people pass away every year. Countries that are unable to manage renal disease testing may find promise in recent developments in ML and DL-based testing for kidney disease.

Using classifier techniques including Gaussian NB, Bernoulli NB, and Random Forest, Bemando et al. studied the association between blood-related disorders and their characteristics. These three algorithms provide statistical findings in a variety of ways by anticipating them. In this experiment, we found that the estimated accuracy of the Nave Bayes algorithm was higher than that of other methods [6].

In the sphere of medicine, Kumar and Polepaka developed a method for sickness forecasting. They used CNN, Random Forest, and other machine learning techniques. These algorithms produce higher classification, recall, precision, and F1-score results for sickness datasets. In this experiment, Random Forest performed statistically and accurately better than other algorithms [7].

A method for forecasting medically connected illnesses was developed by Sing et al. They used a support vector machine classifier for better prediction. The accuracy was between 73 and 91 percent; eventually, the author increased accuracy to 91 percent [8].

In the field of medicine, Desai et al. developed a method for sickness prediction. In this study, the author used back-propagation NN and LR classification methods. These two approaches produce different results, with logistic regression and statistical analysis producing a more precise model than previous techniques [9].

A database for medical illnesses linked to ECG arrhythmias was made by Patil et al. The authors used Cuckoo search-optimized neural networks and support vector machines on a dataset of diseases, and support vector machine estimated 94.44 percent improved accuracy [10].

Liu et al. used a dataset of observed illnesses for statistical analysis. Using machine learning techniques like support vector machines, they predicted superior results for specificity, sensitivity, positive value for prediction, and negative predictive value [11].

Acharya et al. reviewed the medical related illness dataset in order to improve statistical analysis results. They applied algorithms based on machine learning to the ECG dataset and used a variety of machine learning techniques, including CNN, to achieve a classification accuracy of 94% [12].

An approach to statistical analysis was developed by Wasle et al.

The authors used a range of machine learning techniques to analyze the dataset for chronic kidney disease. To improve prediction, they applied Nave Bayes, Decision Trees, and Random Forest, and they found that Random Forest computed classification accuracy that was higher than the other algorithms [13].

Nithya et al. created an approach for categorizing and cluster-based analysis using the dataset for kidney illness. The authors applied the K-Means clustering method to various collections of photos to gather the images that were most similar to one another. The classification accuracy was calculated at 99.61 percent utilizing Artificial Neural Networks for Kidney Disease Image Prediction [14].

In order to analyze datasets for chronic kidney disease, Al Imran et al. looked at the usage of machine learning techniques. The authors used Logistic Regression and Feed Forward Neural Network to get superior outcomes for statistical evaluations such as F1-score, Precision, Recall, and AUC than previous techniques [15].

Using a dataset, Navaneeth and Suchetha developed a technique for forecasting chronic renal illness. They used machine learning techniques like SVM and CNN.

Greater accuracy, sensitivity, and specificity findings were predicted by the authors [16].

The dataset for chronic renal disease was used by Brunetti et al. For the illness dataset, authors employed CNN machine learning and calculated 95% classification accuracy [17].

3 Methodology

The dataset for chronic renal illness was obtained from the UCI machine learning repository, and three machine learning classifiers—logistic regression, decision trees, and support vector machines—were applied. 80% of the training data and 20% of the testing data were used to train and test the generated model. The approach of tenfold cross validation is used to train the classifiers. We employed the bagging ensemble approach to enhance the performance of the created model after analyzing all three machine learning classifiers, and the final outcomes were assessed. Fig. 1 provided a description of the suggested methodology.

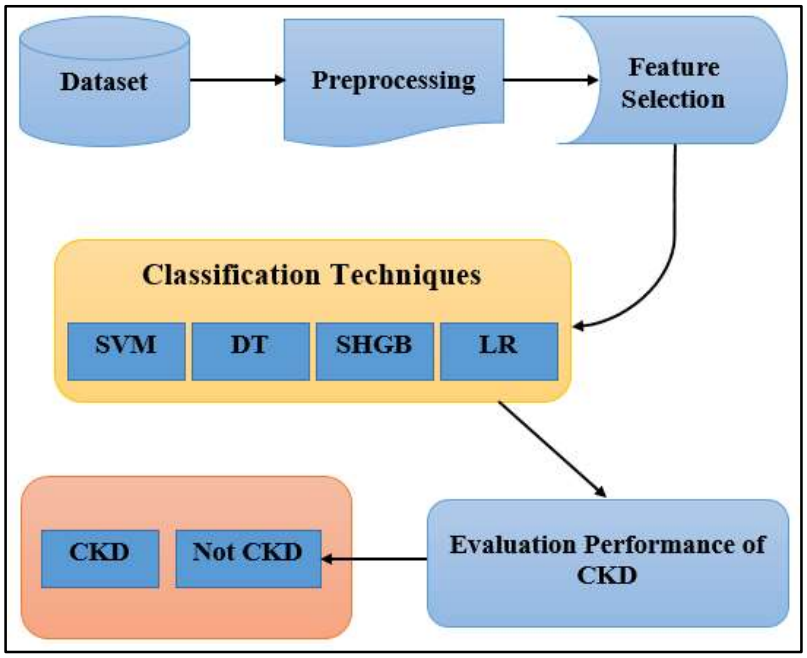

Fig. 1. Proposed model

Data Description

The UCI machine learning repository provided the database for this work. The dataset contains 400 instances (250 CKD and 150 NOT_CKD). There are 14 category attributes and 11 non-categorical attributes listed in Table 1 [18]. Table 1 displays the categorical and non-categorical characteristics of chronic kidney disease with null counting values and data types as attribute values. There are 400 occurrences and 25 (both category and non_categorical) attributes in total for chronic renal disease. From the UCI machine learning repository, information on chronic kidney disease is collected in the form of an electronic medical record. The target variable has the values "1" and "0". The "1" represents normal circumstances, while the "0" represents illness [19].

Table 2. CKD dataset with categorical and non-categorical attributes

Non_categorical attributes				Categorical attributes			
#	Column	Non-null count	Dtype	#	Column	Non-null count	Dtype
0	Id	400 non-null	int64	0	bp	388 non-null	float64
1	Age	391 non-null	float64	1	Sg	353 non-null	float64
2	bgr	356 non-null	float64	2	Al	354 non-null	float64
3	Bu	381 non-null	float64	3	su	351 non-null	float64
4	Sc	383 non-null	float64	4	Rbc	248 non-null	object
5	Sod	313 non-null	float64	5	Pc	335 non-null	object
6	Pot	312 non-null	Float64	6	Pcc	396 non-null	object
7	Hemo	348 non-null	float64	7	Ba	396 non-null	object
8	Pcv	330 non-null	object	8	Htn	398 non-null	object
9	Wc	295 non-null	object	9	Dm	398 non-null	object
10	rc	270 non-null	object	10	Cad	398 non-null	object
11	classification	400 non-null	object	11	Appet	399 non-null	object
dtype: float64(7), int64(1), object(4)				12	pe	399 non-null	object
				13	Ane	399 non-null	object
				14	classification	400 non-null	object
				dtype: float64(4), object(11)			

Machine Learning Classifiers

Logistic Regression Classifier

Pierre François Verhulst created the logistic function, which was formerly known as the "Logistic," as a model of population expansion in the decades between the 1830s and 1840s. On the creation of this function, numerous researchers afterwards worked. Cox and Theil published the first versions of the multinomial "logit" model in 1966 and 1969. The multinomial function logit was associated with discrete choice theory by Daniel McFadden, who demonstrated how it originated from the presumption of relative preferences for irrelevant options being independent. The novel logistic regression idea was theoretically established by this. In many fields, including machine learning, medicine, and social sciences, logistic regression is helpful. A system or model's success or failure can be predicted using this technique, which is very useful in engineering [20]. Here is a quick explanation of logistic regression:

A linear regression predicts values outside the range of (0 to 1), but a logistic regression only predicts the likelihood in two values.

The mathematical formulation of logistic regression is shown below:

$$p = \frac{1}{1 + e^{-(+b_1x_1+b_2x_2+\ldots+b_px_p)}}$$

Decision Tree Classifier

J. R. Quinlan of the University of Sydney developed the decision tree, a supervised learning-based predictive modelling tool, and he wrote about it in his book Machine Learning. This tool utilizes the multivariate analysis method, which can be used to predict, clarify, characterize, and categorize the outcome.

The dataset is divided based on various circumstances, describing cases with more than one cause and explaining the condition in light of many impacts. The decision-tree-generating Iterative Dichotomize version 3 (ID3) algorithms were developed by Quinlan. On the basis of ID3, he then broadened his research and developed C4.5, an enhancement of ID3. C5.0 under GPL [21] is a feature-rich and improved version of C4.5, which is offered for sale by Quinlan. Following a top-down method that incorporates data segmentation, a decision tree is produced from the root. The formula shown below is used to calculate entropy and the Gini index.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

$$Entropy = \sum_{i=1}^{C} -p_i log_2(p_i)$$

For the creation of decision trees, numerous algorithms are utilized.

1. Classification and Regression Tree (CART)
2. ID3
3. CHAID
4. ID4.5

Support Vector Machine Classifier

Support Vector Machine Regression and classification are two uses for the supervised learning tool known as a classifier. The main principle of how SVM functions is that it is a binary classification method that divides the data points to discover a hyperplane in the case of numerous alternative inputs. In the case of high dimensional spaces, it functions well and effectively handles the outliers. Classification is carried out using decision functions, commonly referred to as support vectors. For classification, at least four different types of kernels are employed: linear SVC, linear SVC with RBF kernel, and polynomial SVC [22]. The classifier algorithm is mathematically represented as follows.

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0, 1 - y_i\left(w \cdot x_i - b\right)\right)\right] + \lambda||w||^2$$

4 Results and Discussion

The logistic regression classifier is utilized initially, followed by the decision tree classifier, and finally the support vector machine classifier in the current paper's use of machine learning classifiers. The technique section above describes each of these classifiers in detail.

A confusion matrix is a table-like structure used to explain or assess a classifier's performance. On a dataset for which the true values are already known, this performance description is carried out. Despite the fact that the concepts in the matrix may seem perplexing, the confusion matrix is typically regarded to be straightforward and simple to comprehend [25]. The confusion matrix's terms are defined below.

- True Negative (TN): the case unendingly was projected to be negative.
- True Positive (TP): the case was positive and should be positive.
- FN (False Negative): the case was positive, but the outcome was projected to be negative.

- False Positive (FP): the case was negative, at this point, it was expected to be positive.

Many different constituent variable parameters are needed to build a classification report; these parameters are used to display the values of the parameters used to calculate an accuracy score. Using the formula shown in the table below, recall value—also known as hit rate, True Positive Rate (TPR), or sensitivity—can be calculated. Selectivity, also known as specificity or True Negative Rate (TNR), is calculated using the formula shown in the table below. Precision, also known as positive predictive value (PPV), is calculated using the formula shown in the table below. F1-score, also known as the balance f-score or classic f-measure, is actually the harmonic mean of sensitivity and precision [32–34]. Using the formula shown in Table 2 below, precision is computed. Table 3 provides a full summary of the outcomes from the base classifiers.

Table 2. Table of the Confusion Matrix metric formulas

Type of Metric	Formula
Accuracy	$ACC = \frac{tp+tn}{tp+fp+tn+fn}$
Recall	$Recall = \frac{tp}{tp+fn}$
Precision	$Precision = \frac{tp}{tp+fp}$
F1-score	$F = 2.\frac{precision.recall}{precision+recall}$
Specificity	$Specificity = \frac{tn}{tn+fp}$

The confusion matrix reveals not just a predictive model's performance, but also which classes are successfully predicted, which are incorrectly forecasted, and what types of errors are being made.

```
from sklearn.metrics import classification_report
SHGB_Pred=SHGBmodel.predict(x_test)
SHGBreport = classification_report(y_test, SHGB_Pred)
print(SHGBreport)
```

```
              precision    recall  f1-score   support

           0       0.93      0.88      0.90        48
           1       0.88      0.94      0.91        47

    accuracy                           0.91        95
   macro avg       0.91      0.91      0.91        95
weighted avg       0.91      0.91      0.91        95
```

Fig. 2. Classification report from experiment with accuracy for SHGB

At the point when we utilize the term precision, we typically suggest exactness. The quantity of right expectations partitioned by the all-out number of information tests is the proportion. We considered the precision of each of the four calculations in this estimation are displayed in Fig. 3.

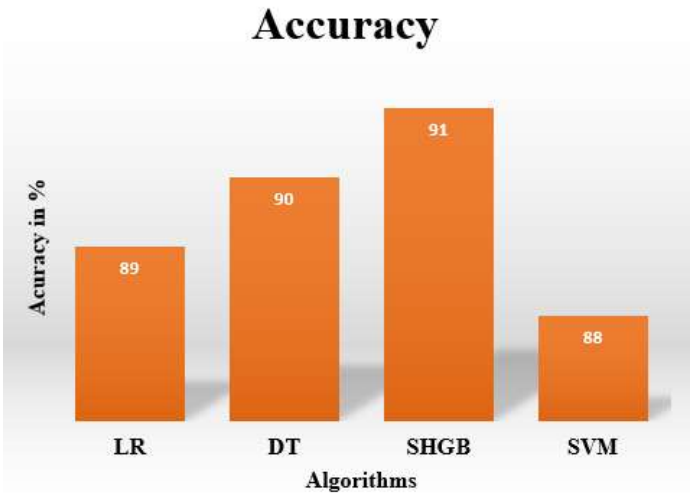

Fig. 3. Accuracy Chart of Machine learning algorithms

In the wake of ascertaining the presentation of proposed models and looking at them all, the best classifier to anticipate Chronic Disease was picked. As per the exploratory information, the SHGB strategy has the greatest exactness of 95%, contrasted with 92%, 91% and 94 percent for the DT, LR and SVM calculations, individually. The outcomes are displayed in Fig. 4.

Algorithms	Accuracy in %
LR	89
DT	90
SHGB	91
SVM	88

Fig. 4. Accuracy Table of Machine learning algorithms

5 Conclusion

We used data on chronic renal disease from the UCI machine learning repository for this study. To evaluate the effectiveness of the prediction model, we created a set of three machine learning classifiers: Logistic Regression, Decision Tree, Histogram Boosting Gradient (SHGB) and Support Vector Machine. The sensitivity, precision, recall, f1-score, support, confusion matrix, and other performance matrices are some of the factors that affect how well a model

performs. Categorical and non-categorical variables from the chronic kidney disease dataset were used to train the created chronic kidney disease prediction model. Following the application of the base classifiers, we discover that the Histogram Boosting Gradient (SHGB) classifier produced higher results in terms of Accuracy, Precision, Recall, and F-score, with values of 91%, 0.93, 0.94, and 0.91, respectively. In comparison to logistic regression, decision tree and support vector machines, the Histogram Boosting Gradient (SHGB) classifier performs better. This can aid both patients and medical professionals in the early detection of chronic kidney disease, potentially saving lives. In the future, the model can be improved by using feature selection techniques to boost forecast accuracy.

References

1. Aljaaf, A.J. 2018 Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC). Wellington. New Zealand
2. A. Nishanth, T. Thiruvaran, Identifying important attributes for early detection of chronic kidney disease. IEEE Rev. Biomed. Eng. 11, 208–216 (2018)
3. A. Ogunleye, Q.-G. Wang, XGBoost model for chronic kidney disease diagnosis. IEEE/ACM Trans. Comput. Biol. Bioinform. 17, 2131–2140 (2020)
4. F. Aqlan, R. Markle, A. Shamsan, "Data mining for chronic kidney disease prediction." in IIE Annual Conference. Proceedings, Institute of Industrial and Systems Engineers, (IISE 2017), pp. 1789–1794
5. N. Borisagar, D. Barad, P. Raval, Chronic kidney disease prediction using back propagation neural network algorithm. Proce. Int. Confe. Commun. Netw. 19–20, 295–303 (2017)
6. C. Bemando, E. Miranda, M. Aryuni, "Machine-Learning-Based Prediction Models of Coronary Heart Disease Using Naïve Bayes and Random Forest Algorithms," in 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM), (IEEE, 2021), pp. 232–237
7. R.P. Ram Kumar, SanjeevaPolepaka, Performance comparison of random forest classifier and convolution neural network in predicting heart diseases, in Proceedings of the Third International Conference on Computational Intelligence and Informatics. ed. by K. SrujanRaju, A. Govardhan, B. PadmajaRani, R. Sridevi, M. Ramakrishna Murty (Springer, Singapore, 2020)
8. H. Singh, N. V. Navaneeth, G. N. Pillai, "Multisurface proximal SVM based decision trees for heart disease classification," in TENCON 2019-2019 IEEE Region 10 Conference (TENCON), (IEEE 2019), pp. 13–18
9. S.D. Desai, S. Giraddi, P. Narayankar, N.R. Pudakalakatti, S. Sulegaon, Backpropagation neural network versus logistic regression in heart disease classification in advanced computing and communication technologies (Springer, Singapore, 2019)
10. D.D. Patil, R.P. Singh, V.M. Thakare, A.K. Gulve, Analysis of ecg arrhythmia for heart disease detection using svm and cuckoo search optimized neural network. Int. J. Eng. Technol. 7(217), 27–33 (2018)

11. N. Liu, Z. Lin, J. Cao, Z. Koh, T. Zhang, G.-B. Huang, W. Ser, M.E.H. Ong, An intelligent scoring system and its application to cardiac arrest prediction. IEEE Trans. Inf Technol. Biomed. 16(6), 1324–1331 (2012)
12. U. Rajendra Acharya, Oh. Shu Lih, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R.S. Tan, A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. 89, 389–396 (2017)
13. R.S. Walse, G.D. Kurundkar, S.D. Khamitkar, A.A. Muley, P.U. Bhalchandra, S.N. Lokhande, Effective use of naïve bayes, decision tree, and random forest techniques for analysis of chronic kidney disease, in International Conference on Information and Communication Technology for Intelligent Systems. ed. by T. Senjyu, P.N. Mahalle, T. Perumal, A. Joshi (Springer, Singpore, 2020)
14. A. Nithya, A. Appathurai, N. Venkatadri, D.R. Ramji, C.A. Palagan, Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images. Measurement (2020). https:// doi. org/ 10. 1016/j. measu rement. 2019. 106952
15. Abdullah Al Imran, Md Nur Amin, and Fatema Tuj Johora. Classification of chronic kidney disease using logistic regression, feedforward neural network and wide & deep learning. In 2018 International Conference on Innovation in Engineering and Technology (ICIET), pages 1–6. IEEE, 2018.
16. B. Navaneeth, M. Suchetha, A dynamic pooling based convolutional neural network approach to detect chronic kidney disease. Biomed. Signal Proce. Control 62, 102068 (2020)
17. A. Brunetti, G.D. Cascarano, I. De Feudis, M. Moschetta, L. Gesualdo, V. Bevilacqua, Detection and segmentation of kidneys from magnetic resonance images in patients with autosomal dominant polycystic kidney disease, in International Conference on Intelligent Computing. ed. by D.-S. Huang, K.-H. Jo, Z.-K. Huang (Springer International Publishing, Cham, 2019)
18. D. Ramos et al., Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building. Energy Rep. 8, 417–422 (2022)
19. H.E. Song et al., Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest. Chemosphere 290, 133388 (2022)
20. H.U. Rongyao et al., Multi-task multi-modality SVM for early COVID-19 diagnosis using chest CT data. Inf. Proc. Manag. 59(1), 102782 (2022)
21. X.U. Ankun et al., Artificial neural network (ANN) modeling for the prediction of odor emission rates from landfill working surface. Waste Manag. 138, 158–171 (2022)
22. D.C. Yadav, S. Pal, An Ensemble Approach on the behalf of Classification and Prediction of Diabetes Mellitus Disease Emerging Trends in Data Driven Computing and Communications (Springer, Singapore, 2021)
23. D.C. Yadav, S. Pal, Performance based evaluation of algorithms on chronic kidney disease using hybrid ensemble model in machine learning. Biomed. Pharmacol. J. 14(3), 1633–1646 (2021)

24. D.C. Yadav, S. Pal, Discovery of Thyroid Disease Using Different Ensemble Methods with Reduced Error Pruning Technique, in Computer-aided Design and Diagnosis Methods on the behalf of Biomedical Applications. ed. by G.R. Varun Bajaj, V.B. Sinha, G.R. Sinha (CRC Press, Boca Raton, 2021)
25. A. Zoda et al., Inferring genetic characteristics of Japanese Black cattle populations using genome-wide single nucleotide polymorphism markers. J. Animal Genet. 50(1), 3–9 (2022)
26. G.M. Ifraz, M.H. Rashid, T. Tazin, S. Bourouis, M.M. Khan, Comparative analysis for prediction of kidney disease using intelligent machine learning methods. Comput. Math. Methods Med. (2021). https:// doi. org/ 10. 1155/ 2021/ 61414 70
27. S. Krishnamurthy, K.S. Kapeleshh, E. Dovgan, M. Luštrek, B.G. Piletič, K. Srinivasan, Y.C. Li, A. Gradišek, S. Syed-Abdul, "Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan." medRxiv. (2020). https:// doi. org/ 10. 1101/ 2020. 06. 25. 20139 147
28. Z.U. Rehman, M.S. Zia, G.R. Bojja, M. Yaqub, F. Jinchao, K. Arshid, Texture based localization of a brain tumor from MRimages by using a machine learning approach. Med. Hypotheses 141, 109705 (2020)
29. X. Han, X. Zheng, Y. Wang, X. Sun, Y. Xiao, Y. Tang, W. Qin, Random forest can accurately predict the development of end stage renal disease in immunoglobulin a nephropathy patient. Annals Transl. Med. (2019). https:// doi. org/ 10. 21037/ atm. 2018. 12. 11
30. E.H.A. Rady, A.S. Anwar, Prediction of kidney disease stages using data mining algorithms. Inform. Med. Unlocked (2019). https:// doi. org/ 10. 1016/j. imu. 2019. 100178
31. Z. Dong, Q. Wang, Y. Ke, W. Zhang, Q. Hong, C. Liu, X. Chen, Prediction of 3-year risk of diabetic kidney disease using machine learning based on electronic medical records. J. Transl. Med. 20(1), 1–10 (2022)
32. D.C. Yadav, S. Pal, Prediction of thyroid disease using decision tree ensemble method. Human-Intell. Syst. Integr. 2(1), 89–95 (2020)
33. V. Chaurasia, S. Pal, Applications of machine learning techniques to predict diagnostic breast cancer. SN Compu. Sci. 1(5), 1–11 (2020)
34. Chaurasia, V., & Pal, S. (2014). Performance analysis of data mining algorithms for diagnosis and prediction of heart and breast cancer disease. Review of research. 3(8).