

A REVIEW OF CHRONIC KIDNEY DISEASE DIAGNOSIS USING DATA MINING TECHNIQUES

Oyenike Adunni OLUKIRAN¹, Elijah Olusayo OMIDIORA², Stephen Olatunde OLABIYISI³, Moses Kehinde AREGBESOLA⁴, Babatunde Julius OLORUNSOLA⁵

^{1,4}Department of Mathematical and Computing Sciences, Faculty of Applied Sciences, KolaDaisi University Ibadan, Oyo State, Nigeria

^{2,3}Ladoke Akintola University of Technology, Ogbomosho, Nigeria
Department of Computer Science

⁵Department of Computer Science, University of Ilorin, Ilorin, Nigeria

¹olukiran.oyenike@gmail.com, ²eomidiora@lautech.edu.ng, ³soolabiyisi@lautech.edu.ng,

⁴kennie747@yahoo.com, ⁵juliusbabatunde140@gmail.com

Keywords: Chronic Kidney Disease, Data Mining, Diagnosis, Prognosis

Abstract: Chronic Kidney Disease (CKD) has become a global health challenge, it is a situation where kidneys are damaged and toxic can no longer be filtered. The application of data mining approaches enables researchers to predict or diagnose CKD. In healthcare, the data mining technique has gained so much attention for obtaining accurate results in medical diagnosis. The technique of data mining involves the dispensation of an inter-mixture approach to know the decision-making knowledge in the database in a method that can be used in forecasting, prognosis and diagnosis. The knowledge of data mining allows the medical personnel to conduct treatment at an early stage and also enables patients to know the present state of their health condition at an earlier stage and follow necessary diet and prescriptions towards the treatment management of CKD. This paper carried out a literature review on the application techniques of data mining for the diagnosis of CKD and also future research directions were briefly mentioned.

1. INTRODUCTION

CKD is referred to as insistent changes in the structure of a kidney, function which leads to structural abnormalities such as lumps, tumours, deformities and degenerate that are obvious on imaging [1]. When an individual is being diagnosed with CKD, this can affect the working capability and quality of living of the diagnosed person. CKD may lead to other chronic illnesses like anemia, high blood pressure, diabetes, weak bones and other disorders due to damage of nerve and bad nutritional condition. High-risk individuals are the people that suffer from diabetes, hypertension, and hereditary. The CKD patients can be managed by early diagnosis, management, or treatment of the fundamental cause to reduce the development and the responsiveness to secondary processes that lead to nephron loss [2]. The management of blood pressure and inhibition of the renin-angiotensin system is the predominant aspects of therapy.

The risk factors for CKD development entail the low nephron at birth, damage of nephron as a result of age increase and chronic kidney destructions caused by exposure to toxins or sicknesses such as obesity and type II diabetes mellitus [3]. The majority of people with CKD tend to have cardiovascular diseases. For those that develop to End-Stage Renal Disease (ESRD), the restricted ease of access to renal replacement therapy is an issue in numerous countries [4].

Renal fibrosis is a general pathological appearance of CKD, irrespective of the originating insult [5]. The Kidney Disease Improving Global Outcomes (KDIGO) initiative categorises a person as getting CKD if defects in structure of kidney continue for more than three months. KDIGO defines a severity classification, defining several CKD phases using glomerular filtration rate (GFR; either estimated (eGFR) or measured (mGFR)) and the extent of

albuminuria. GFR and albuminuria are employed to categorise CKD because GFR is a properly-known marker of renal excretory function and albuminuria is an indicator of glomerular injury. Among several techniques employed in the diagnosis and prediction of CKD is the data mining techniques [6]. The technique of data mining is an inter-disciplinary domain that involves the discovery of properties or patterns of datasets [7]. The technique considers the extraction process of significant data from large datasets. The extracted data can be used in defining rules to begin a process in decision-making. Many application areas such as engineering, scientific, private or commercial companies, educational institutions, and public organizations, healthcare have used data mining to solve myriad of problems. In medicine, it is predominantly used for disease diagnosis or prediction [8]. Several techniques of data mining are available for the prediction of diseases namely classification, clustering, association rules, summarizations and regression.

The two desired models in data mining are predictive and descriptive, under these models, many tasks are applied [9]. Considering various historical data, a predictive model gives an estimation of values for data with known outcomes obtained from numerous data. A descriptive type finds patterns or connections that exist in data, unlike the predictive model, a descriptive model assists as in a way to search the data properties observed and not new properties prediction [10]. Medical data mining possesses a great impact in the medical domain towards the extraction of hidden patterns in the dataset. Application of effective approaches in data mining is proven to obtain information that is hidden inpatient data from hospital or laboratory records that further can doctors in maximising the detection rate for diagnosis of disease severity stage.

2. LITERATURE REVIEW

2.1 Data Mining

Data mining is the drawing out or in-depth excavating into data that are in diverse arrangements to access and gather information on a particular pattern [11]. The data mining approach allows huge datasets to be arranged first, then thereafter patterns are recognised and relations are produced to perform analysis on

data in solving problems. Data mining is the knowledge discovery in the database that processes data and converts it into meaningful information [12]. The results of the predictive model are suitable for analysis in a wide field such as data science, bioinformatics, health management, finance and other medical fields [13][14]. The hidden knowledge can be discovered depending on the data mining technique employed. There are many approaches used in CKD prediction or diagnosis using data mining methods. In healthcare, data mining is an intriguing domain where computational intelligence can be used in analyzing patients' records for the discovery of new knowledge useful in medical decision-making [15]. The knowledge induced is expected to improve diagnosis accuracy, treatment of disease successfully and also reduce the level of errors [16].

2.2 Techniques in Data Mining

These techniques are approaches employed in sorting data for the identification of patterns, they include association, classification, prediction and clustering [17].

- (i) **Association:** This technique looks for patterns based on the connection of a particular event to other events. It is a commonly used approach for predicting diseases as it gives the relationship of unlike features for studying and grouping outpatient with all risk factors needed for prediction [9].
- (ii) **Clustering:** This is an unsupervised learning algorithm where no class labels are provided. It locates and visually documents collections of facts that were not previously known.
- (iii) **Classification:** The technique is supervised learning, where the class labels of some training samples are supplied. Mathematical approaches such as linear programming, Naïve Bayes, Decision Trees and Artificial Neural Network are employed for the classification method [9].
- (iv) **Prediction (Forecasting):** It determines how data can result in a reasonable future prediction. The prediction technique fits in the prognostic model of data mining [8].

- (v) **Sequence:** It searches for patterns where one occurrence results in another future occurrence.
- (vi) **Summarization:** The presentation of the summary of data generated in an easily comprehensible and informative manner.
- (vii) **Regression:** This is a learning technique used to fit an equation to a particular dataset.

2.3 Classification Algorithms in Data Mining

Technique in which data is being categorised into a given number of classes [18][18]. A classification model tries to conclude the input values given for training. The classification objective is to adequately predict the target class for individual cases in the data. For example, a classification algorithm could be applied to identify a patient suffering from CKD and other diseases [19]. Techniques of classification can process a huge dataset [20]. Classification predicts categorical class labels and data is classified applying training set and class labels. Several classification algorithms include: Decision Tree, ID3, k-nearest neighbour (KNN), Naive Bayes, Support Vector Machine (SVM), and Artificial Neural Network (ANN) are used in data mining.

2.3.1 Support Vector Machine (SVM)

This is a supervised learning algorithm that performs classification, regression and even outlier detection. The linear SVM classifier operates by sketching a straight line between two classes [21]. The main function of SVM is to identify a hyperplane in N-dimensional space (N number of features) which specifically categorises data points. SVM learns by example to assign labels to objects. For instance, fraudulent credit card activity can be identified by inspecting activity reports of numerous fraudulent and non-fraudulent credit cards [22]. SVMs have been productively employed in wide biological applications and data mining. A common biomedical application of SVMs is the automatic classification of microarray gene expression profiles.

2.3.2 Artificial Neural Network (ANN)

It is a model based on the method the human brain analyses and processes information [23]. ANN comprises one or more connections of artificial neurons and that processes information

by a connectionist approach [24]. It is the base of artificial intelligence (AI) that handles problems that are sometimes impossible or difficult by human or statistical standards. ANNs are predominantly part of machine learning to identify patterns in data, this enables ANNs to be considered as an optimal solution for classification (data sorting into fixed categories), clustering (discovery of similar features among data and drawing that data together into classes) and making predictions from data (such as identifying the risk factors in CKD). The ANN structure is given in Figure 1.

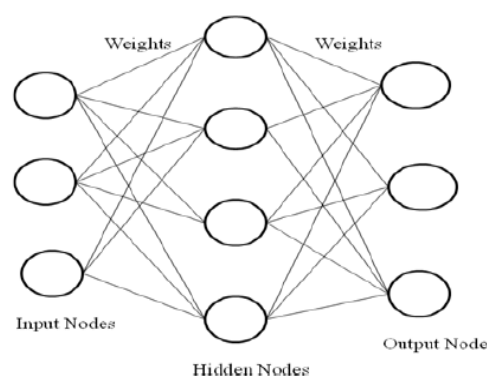


Figure 1: Structure of ANN [25]

2.3.3 Decision Tree (DT)

DT is a very unique type of probability tree that allows the making of a decision about some kind of process [26]. DT displays a technique that only contains conditional control statements. DT remains one of the classification techniques in data mining that is employed for decision support systems and machine learning processes [27]. This technique shows a significant part in process of data mining and data analysis [28]. Generally, the structure of the decision tree allows the applicability to understand the structure of trained knowledge models. DT is also a model that can be applied in prediction which is very effective for clustering, classification and predictive tasks [29].

2.3.4 Naïve Bayes (NB)

This is a learning algorithm that falls under supervised learning approach which uses the Bayes' theorem with strong independence assumptions between the features to obtain results [30], which means the algorithm assumes that individual input variable is independent (nothing to do with characteristics of other

classes). NB technique analyses the association that exists between attribute and the class for instance to produce a conditional probability for the connection between the class and attribute values [31]. This technique has been identified to work effectively with some specific datasets and when applied together with attribute selectors eliminates irrelevant and redundancy.

2.3.5 K-Nearest Neighbour (KNN)

KNN is a method of instance-based learning that represents one of the simplest algorithms used in learning algorithms for regression and classification problems [32]. This method is also called a classification based on memory as the samples required for training must be in memory at run-time [18]. KNN algorithms use data and classify new data points using similarity measures (distance function). It is simply based on the idea that objects that are near each other will also have the same features. Classification is achieved through a majority vote to its neighbors.

2.5 Chronic Kidney Disease (CKD)

CKD is also referred to as chronic kidney failure that describes the gradual loss or damage of kidney structure, which may not be sensitive and specific for the cause of disease but may lead to a reduction in function of kidney [33]. Kidney filters wastes and excess fluids from the blood, which are then excreted in the urine. When chronic kidney disease reaches an advanced stage, dangerous levels of fluid, electrolytes and wastes can build up in the body. The progression of CKD can be seen as a function of numerous parameters including underlying renal diseases, blood pressure, hypertension, proteinuria, and age [34].

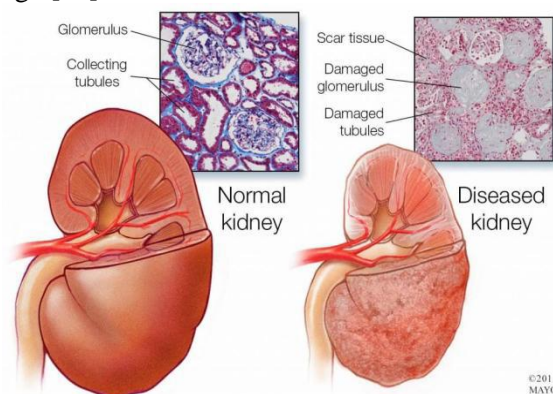


Figure 2: Structure of Normal and Diseased Kidney [35]

The early CKD diagnosis involves the great attention of physicians in knowing the suitable time to apply medical treatments and to control identified risk factors that reflect on the disease progression to End-Stage Renal Disease (ESRD) such as hypertension, proteinuria and hypophosphatemia [3]. The structure of normal and diseased kidneys is shown in Figure 2.

3. RELATED WORK

[36] developed a system for predicting CKD disease based on the CKD parameters. The CKD disease predictive identifies the disease and its stages efficiently and economically. Naive Bayes algorithms and K-Nearest Neighbour (KNN) were used for classification purposes. When multiple patient data were analyzed, the experimental result recorded an accuracy of 97% using KNN and an accuracy of 91% using Naive Bayes.

A new approach called neuro-fuzzy system for CKD diagnosis was applied by combining the fuzzy system and neural networks [16]. The system was made to generate outputs by mathematical computation and not based on a theory of probability. Results produced using mathematical computation gave higher accuracy. [37] presented a predictive model to predict CKD at an early stage. The dataset showed input parameters collected from the CKD patients and the model were trained and validated for the given input parameters. Decision tree, Random Forest and Support Vector Machine were used to classify CKD. The models were evaluated based on the accuracy of prediction. The results showed that Random Forest Classifier outperformed Decision trees and Support Vector machines.

[38] focused on the development of identification and analyzing techniques for life-threatening diseases like Chronic renal disorder (CKD) by applying data mining classification techniques such as Naïve Bayes, Artificial Neural Network(ANN), Decision Tree (C4.5) to predicts phases of Chronic urinary organ disease(CKD) at the earlier stage. The results showed that ANN gave an accuracy of 98.90% and Naïve Bayes recorded an accuracy of 96.75%.

Classification of CKD at different stages according to severity was proposed by [39]. Genetic Algorithm and C4.5 Decision Tree were used to classify CKD. The models are evaluated

using four different measures. Results obtained revealed that the genetic algorithm gave better accuracy for predicting chronic kidney disease.

[40] used several learning algorithms; C4.5, SVM and NB to predict patients with chronic kidney failure disease and patients who are not suffering from this disease. Simulation results showed that the C4.5 classifier produced the best performance in prediction in terms of accuracy and minimum execution time.

[35] analyzed and predicted chronic kidney disease by determining the unknown pattern of the relationship applying attribute selection and machine learning approach. Naive Bayes classifier and C4.5 Decision Tree were employed for classification. A dataset from UCL Irvine repository was considered to evaluate the model. Based on certain features, the model predicted whether a person has been diagnosed of having a CKD or Not CKD. While performing comparative analysis, it was shown that the C4.5 Decision Tree recorded a higher accuracy rate in prediction.

[41] applied different classification algorithms namely Naïve Bayes, Multi-layer perceptron and Support Vector Machine (SVM) to diagnose Chronic Kidney Disease. Dataset was taken from the machine learning repository of UCI. Methods of pre-processing such as missing value replacement, unsupervised discretization and normalization were used on the Chronic Kidney Disease dataset to improve accuracy. The experimental result showed that SVM gave the best accuracy of 98.75% in 0.13s.

4. SUMMARY AND DISCUSSION

Chronic Kidney Disease diagnosis is generally a difficult task, expensive, time-wasting and often risky. The early detection of this disease remains an important approach towards diagnosis and management of CKD especially in developing countries where the diseases are generally diagnosed in later stages. Several techniques in data mining and machine learning have been proposed for CKD diagnosis and management, but there is a need to develop an effective and robust diagnostic model to reduce time consumption and error in the CKD diagnosis. Future research direction should be focused on

the improvement of attribute selection/preprocessing and classification stage to produce an enhanced diagnostic model for CKD. The performance of data mining techniques together with machine learning algorithms depends majorly on the features used for diagnosis or prediction purposes [42]. If features are not adequately selected, it can lead to misclassification. Feature reduction remains one of the significant phases in the machine learning algorithm, this technique consists of feature extraction and selection that produces an approximation to the original feature in fewer dimensions, while still maintaining the same structure of original features [43]. while feature selection is the process of selecting a subset of relevant features for model construction, thus reducing training times, simplifying the models (to make interpretation easier), and improving the chances of generalization, avoiding over-fitting. The basic difference between feature selection and extraction is that feature selection retains a subset of the original features while feature extraction produces new ones [44]. With these aforementioned advantages of feature selection and machine learning algorithms. In the future study, it is necessary to develop an improved CKD predictive model through the combination of a nature-inspired computational algorithm for feature selection such as harmony search, cuckoo, bat algorithm, ant colony optimization and so on and machine learning techniques.

5. CONCLUSION

Among many approaches that have been employed by researchers in the diagnosis and prediction of CKD are data mining techniques. Data mining is an interdisciplinary area that involves the discovery of properties or patterns of datasets. The technique of data mining obtains significant information from large datasets. The extracted information can be used in defining rules to begin a process in decision-making. This technique has been used to remove the useful information from the data storage and further use the removed information for various purposes such as diagnosis of diseases. This paper presents a review of data mining techniques used for chronic kidney disease prediction and diagnosis.

Summary of the Past Techniques used in CKD using Data Mining

Author	Methodology	Advantage	Disadvantage
[39] Devi and Suganyadevi (2020).	Applied data mining techniques like Classification Techniques like Naive mathematician, Artificial Neural Network (ANN), C4.5 to predict phases of Chronic urinary organ disease (CKD). at the earlier Stage.	The Probabilistic Neural Networks technique showed the highest precision. like Chronic renal disorder (CKD)	The decision support system showed the least precision comparatively
[37] Deepika, Rao, Rampure, Prajwal, Gowda (2020)	Applied Naïve Bayes and KNN to detect the CKD	The model built was able to perform in a real-time interval, also an early-stage detection was considered.	Only two machine learning algorithms were employed
[12] Revathy, Bharathi, Jeyanthi and Ramesh, M. (2019).	Used data preprocessing, data transformation and different learning algorithms to predict CKD	The results of the framework show promising results of better prediction at an early stage of CKD	The diagnosis was not achieved in real-time application
[36] Kapoor, Verma, and Panda (2019)	analyzed and predicted chronic kidney disease by determining the unknown pattern of the relationship applying attribute selection and machine learning approach.	Results of clustering which obtained from 300 patients showed that FCM based clustering algorithms achieve higher accuracy than most existing algorithms	The approach predicted only whether a person is diagnosed with a CKD or Not CKD
[41] Boukenze, Mousannif and Haqiq (2016)	used multiple machine learning algorithms that are Support Vector Machine (SVM), Decision Tree (C4.5), and Bayesian Network (BN),	Predict kidney disease by and chose the most efficient one.	Used some selected learning algorithms
[43] Gopika and Vanitha (2017)	Proposed new algorithm Improved Hybrid Fuzzy C-Means (IHFCM) which is an improvisation of FCM with Euclidean distances to predict kidney diseases in patients	Improved Hybrid Fuzzy C-Means (IHFCM) performed efficiently	Only a lazy learning algorithm, Euclidean distance was used for the prediction task.

REFERENCES

- [1] J. Tabassum, S., Mamatha B.& Majumdar, "Analysis and Prediction of Chronic Kidney Disease using Data Mining Techniques," *Int. J. Eng. Res. Comput. Sci. Eng.*, vol. 4, no. 9, pp. 25–32, 2017.
- [2] M. K. Ameta, M. A., & Jain, "Data Mining Techniques for the Prediction of Kidney Diseases and Treatment : A Review," *Int. J. Eng. Comput. Sci.*, vol. 6, no. 2, pp. 20376–20378, 2017.
- [3] A. S. El-houssainy, A. R., & Anwar and A. S. Anwar, "Informatics in Medicine Unlocked Prediction of kidney disease stages using data mining algorithms," *Informatics Med. Unlocked*, vol. 15, pp. 1–7, 2019.
- [4] A. Kaur, G., & Sharma, "Predict Chronic Kidney Disease Using Data Mining Algorithms in Hadoop," *Int. J. Adv. Electron. Comput. Sci.*, vol. 5, no. 4, pp. 6–13, 2018.
- [5] S. Zeynu, S., & Patil, "Survey on Prediction of Chronic Kidney Disease Using Data Mining Classification Techniques and Feature

- Selection,” *Int. J. Pure Applied Math.*, vol. 118, no. 8, pp. 149–156, 2018.
- [6] S. Vijayarani, S. & Dhayanand, “Data Mining Classification Algorithms for Kidney Disease Prediction,” *Int. J. Cybern. Informatics*, vol. 4, no. 4, pp. 13–25, 2015.
- [7] D. Chandna, “Diagnosis of Heart Disease Using Data Mining Algorithm,” *International Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 1678–1680, 2014.
- [8] T. Karthikeyan, B. Ragavan, and V. A. Kanimozhi, “A Study on Data mining Classification Algorithms in Heart Disease Prediction,” *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 5, no. 4, pp. 1076–1081, 2016.
- [9] D. C. Bindushree, “Prediction of Cardiovascular Risk Analysis and Performance Evaluation Using Various Data Mining Techniques: A Review,” *Int. J. Engineering Res.*, vol. 5013, no. 5, pp. 796–800, 2016.
- [10] A. Ram, Shrawan and Doegar, “A Comparative Study of Data Mining Techniques for Predicting Disease Using Statlog Heart Disease Database,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 6, pp. 1202–1210, 2015.
- [11] I. Tarimer and B. C. Karadag, “Comparison with Classification Algorithms in Data Mining of a Fuel Automation System’s Sales Data,” *Glob. Econ. Rev.*, vol. 5, no. 1, pp. 245–254, 2020.
- [12] M. Gopika, S. & Vanitha, “Survey on Prediction of Kidney Disease by using Data Mining Techniques,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 6, no. 1, pp. 198–201, 2019.
- [13] K. A. Otunaiya and G. Muhammad, “Performance of Data mining Techniques in the Prediction of Chronic Kidney Disease,” *Comput. Sci. Inf. Technol.*, vol. 7, no. 2, pp. 48–53, 2019.
- [14] H. Karim and K. Zand, “A Comparative Survey on Data Mining Techniques for Breast Cancer Diagnosis and Prediction,” *Indian J. Fundam. Appl. Life Sci.*, vol. 5, no. 1, pp. 4330–4339, 2015.
- [15] R. Ada, & Kaur, “A Study of Detection of Lung Cancer Using Data Mining Classification Techniques,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 3, pp. 131–134, 2013.
- [16] K. V Sharma, N., & Rohit., “International Journal of Advanced Research in Prediction of Kidney Disease by using Data Mining Techniques,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 6, no. 9, pp. 66–70, 2016.
- [17] M. A. Khaleel, S. K. Pradhan, and G. N. Dash, “Finding Locally Frequent Diseases Using Modified Apriori Algorithm,” *Int. J. Adv. Res. Comput. Communication Eng.*, vol. 2, no. 10, pp. 3792–3797, 2013.
- [18] T. Sanlı, Ç. Sıcakyüz, and O. H. Yüregir, “Comparison of the accuracy of classification algorithms on three data-sets in data mining,” *Int. J. Eng. Sci. Technol.*, vol. 12, no. 3, pp. 81–89, 2020.
- [19] S. Ponmani, R. Samuel, and P. Vidhupriya, “Classification Algorithms in Data Mining – A Survey,” *A Comp. Study Classif. Tech. Data Min. Algorithms*, vol. 6, no. 1, pp. 1–6, 2017.
- [20] N. C. S. Reddy, K. S. Prasad, and A. Mounika, “Classification Algorithms on Data mining : A Study,” *Int. J. Comput. Intell. Res.*, vol. 13, no. 8, pp. 2135–2142, 2017.
- [21] A. Meligy and M. Al-Khatib, “A grid-based distributed SVM data mining algorithm,” *Eur. J. Sci. Res.*, vol. 27, no. 3, pp. 313–321, 2009.
- [22] J. Nayak, B. Naik, and H. S. Behera, “A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications & Challenges,” *Int. J. Database Theory Appl.*, vol. 8, no. 1, pp. 169–186, 2015.
- [23] S. Sergyán, “A new approach of face detection-based classification of image databases,” *Acta Polytech. Hungarica*, vol. 6, no. 1, pp. 175–184, 2009.
- [24] V. Krishnaiah, G. Narsimha, and N. S. Chandra, “Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 1, pp. 39–45, 2013.
- [25] M. M. Kasar, D. Bhattacharyya, and T. Kim, “Face Recognition Using Neural Network : A Review,” *Int. J. Secur. Its Appl.*, vol. 10, no. 3, pp. 81–100, 2016.
- [26] B. N. Patel, S. G. Prajapati, and K. I. Lakhtaria, “Efficient Classification of Data Using Decision Tree,” *Int. J. Data Min.*, vol. 2, no. 1, pp. 6–12, 2012.
- [27] Seema, M. Rathi, and Mamta, “Decision Tree: Data Mining Techniques,” *Int. J. Latest Trends Eng. Technol.*, vol. 1, no. 3, pp. 150–155, 2012.
- [28] D. Singh, H. Naveen, and J. Samota, “Analysis of Data Mining Classification with Decision Tree Technique,” *Glob. J. Comput. Sci. Technol.*, vol. 13, no. 13, pp. 1–6, 2013.
- [29] S. Govindu and B. Yugandhar, “Customer Relationship Management Based On Decision Tree Induction,” *Int. J. Comput. Sci. Technol.*, vol. 3, no. 2, pp. 1–4, 2015.
- [30] L. Marlina, A. Putera, and U. Siahaan, “Data Mining Classification Comparison (Naïve Bayes and C4 . 5 Algorithms),” *Int. J. Eng. Trends Technol.*, vol. 38, no. 7, pp. 380–383, 2016.
- [31] L. Marlina, M. Lim, and A. P. Utama Siahaan, “Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms),” *Int. J. Eng. Trends Technol.*, vol. 38, no. 7, pp. 380–383, 2016.
- [32] S. B. Imandoust and M. Bolandraftar, “Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background,” *Int. J. Eng. Res.*

- Appl., vol. 3, no. 5, pp. 605–610, 2013.
- [33] P. M. Patil, “Review on Prediction of Chronic Kidney Disease Using Data Mining Techniques,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, no. 5, pp. 135–141, 2016.
- [34] A. Tripathi, A. Nadaf, and A. K. Yadav, “Identification of the Stages of Chronic Kidney Disease Using Data Mining Approach,” *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 2, no. 8, pp. 463–467, 2020.
- [35] S. Kapoor, R. Verma, and S. N. Panda, “Detecting Kidney Disease using Naïve Bayes and Decision Tree in Machine Learning,” *Int. J. Innovative Technol. Explore. Eng.*, vol. 9, no. 1, pp. 498–501, 2019.
- [36] G. Deepika, B., Vasudeva R.K., Rampure, D.N., Prajwal, P., and Gowda, “Early Prediction of Chronic Kidney Disease by Using Machine Learning Techniques,” *Am. J. Comput. Sci. Eng. Surv.*, vol. 8, no. 2, pp. 2–5, 2020.
- [37] M. R. S.Revathy, B.Bharathi, P.Jeyanthi, “Chronic Kidney Disease Prediction using Machine Learning Models,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 6364–6367, 2019.
- [38] M. R. Devi and S. Suganyadevi, “Scrutiny Of Prophecy Chronic Kidney Diseases Through Data Mining Techniques,” *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, pp. 2860–2862, 2020.
- [39] P. Suguna and S. Prema, “J48 Decision Tree and Novel Genetic Algorithm Framework for Chronic Kidney Disorder (CKD),” *Int. J. Adv. Res. Sci. Eng. Technol.*, vol. 4, no. 11, pp. 4876–4884, 2017.
- [40] A. Boukenze, B., Mousanni, H., & Haqiq, “Performance of Data Mining Techniques to Predict in Healthcare Case Study: Chronic Kidney Failure Disease,” *Int. J. Database Management Syst.*, vol. 8, no. 3, pp. 1–9, 2016.
- [41] C. S. Kumar and P. Thangaraju, “Improving Classifier Accuracy for diagnosing Chronic Kidney Disease Using Support Vector Machines,” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 3697–3706, 2019.
- [42] N. Varghese and V. Verghese, “A Survey of Dimensionality Reduction and Classification,” *Int. J. Comput. Sci. Eng. Surv.*, vol. 3, no. 3, pp. 45–54, 2012.
- [43] S. Telgaonkar, A. H & Deshmukh, “Dimensionality Reduction and Classification through PCA and LDA,” *Int. J. Comput. Appl.*, vol. 122, no. 17, pp. 4–8, 2015.
- [44] P. K. Kumar, G., & Bhatia, “A Detailed Review of Feature Extraction in Image Processing Systems,” in *International Conference on Advanced Computing & Communication*, 2014, pp. 5–12.