Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

MACHINE LEARNING IN HEALTHCARE: PREDICTING CHRONIC KIDNEY DISEASE THROUGH FEATURE-DRIVEN HEURISTIC MODELS

Farzana Hussain[1], Sana Shaiza Shams[2], Malak Roman[*3], Masood Anwar[4], Fardad Ali Shah[5], Ayesha Bint E Meraj[6], Sajjadur Rahman[7], Muhammad Azhar Uddin[8]

[1,2,5,6,7,8]BS-Computer Science, Department of Computer Science, University of Chitral, Pakistan
*[*3,4]Lecturer, Department of Computer Science, University of Chitral-KPK, Pakistan*

[*3]malak_5116@uoch.edu.pk

DOI: https://doi.org/10.5281/zenodo.17091794

Keywords
Machine Learning, Kidney Diseases Prediction, Genetic Algorithm, K-Nearest Neighbor, Naïve Bayes, Multilayer Perceptron.

Article History
Received: 17 June 2025
Accepted: 27 August 2025
Published: 10 September 2025

Corresponding Author: *
Malak Roman

Abstract

Chronic kidney disease (CKD) is amid the substantial backer to morbidity and mortality from non-infectious syndromes that can affect 10 to 15% of the universal inhabitants. Premature and accurate recognition of the stages of CKD is whispered to be vibrant to minimize the influences of patient's healthiness complications, for instance hypertension, low blood count-anemia, bone disorder, poor nutritional vigor, neurological, and acid base abnormalities worries with timely intrusion through proper medications. Machine learning models can very well help clinicians to accomplish exactly this goal because of their prompt recognition performance. There have been different studies conducted based on machine learning methods in the identification of CKD during its premature phase. They did not pay more attention to the prediction of the given stage. In this research, the binary, and multi classification for phase prediction has been performed. The prediction models used include K-Nearest Neighbor 'KNN', Naïve Bayes 'NB', and Multilayer Perceptron 'MLP'. An evaluation of the models was done and the results from the experiment indicated that KNN has a better performance of 99.17% than NB and MLP.

INTRODUCTION

Kidneys are critical organs liable for sustaining the bodies inside environment by filtering waste away commodities and superfluous fluids from the blood. Their development begins early during fetal life, with functional activity starting around the 12th week of gestation. This early onset of function highlights the critical role of the kidneys in regulating fluid balance even before birth. Given their essential physiological role, any impairment in kidney function can have serious health implications throughout life. Chronic kidney disease (CKD) is a condition in which the kidneys gradually lose their ability to filter waste and excess water from the blood. Over time, this progressive decline in kidney function can lead to kidney failure and, ultimately, death. Often referred to as a silent killer, CKD is typically asymptomatic in its early stages [1].

Chronic kidney disease (CKD) progresses through 5-stages based on kidney function (eGFR). Stage-1 shows kidney damage but normal function (eGFR

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

>90), while Stage-2 indicates mild function loss (eGFR 60-89). Stage-3 splits into 3a (eGFR 45-59) managed by GPs and 3b (eGFR 30-44) needing specialist care, Stage-4 reveals unadorned impairment (eGFR 15-29), and Stage 5 means kidney failure (eGFR <15) requiring dialysis or transplant. Early stages often remain stable with proper care, while later stages need intensive treatment [2]. According to evaluations by the World Health Organization-WHO, millions of individuals worldwide are pretentious by this disease. The major causes of CKD include hypertension, diabetes, and genetic predisposition. Diagnosis is carried out through various methods such as serum creatinine levels, urine analysis, glomerular filtration rate (GFR), and imaging or kidney biopsy [3].

Technology has seamlessly woven itself into the fabric of our everyday existence. It is no longer just a tool; it is an invisible assistant, a primary source of information, and a central hub for our social and professional interactions [4]. In stream years, machine learning 'ML' has appeared as an effective tool in the prediction, diagnosis, and medication of chronic-kidney disease (CKD) [5]. ML algorithms have demonstrated promising results in identifying early patterns of CKD progression by analyzing complex medical data. This integration of ML into nephrology supports early intervention and enhances the precision of clinical decision-making [6]. Closely related to ML, data mining techniques further enrich the healthcare domain by enabling the extraction of meaningful patterns and trends from large-scale medical datasets. These techniques, widely applied across domains such as telecommunications and finance, are particularly transformative in healthcare, where they help identify inefficiencies, optimize resource allocation, and reduce costs [7].

Although machine learning and data mining practices have been employed in chronic kidney disease (CKD) research, most studies rely on standalone models, which can limit prediction accuracy. This study proposes a comparison of the efficiencies of three distinct models 'Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), and Bayesian Networks (BN)' with a Genetic Search Algorithm (GSA) for feature selection. By utilizing the UCI Kidney Disease Dataset, the study intentions to detect the highly effective model for enhancing early detection and supporting clinical decision-making through more precise and data-driven insights.

LITERATURE REVIEW

Snegha, J. et al., [8] performed research work on chronic kidney disease prediction in 2020. The data set was obtained from Kaggle online data bank having 400 samples. After data preprocessing and data cleaning the Random Forest, and Back Propagation Neural Network (BPNN) were implemented as predictive algorithms. Result shows that BPNN performs best with 98.40% accuracy than Random Forest algorithm.

Vijayarani, S., & Dhayanand, S. [9], implemented Naïve Bayes, and Support Vector Machine (SVM) on a synthetic Kidney Function Test (KFT). The dataset has 584 patient cases. After data preprocessing with data transformation and classification mapping SVM and Naïve Bayes machine learning algorithms were implemented. Result reveal that SVM outperformed nave bayes in terms of accuracy.

Rady, A., & Anwar, A. [10] instigated Probabilistic Neural Network (PNN), Multilayer Perceptron (MLP), SVM, and Radial Basis Function (RBF) models for predictive machine learning algorithms for 'Prediction of Kidney Disease Stages Using Data Mining Algorithms'. The proposed models were trained and tested on the dataset, which is downloaded from UCI-machine learning repository, having 361 CKD patients with 25 attributes. In the preprocessing stage missing values were handled with median replacement method. Among the predictive algorithms probabilistic neural network achieved the highest accuracy of 96.7%.

Islam, M. A. et al., [11] examined 12 machine learning algorithms for chronic kidney disease prediction, Including XGboost, AdaBoost, Decision Tree, SVM, Random Forest, Gradient Boosting, LGBM, CatBoost, KNN, Hybrid machine learning, Extra tree, Stochastic Gradient Boosting and ANN. Dataset was downloaded from UCI-database having a total of 400 instance's and 25 attributes including class label. For dataset preprocessing means and mode imputation, for optimal feature collection uses PCA. Among the 12 ML models XGBoost performed better PCA than other algorithms with the accuracy of 99.16%.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Dutta, S. et al., [12] signified their exploration on diagnosing chronic kidney disease using Logistic Regression, Decision Trees, and Random Forests as a data mining procedure. The data set has 400 patient's records, with 250 CKD-positive and 150 CKD-negative cases. The findings indicated that Logistic Regression provided the best classification results for CKD detection, making it a reliable approach for early diagnosis and clinical application. Logistic regression algorithm performance with 1.0 accuracy achieved.

Qin, J. et al., [13] applied six machine learning classifiers for chronic kidney disease (CKD) prognostication i.e., Support Vector Machine, Random Forest, Logistic Regression, K-Nearest Neighbors, feedforward neural network (FNN) and Naïve Bayes. For missing values preprocessing K-Nearest neighbors were implemented, which boosts the performance and accuracy of Random Forest reaching 99.75% among the all-applied algorithms.

Roman. M et al., [14] investigated stroke disease prediction based using K-Nearest Neighbor and Decision Tree algorithms. Data was gathered from two different hospitals in Peshawar city, having 12 features. Genetic Search, and Chi-Square were used for best feature range to achieve better expectations. Based on their results, they found that K-NN along with Genetic Search yielded an optimal accuracy of 97.5%, better than Decision Tree models.

Al-Batah, M. S., et al., [15] built an intelligent system for heart disease forecasting for Jordanian hospitals. They used one dataset of 1,025 samples each with 14 features. Various algorithms were tested such as Logistic Regression, Neural Networks, Naïve Bayes, and Random Forest. They found Random Forest produced an accuracy of 98.4%, which was the best among those investigated.

Krishnaiah, V. et al., [16] carried out a study to determine the function of data mining methods in predicting heart conditions and to identify a more effective mining method for categorising cardiac disorders. For predicting cardiac diseases, data mining methods such as Fuzzy-KNN, K-Nearest Neighbour, C4.5, J48, Neuro-Fuzzy, K-means, and Neural Network were compared. It was found that fuzzy-KNN produced better and more accurate results than the others.

Prasertsakul, T. et al., [17] provided a classification algorithm for stroke patients' rehabilitation after extracting data related to them. The MATLAB program is used to handle data from the Edinburgh University website in order to train and test the Artificial Neural Networks (ANN) and Decision Tree (D.T.) models. Results show that the decision tree outperformed with the accuracy of 98.2 % than neural network (with accuracy of 96.4%).

The implementation of artificial neural networks (ANN), naïve bayes (NB), and k-nearest neighbour algorithms (KNN) is extensively studied in the literature, which also provides us with a thorough understanding of how these algorithms can be used to predict renal illnesses in specific patients. It is observed that genetic search is significant in selecting the best attributes to give us better results for our data.

RESEARCH METHODOLOGY

An intelligent integrated model was developed for kidney diseases forecasting. The UCI machine learning repository provides access to the kidney illness dataset. Two datasets, one for training and the other for testing the model were created following data pre-processing. The best and most efficient feature selection in pre-processing is achieved with the genetic search algorithm. Figure 1 illustrates the individual applications of the K-nearest neighbour, Naïve Bayes, and artificial neural network classification models. After training the model, the accuracy of the suggested integrated model was determined using test data.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

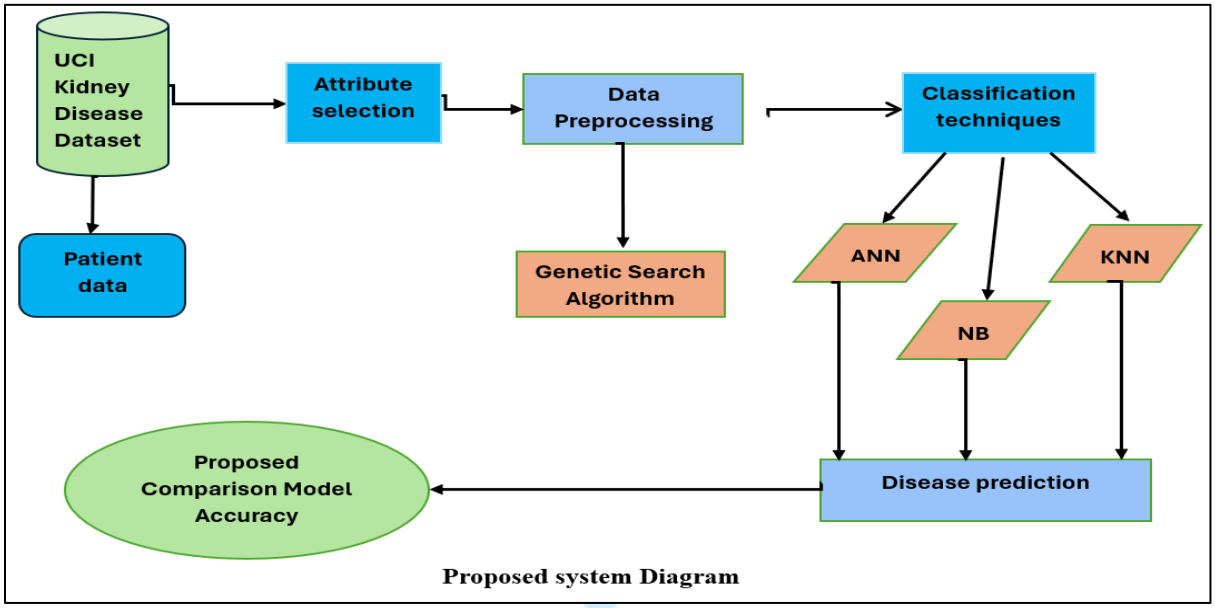

Fig1: Proposed Classifier Model Architecture

Data Collection

The persistent kidney disease data set obtained from 'UCI learning repository database', consists of 24 attributes with a class label. Table 1 presents a summary of the collected data, listing and key values attributes.

Feature Selection

Data reliability is enhanced in this stage. Data cleaning, integration, reduction, and removal of outliers' tasks are performed at this stage. Feature subset evaluation is used to dig out the least number of parameters. The selected attributes presented results like attained with all of attributes [18]. It is extraordinarily helpful to eradicate unrelated attributes and increase effectiveness.

ATTRIBUTE DESCRIPTION					
S.No	Attribute	Value	S.No	Attribute	Value
1	Age	Numerical	13	Sodium	Numerical
2	Blood pressure	Numerical mm/Hg	14	Potassium	Numerical
3	Specific gravity	Nominal	15	Hemoglobin	Numerical
4	Albumin	Nominal	16	Packed cell volume	Numerical
5	Sugar	Nominal	17	White blood cellcount	Numerical
6	Red blood	Nominal	18	Red blood count	Numerical
7	Pus cell	Nominal (Normal=0, abnormal=1)	19	Hypertension	Nominal htn =yes 1, no 0
8	Pus cell clumps	Nominal	20	Diabetes mellitus	Nominal dm = yes 1, no 0
9	Bacteria	Nominal	21	Coronary artery disease	Nominal cad= yes 1, no 0
10	Blood glucose random	Numerical	22	Appetite	Nominal Appt=yes 1, no 0)
11	Blood urea	Numerical	23	Pedal edema	Nominal Pe =yes 1, no 0
12	Serum creatinine	Numerical	24	Anemia	Nominal yes (1), no (0)

Table 1: Attributes Description

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Genetic Search Algorithm:

Genetic Algorithms (GAs) are search methods based on natural selection and genetics, arose by John Holland in the 1970s [19], using populations of potential solutions (called individuals or chromosomes). Through iterative generations, GAs improves these solutions by mimicking biological evolution. The process begins with randomly created solutions. Every single solution's quality is evaluated using a fitness function that measures how well it solves the problem [20]. During each generation, three key operations are applied i.e. selection, crossover, and mutation. Genetic algorithm workflow feature selection. Rectangles represent operations, diamond indicate decision points, and dashed lines show elitism preservation in figure 2.

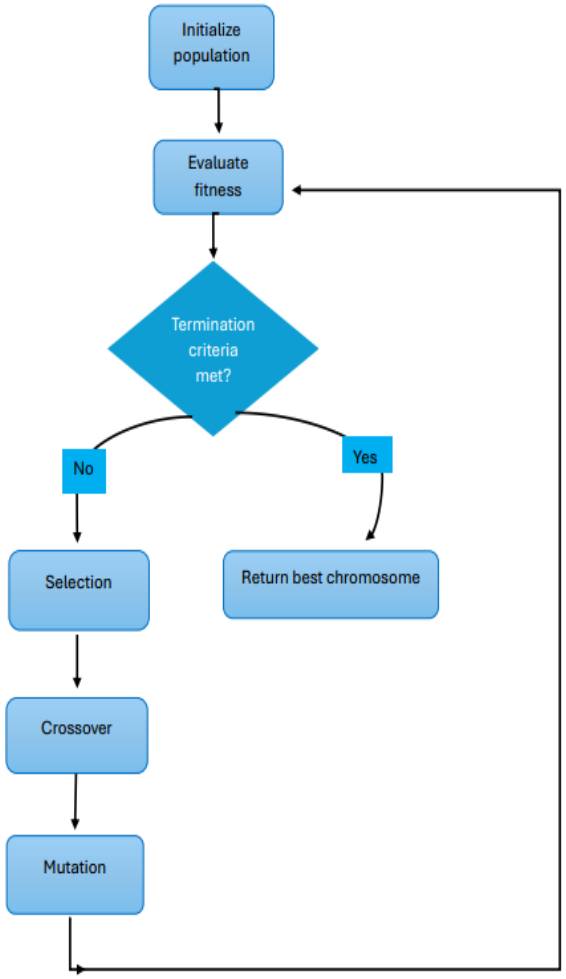

Fig 2. Flow chart of Genetic system

Gas work particularly well for complex problems because they [21]:

- Can find optimal solutions even for problems with multiple possible answers (multi-modal functions)
- Work with both continuous and discrete (separate) values
- Examine many potential solutions simultaneously (multi-point search).

Each potential solution is represented as a string called a chromosome. The population consists of multiple chromosomes with their fitness scores. Each iteration produces a new generation of improved solutions [22]. At the preprocessing stage genetic algorithm selects best 16 attributes out of totally 25 attributes as shown in table 2.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

S. NO	ATTRIBUTE	S. NO	ATTRIBUTE
1	Blood pressure	5	Blood glucose random
2	Specific gravity	6	Serum creatine
3	Albumin	7	Sodium
4	Red blood cell	8	Potassium
9	Packet cell volume	10	Appetite
11	White blood cell count	12	Peddle edema
13	Hypertension	14	Anemia
15	Diabetes mellitus	16	Hemoglobin

Table 2: Attributes selected by Genetic Algorithms

CLASSIFICATION MODELS:

The construction of a model that can be used to classify a group of items which will later be used to assign class labels or the attributes of yet unknown objects in the future is known as classification [23]. In the organized intelligent integrated model k-nearest neighbor, artificial neural network and naïve bayes classifier were used for kidney diseases forecasting.

K-Nearest Neighbor

The K-Nearest Neighbors algorithm represents a fundamental instance-based learning approach that operates through direct comparison between test samples and training data [22]. As a lazy learner, K-NN performs classification by identifying the k^{th} most similar training samples in feature space using distance metrics, with Euclidean distance being most prevalent for continuous data. The algorithm supports both classification through majority voting and regression via neighbor averaging, where the choice of k significantly impacts results - smaller values (k=1) create complex boundaries while larger values (k=5) produce smoother decision surfaces, as illustrated by cases where a test point's classification changes based on neighborhood size [24].

To measure similarity or proximity between data points, Manhattan Distance, Euclidean Distance, and Minkowski Distance formulas are used. We employed Euclidean Distance formula as shown in equation 1.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Equ.1 Euclidean Distance (L_2 Norm)

These metrics are fundamental in algorithms like k-Nearest Neighbors (k-NN) for comparing feature vectors. The Euclidean distance is particularly common for measuring the separation between training samples and test instances [25].

Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANN) are computing systems inspired by the brain's biological neurons, forming the core of deep learning. Each neuron activates based on the weighted sum of its inputs,

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

processed through an "activation function" to produce The network's behavior depends on three key components:

a. "Activation functions", which shape neuron responses.
b. "Learning rules", which adjust connection weights during training; and
c. "Network architecture", defining layer organization (input, hidden, and output layers) and connections [26]

ANN process data through their input layer (structured in matrices), transform features into hidden layers via summation and activation functions, in addition to deliver results through the output layer [27]. This structure enables ANN to efficiently learn patterns, perform classifications, and make data-driven predictions. By optimizing activation functions, learning rules, and architecture, ANN achieves powerful computational capabilities for complex problem-solving [28].

Naïve Bayesian Network

Naive Bayes is a straightforward probabilistic classifier that applies Bayes' Theorem under the assumption of feature independence. This model is more accurately described as an independent feature model, as it assumes that the presence of one feature is independent of the others, which simplifies the computation of the posterior probability [29]. Despite the unrealistic nature of this assumption in real-world scenarios, Naive Bayes performs effectively in many supervised classification tasks, often requiring a relatively small amount of training data to estimate the necessary parameters (means and variances) for each class, rather than needing to compute the full covariance matrix as shown in equation 2 [30].

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Equ. 2 Bayes' Theorem is mathematically representation

Whereas,

an output.

- P(A|B) 'is the posterior probability of class A given the feature vector B',
- P(B|A) 'is the likelihood of observing the feature vector B given class A',
- P(A) 'is the prior probability of class A',
- P(B) 'is the evidence or the total probability of feature vector B, which is constant across all classes'.

The simplicity of this approach, combined with the effectiveness in predicting outcomes, makes Naive Bayes a valuable tool, especially for tasks such as kidney disease prediction.

RESULTS AND DISCUSSION

In the context of healthcare, particularly in the early revealing of CKD, the precision of a diagnostic model is not merely a technical matter, it can define the trajectory of a patient's life. A false-negative-FN could indicate a missed identification, while a false positive-FP could lead to avoidable psychological and medical stress. With this life-saving responsibility in mind, this work evaluated the execution of three machine learning classifiers Multilayer Perceptron-(ANN), K-Nearest Neighbors (KNN), and Naive Bayes (NB). Genetic Search was employed as a feature selection strategy, reducing the dimensionality by selecting 16 most relevant features. The classifiers were assessed using standard evaluation metrics including accuracy, precision, recall (sensitivity), false positive rate (FPR), and false negative rate (FNR) [31]. These metrics were computed using the following formulas:

$$Accuracy = \frac{No\ of\ correct\ predictions}{Total\ no\ of\ predictions} \quad \#(equ.3)$$

$$Precision = \frac{True\ Positive}{True\ positive + False\ Positive} \quad \#(equ.4)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad \#(equ.5)$$

$$F1_{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \#(equ.6)$$

The K-Nearest Neighbors (KNN), classifier demonstrated excellent performance across all metrics. It achieved an accuracy of 99.17%, correctly predicting 119 out of 120 instances. The absence of false negatives is particularly important for CKD diagnosis, where missing a positive case can lead to serious medical consequences. These results make

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

KNN a promising candidate for high-stakes medical applications.

The Naïve bayes classifier also performed well, with an accuracy of 98.33% and a kappa statistic of 0.9659. Out of 120 instances. Although Naïve bayes maintained a perfect recall like KNN, its slightly higher false positive rate indicates a small risk of over-diagnosing CKD, which could be addressed through hyperparameter tuning.

ANN, in contrast, had the lowest performance among the three classifiers. It achieved an accuracy of 95%, with 114 correct predictions. This performance indicates the classifier maintains perfect specificity while being slightly conservative in positive case identification, making particularly suitable for applications where false positive is more costly than false negative, though the 8.57% false negative rate may require mitigation in critical use cases. The comparison in table 2 summarizes the performance metrics of each classifier.

Matric	KNN	Naïve Bayes	MLP	Best Performer
Accuracy	99.17%	98.33%	95.00%	KNN
Precision(yes)	100%	100%	100%	All equal
Recall(yes)	98.57%	97.14%	91.43%	KNN

Table 2: Classifiers Result Comparison

The KNN classifier outclassed others classifier models, achieving the highest accuracy and most balanced error profile. The MLP demonstrated perfect precision however, exhibited conservative classification behavior with elevated false negatives. Naïve Bayes provided a robust middle ground, balancing performance and computational efficiency. This analysis recommends KNN for general applications, with MLP preferred when false positive must be strictly avoided.

Accuracy is a central evaluation metric for classification models, demonstrating the proportion of all correct predictions out of the total predictions made [32] figure 3 graphically presents a comparative assessment of classification in relationships of accuracy across three machine learning algorithms: K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Naïve Bayes (NB). The results demonstrate a clear performance hierarchy, with KNN achieving superior accuracy (99%), followed by NB (98%), and ANN (95%). KNN's compelling performance likely stems from its ability to model complex, non-linear decision boundaries in the data. NB's moderate accuracy, while ANN's comparatively weaker performance indicates its independence assumptions may not adequately capture variable dependencies present in the dataset.

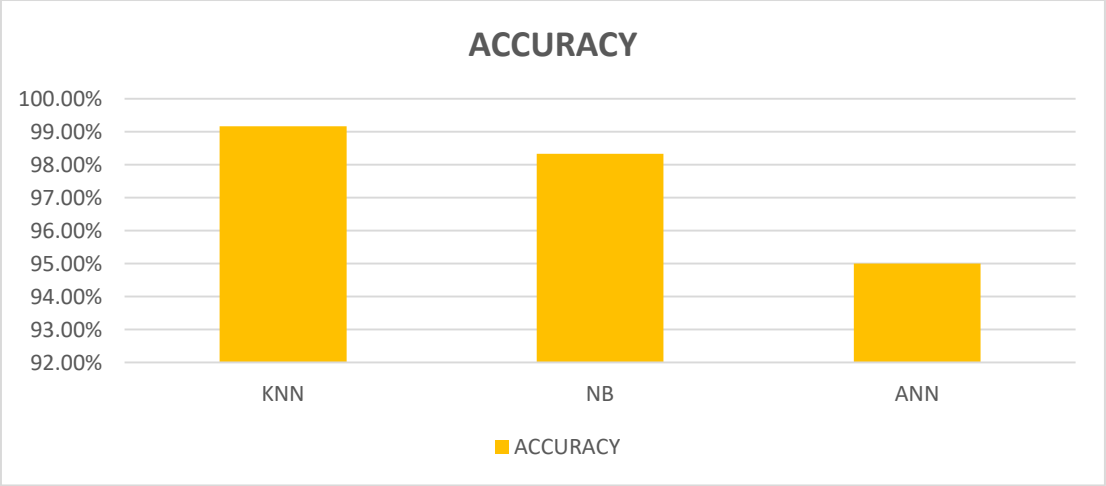

Fig 3. Visual representation of ML algorithms

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

FUTURE SCOPE

Future research could leverage data mining techniques to enhance kidney disease prediction by discovering hidden patterns in large-scale medical datasets. Feature selection and association rule mining could identify key risk factors, while clustering methods may reveal patient subgroups for personalized treatment. Temporal data mining could track disease progression, and anomaly detection might flag early warning signs. Integrating these data mining approaches with machine learning could improve both prediction accuracy and clinical interpretability for better decision-making in kidney care.

REFERENCES

[1] Wang, W., Chakraborty, G., & Chakraborty, B. (2020). Predicting the risk of chronic kidney disease using machine learning algorithm. *Applied Sciences*, *11*(1), 202.

[2] Oh, K. H., Kang, M., Kang, E., Ryu, H., Han, S. H., & Yoo, T. H. (2020). The KNOW-CKD Study: What we have learned about chronic kidney diseases. *Kidney Research and Clinical Practice*, *39*(2), 121.

[3] Dutta, S., Sikder, R., Islam, M. R., Al Mukaddim, A., Hider, M. A., & Nasiruddin, M. (2024). Comparing the Effectiveness of Machine Learning Algorithms in Early Chronic Kidney Disease Detection. *Journal of Computer Science and Technology Studies*, *6*(4), 77-91.

[4] Rahman, M. A. U., Roman, M., Ahmad, S., Jan, M. S., & Nawab, H. U. (2021). Optimizing Collaboration: Insights into Inter-Team Coordination and Self-Management in Distributed Agile Software Development. *Webology (ISSN: 1735-188X)*, *18*(5).

[5] Ullah, M. A., Ullah, A., Roman, M., Anwar, M., Siddiqi, M. U. M., Jaffar, H., & Ali, J. (2025). A Comparative Study of Machine Learning Algorithms for Cardiovascular Risk Prediction: Support Vector Machine, Gradient Boosting, And Rotation Forest. *Spectrum of Engineering Sciences*, *3*(8), 481-491.

[6] Sanmarchi, F., Fanconi, C., Golinelli, D., Gori, D., Hernandez-Boussard, T., & Capodici, A. (2023). Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review. *Journal of nephrology*, *36*(4), 1101-1117.

[7] Roman, M., Nawab, H. U., Ahmad, S., & Khan, I. A. (2022). K-Nearest Neighbor and Fuzzy K-Nearest Neighbor Algorithm Performance Analysis for Heart Disease Classification. *Webology (ISSN: 1735-188X)*, *19*(1).

[8] Snegha, J., Tharani, V., Preetha, S. D., Charanya, R., & Bhavani, S. (2020, February). Chronic kidney disease prediction using data mining. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (pp. 1-5). IEEE.

[9] Vijayarani, S., & Dhayanand, S. (2015). Data mining classification algorithms for kidney disease prediction. *Int J Cybernetics Inform*, *4*(4), 13-25.

[10] Rady, E. H. A., & Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. *Informatics in medicine unlocked*, *15*, 100178.

[11] Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of pathology informatics*, *14*, 100189.

[12] Dutta, S., Sikder, R., Islam, M. R., Al Mukaddim, A., Hider, M. A., & Nasiruddin, M. (2024). Comparing the Effectiveness of Machine Learning Algorithms in Early Chronic Kidney Disease Detection. *Journal of Computer Science and Technology Studies*, *6*(4), 77-91.

[13] Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2019). A machine learning methodology for diagnosing chronic kidney disease. *IEEE access*, *8*, 20991-21002.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

[14] Roman, M., Naz, I., Luqman, M. A., Ali, J., Jan, M. S., & Nawab, H. U. (2024). Stroke Disease Prediction Using K-Nearest Neighbor and Decision Tree Algorithms with Machine Learning Pre-Processing Techniques. *Migration Letters*, *21*(S4), 2015-2027.

[15] Al-Batah, M. S., Alzboon, M. S., & Alazaidah, R. (2023). Intelligent Heart Disease Prediction System with Applications in Jordanian Hospitals. *International Journal of Advanced Computer Science and Applications*, *14*(9).

[16] Krishnaiah, V., Narsimha, G., & Chandra, N. S. (2016). Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. *international journal of computer applications*, *136*(2), 43-51.

[17] Prasertsakul, T., Kaimuk, P., & Charoensuk, W. (2014, November). Defining the rehabilitation treatment programs for stroke patients by applying Neural Network and Decision Trees models. In *The 7th 2014 Biomedical Engineering International Conference* (pp. 1-5). IEEE.

[18] Roman, M., Ullah, A., Ullah, M. A., Hussain, F., Shams, S. S., Bint-e-Meraj, A., & Ali, S. (2025). Predicting Academic Success: A Machine Learning Approach Using Decision Tables and Random Forests Algorithms. Spectrum of Engineering Sciences, 3(5), 205-213.

[19] Holland, J. H. (1975). Adaptation in natural and artificial systems. *University of Michigan Press google schola*, 29-41.

[20] Haldurai, L., Madhubala, T., & Rajalakshmi, R. (2016). A study on genetic algorithm and its applications. *Int. J. Comput. Sci. Eng*, *4*(10), 139-143.

[21] Gaidhani, C. R., Deshpande, V. M., & Bora, V. N. (2014). Image steganography for message hiding using genetic algorithm. *International Journal of Computer Science and Engineering (IJCSE)*, *2*(3).

[22] Radwan, A. A., Latef, B. A. A., Ali, A. M. A., & Sadek, O. A. (2006). Using genetic algorithm to improve information retrieval systems. *World Academy of Science and Engineering Technology*, *17*(2), 6-13.

[23] Mining, M. D. Data Mining: Concepts and Techniques (2nd.edition).

[24] Naveen, S., & Ramamoorthy, M. (2025, April). Prediction of diabetes in women using XGB algorithm compared with KNN algorithm using machine learning. In *AIP Conference Proceedings* (Vol. 3270, No. 1, p. 020024). AIP Publishing LLC.

[25] Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). *Morgan Kaufmann*.

[26] Da Silva, I. N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L. H. B., dos Reis Alves, S. F., da Silva, I. N., ... & dos Reis Alves, S. F. (2017). *Artificial neural network architectures and training processes* (pp. 21-28). Springer International Publishing.

[27] Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PloS one*, *14*(2), e0212356.

[28] Mondal, S., Maity, R., & Nag, A. (2025). An efficient artificial neural network-based optimization techniques for the early prediction of coronary heart disease: comprehensive analysis. *Scientific Reports*, *15*(1), 4827.

[29] Olofintuyi, S. S. (2023). Breast Cancer Detection with Machine Learning Approach. *Fudma Journal of Sciences*, *7*(2), 216-222.

[30] Vijayarani, S., & Dhayanand, S. (2015). Data mining classification algorithms for kidney disease prediction. *Int J Cybernetics Inform*, *4*(4), 13-25.

[31] Anwar, M., Rahman, T., & Roman, M. (2025). Voice-Activated Smart Environments: Deep Learning Approach for Pashto Speech Command Processing. *Spectrum of Engineering Sciences*, *3*(5), 54

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

[32] Shah, F. A., Meraj, A. B. E., Roman, M., Anwar, M., Zaib, A., Shams, S. S., & Hussain, F. (2025). Enhancing Weather Forecasting Accuracy: A Machine Learning Approach Using Genetic Algorithm and Random Forest. *Global Research Journal of Natural Science and Technology, 3(2),* 283-298.