INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Improving the Classification Performance for the Effective Diagnosis of Diseases Using Optimized Weighted KNN on Clinical Dataset

C Sujdha[1]*, R Thriumalai Selvi[2], S Bharathidason[3]

1 Research Scholar, PG & Research Department of Computer Science, Government Arts College, Nandanam, Chennai-35, Tamil Nadu, India
2 Associate Professor, PG & Research Department of Computer Science, Government Arts College, Nandanam, Chennai-35, Tamil Nadu, India
3 Assistant Professor, Department of Computer Science, Loyola College (Autonomous), Chennai-34, Tamil Nadu, India

OPEN ACCESS

Received: 26-10-2024
Accepted: 09-01-2025
Published: 07-02-2025

Citation: Sujdha C, Selvi RT, Bharathidason S (2025) Improving the Classification Performance for the Effective Diagnosis of Diseases Using Optimized Weighted KNN on Clinical Dataset. Indian Journal of Science and Technology 18(3): 215-222. https://doi.org/10.17485/IJST/v18i3.3527

* **Corresponding author**.

chvisuja23@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background: This study introduces an effective technique for diagnosing diseases more accurately using the improved machine learning algorithm on clinical datasets. Early detection and prediction of diseases can significantly improve survival rates. **Objective:** The aim of this paper is to propose a novel approach to enhance the conventional KNN performance for predicting diseases more precisely. **Method:** Optimized feature and distance weights and Bayesian inspired refinements are employed in the proposed approach which is evaluated by conducting experiments on five widely-used UCI and Kaggle datasets and compared with 9 contemporary variants of KNN as well as the traditional KNN. The accuracy, precision, and recall measures are used for performance comparison. **Findings:** This research highlights the potential of machine learning, particularly the proposed OWKNN, in enhancing medical diagnostics. Improving the performance of the machine learning algorithms for medical diagnosis enables early detection and prompt intervention, thereby potentially improving patient outcomes. **Novelty:** The improved version of the KNN algorithm has been invented by integrating the feature, distance weights optimized through Genetic Algorithm and refined with the Bayesian framework. The findings demonstrate the usefulness of the OWKNN in accurately predicting diseases and offering a more reliable avenue for enhancing precise detection strategies. **Result and Discussion:** Experimental results show that the proposed method consistently outperformed the traditional KNN and considered variants of KNN in terms of average accuracy, precision, and recall scores of 91.79%, 93.16%, and 88.86% respectively across various datasets. Thus, this work comes up with an improved version of the machine learning algorithm (KNN) for effective medical diagnostics.

Keywords: KNN Classification; Optimized-Weighted KNN; Mahalanobis Distance; Mutual Information; Bayesian framework; Genetic Algorithm

1 Introduction

Medical diagnosis plays a major role in healthcare where accurate and early detection of diseases can significantly improve patient outcomes and survival rates. The emergence of machine learning (ML) has transformed medical diagnostics by enabling the analysis of complex clinical datasets to uncover patterns and insights beyond human capabilities [1].

The K-Nearest Neighbor (KNN) algorithm is a popular machine learning technique for disease classification because of its ease of use and ability to find relationships in labeled data. The applicability of classical KNN to large-scale and complex medical datasets has been limited by its drawbacks such as high computational cost, sensitivity to irrelevant features, and equal treatment of neighbors [2].

Recent research has attempted to address these issues by incorporating techniques such as Mutual Information (MI)-based feature selection, distance weight system, alternative distance metrics and optimization algorithms.

A novel method for disease classification integrated stepwise feature selection with a new weighting method gives about 10% improvement in accuracy [2]. Another paper improved KNN by assigning weights to features and using distance-based voting and got an average improvement of 2% across multiple datasets [3].

In breast cancer diagnosis, MIGA was used on the Breast Cancer Wisconsin dataset and got 99% accuracy by selecting features based on mutual information and Genetic algorithm [4]. For diabetes classification, the Tuned Fuzzy KNN (TFKNN) algorithm performed well and got 90.63% accuracy and also good in specificity, precision, and AUC [5].

KNN variants have been tested on datasets from Kaggle, UCI, and OpenML in which the Hassanat variant got an average accuracy of 83.62% for disease classification [6]. In another approach, mutual information and distance-based weighting were integrated into KNN to handle complex relationships in medical datasets and improve prediction accuracy [7]. For breast cancer severity classification, a hybrid model combining deep learning-based metaheuristic optimization algorithm with KNN performed better than Gaussian Naïve Bayes and linear SVM [8].

Dashdondov et al. worked on specific diseases like diabetes prediction using Mahalanobis distance to improve classification accuracy, especially in challenging scenarios like the COVID-19 pandemic [9]. Lung cancer diagnosis benefited from weighted KNN models with Pearson correlation and advanced data preprocessing techniques with accuracy above 98% [10,11].

Bayes-Decisive Linear KNN (BLA-KNN) method addressed broader classification challenges like interpretability and dependency issues in disease classification tasks with consistent results across multiple benchmark datasets [12]. K-means clustering and Relief feature selection-based models showed high performance in heart disease prediction with accuracy above 93% [13].

These papers show the evolution of KNN and its applications in healthcare. Though these are promising, they lack a comprehensive framework that integrates feature importance, distance weighting and metaheuristic weight optimization. Also, the prior knowledge in clinical datasets which can be leveraged through techniques like the Bayesian framework is not utilized properly. This gap shows the need for a robust and reliable KNN specifically designed for clinical datasets to improve diagnostic performance.

In this paper, we propose the Optimized Weighted K-Nearest Neighbor (OWKNN) algorithm which is a new approach that extends the traditional KNN by using mutual information for feature weighting, Mahalanobis distance for correlation aware distance weighting, and Genetic Algorithm for weight optimization. Also, Bayesian framework-based refinement is used to improve the predictive reliability. This work fills the gap in existing methods and shows the performance of OWKNN in improving disease prediction accuracy.

2 Methodology

The primary goal of this study is to present a novel optimized-weighted KNN algorithm that has a high accuracy rate and is capable of efficiently classifying test instances. This section outlines the dataset used, methods, and techniques employed throughout the study and discusses the benefits of the proposed approach.

2.1 Dataset

This study experiments with diverse medical datasets, including cardiovascular, metabolic, renal, and oncology-focused data to address various diagnostic challenges. The diversity in feature representation, dataset sizes, and class distributions ensures the OWKNN algorithm's robustness and adaptability to different clinical scenarios. Table 1 presents the details of used datasets taken from Kaggle and UCI Machine Learning Repository in this study.

Table 1. A brief list of five disease datasets considered in this study

ID	Datasets	Features	Data size	References
D1	Heart Attack Possibilities	13	303	(14)
D2	Heart Failure Outcomes	12	299	(15)
D3	Diabetes	8	768	(16)
D4	Chronic Kidney Disease Preprocessed	24	400	(17)
D5	Breast Cancer	5	569	(18)

In all the above-mentioned datasets, the instances are divided into training and test sets. 80% of instances is allocated to training and the remaining 20% is allotted to testing for all datasets. This division ensures a consistent and standardized evaluation process across all datasets, contributing to the robustness of the study's findings.

2.2 Weighted KNN with Mahalanobis Distance

The Mahalanobis distance metric is added to the conventional KNN classifier to improve it by taking scale variations and feature correlations into consideration. The Mahalanobis distance between a training instance x_i and a test instance x is calculated as follows:

$$d(x, x_i) = \sqrt{(x - x_i)^T V^{-1} (x - x_i)} \tag{1}$$

where V^{-1} is the inverse of the covariance matrix V of the training data. This metric adapts the distance based on the data distribution, leading to more accurate neighbor selection.

2.3 Feature Importance via Mutual Information

Mutual Information (MI), which measures the degree of dependence between a feature and the target variable, is used to incorporate feature importance. Given a feature X_i and a target variable Y, the mutual information $I(X_i; Y)$ can be computed as follows:

$$I(X_i; Y) = \sum_{x_i \in X_i} \sum_{y \in Y} p(x_i, y) \log \left(\frac{p(x_i, y)}{p(x_i) p(y)} \right) \tag{2}$$

The features are weighted using the computed mutual information scores, which increases the impact of significant characteristics on the classifier's decision-making process.

2.4 Optimization through Genetic Algorithm

A Genetic Algorithm (GA) is used to optimize the weights given to the features and neighbors. By determining the ideal weight set, the optimization process seeks to enhance classification accuracy. Encoded as a vector of weights w for the neighbors and features, each member of the GA population represents a potential solution. The fitness function is defined as the classification accuracy:

$$Fitness(w) = \frac{1}{n_{test}} \sum_{i=1}^{n_{est}} I(\hat{y}_i = y_i) \tag{3}$$

Where $\hat{y}_i$ is the predicted label for the i^{th} test instance, and y_i is the true label. The GA operations, including selection, crossover, and mutation, evolve the population over several generations to find the optimal weights.

2.5 Weighted Likelihood Calculation

A weighted score L(c) is computed for each class c, representing the likelihood that a given sample belongs to that class. This score incorporates an optimal neighbor weight, the inverse Mahalanobis distance, and feature importance are used to calculate a weighted score for each class c. Based on contributions from its closest neighbors, this score shows how likely it is that a

particular sample belongs in class C. Here is the calculation for the likelihood for class C:

$$L(c) = \sum_{i=1}^{k} \left(\frac{w_i}{d_i + \epsilon} \right) \times \sum_{j=1}^{m} \left(f_j \cdot w_{f_j} \right) \tag{4}$$

Where,

k is the number of nearest neighbors

w_i is the optimized weight for the i^{th} nearest neighbor

d_i is the Mahalanobis distance of the ith nearest neighbor

ε is a small constant added to avoid division by zero

m is the total number of features

f_j is the mutual information (feature importance) of the j^{th} feature.

w_{fj} is the optimized weight for the j^{th} feature.

2.6 Likelihood-based Refinement in Bayesian Framework

This Likelihood-based Refinement in the Bayesian Framework leverages the Bayesian principle of updating prior probabilities with the weighted likelihoods. While it does not use raw conditional probabilities directly, it adapts the Bayesian framework to handle high-dimensional and complex data by incorporating KNN, feature importance, and optimization techniques. This refined methodology balances theoretical consistency with practical constraints, enabling effective and interpretable class probability estimation.

The posterior probability is calculated using the Bayesian framework:

$$P(c \mid x) = \frac{L(c) \cdot P(c)}{\sum_{c'} L\left(c'\right) \cdot P\left(c'\right)} \tag{5}$$

Here:

L(c) is the weighted likelihood for class c, computed using Equation (4).

P(c) is the prior probability of class c, derived from the training data.

$\sum_{c'} L\left(c'\right) \cdot P\left(c'\right)$ is the normalization factor ensuring that the posterior probabilities sum to 1.

Classification with posterior probability involves assigning a sample to the class with the highest posterior probability P(c|x). This probability reflects the likelihood of the sample belonging to each class, updated using prior information and evidence, ensuring the decision is both data-driven and probabilistically sound.

2.7 Performance comparison

We compare the proposed OWKNN with 9 other KNN variants and the traditional KNN. We run experiments on 5 datasets and measure the methods' accuracy, precision and recall. The performance differences between KNN variants are due to their design and how they interact with the datasets. Methods that rely on distance metrics and local adaptiveness performed poorly on smaller datasets due to a lack of neighborhood diversity. Clustering based variants struggled with datasets that have high feature correlation where clustering may not capture discriminative features. These challenges highlight the need for tailored solutions like OWKNN that use optimized feature weighting and Bayesian refinement to address these issues.

2.7.1 The Classic KNN

Classic KNN: The classic KNN is a basic supervised learning algorithm that uses a fixed k to classify points based on the majority class of their nearest neighbors. Simple but can be computationally expensive and sensitive to irrelevant features [19].

2.7.2 KNN variants used for benchmarking

Adaptive KNN (A-KNN): The adaptive KNN selects the optimal k for each testing data point by inheriting the optimal k from the nearest neighbor in the training dataset. This improves accuracy by tailoring k to local data distributions [20].

Locally Adaptive KNN with Discrimination Class (LA-KNN): This variant selects the optimal k by considering the number and distribution of neighbors from the majority and second majority classes in the k-neighborhood. This improves classification by focusing on local discrimination features [21].

Fuzzy KNN (F-KNN): F-KNN assigns membership values to each class in the k-nearest neighbors using fuzzy logic. The class with the highest membership value is selected, this improves accuracy by considering uncertainty in data [22].

K-means Clustering-Based KNN (KM-KNN): This approach combines k-means clustering and 1NN by clustering the training data and using cluster centroids as a reduced training dataset. This reduces computational complexity while maintaining accuracy[23].

Weight-Adjusted KNN (W-KNN): W-KNN assigns weights to attributes or data points using a kernel function, so closer neighbors get more importance. This helps in classification of multi-attribute datasets[24].

Hassanat Distance KNN (H-KNN): This variant replaces the traditional distance metric with Hassanat distance which uses maximum and minimum vector points for better neighbor selection. It performs well in datasets with varying feature scales[25] .

Generalized Mean Distance KNN (GMD-KNN): GMD-KNN calculates distances using generalized mean vectors derived from sorted k-nearest neighbor lists for each class. It iterates over these distances to improve accuracy for complex datasets[26].

Mutual KNN (M-KNN): M-KNN focuses on mutual neighbors by trimming the training dataset to remove noise and anomalies. It uses mutual nearest neighbors for classification to reduce errors and improve robustness[27].

Ensemble Approach KNN (EA-KNN): EA-KNN uses an ensemble method to solve the fixed k problem, uses multiple k values up to a maximum (Kmax) with inverse logarithmic weighting for neighbors. This approach improves flexibility and accuracy in classification[28] .

The performance of the above KNN variants evaluated in[6] is compared with the proposed OWKNN method, as presented in Tables 2, 3 and 4.

2.7.3 Performance Comparison Measures

The performance measures used to evaluate the OWKNN and its KNN variants include accuracy, precision, and recall. Accuracy is the ratio of correct predictions (true positives and true negatives) to total predictions,

Accuracy = $\frac{TP+TN}{TP+FN+TN+FP}$

precision measures the proportion of true positives among all positive predictions.

Precision = $\frac{TP}{TP+FP}$

Recall evaluates the proportion of true positives identified among all actual positives.

Recall = $\frac{TP}{TP+FN}$

These metrics, derived from true positive (TP), true negative (TN), false positive (FP), and false negative (FN) classifications, provide a comprehensive assessment of the algorithms' classification performance.

3 Proposed Algorithm

The proposed Optimized-Weighted K-Nearest Neighbor (OW-KNN) algorithm addresses some of the key limitations of the traditional KNN algorithm, such as sensitivity to unimportant features and its equal weighting of neighbors, which can degrade the classification performance. OWKNN introduces feature importance and distance weighting which can be optimized using Genetic Algorithm (GA) and refined based on the Bayesian framework. This proposed approach improves the classification performance of the KNN algorithm on various clinical datasets.

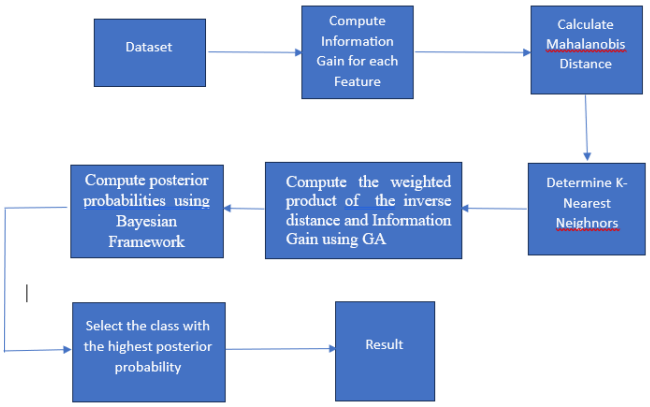

Fig 1. Proposed OW-KNN method

From Figure 1 the proposed Optimized weighted k Nearest Neighbors (OW-KNN) algorithm includes the following key steps:

Step 1: Compute the importance of each feature using the Mutual Information.
Step 2: Compute Mahalanobis distance using Equation (1).
Step 3: Identify k Nearest Neighbors Based on distance calculated.
Step 4: Compute Weighted Likelihood: For each class C_i, calculate the weighted product of the inverse distance and feature weight, optimized by Genetic Algorithm for both distance and feature importance. (Using Equation (4))
Step 5: Bayesian framework-based Refinement: Apply Bayesian framework to calculate posterior probabilities using Equation (5).
Step 6: Class Assignment: Assign the class with the highest posterior probability to the test sample.

4 Results and Discussion

This study tested various KNN versions including the proposed Optimized Weighted KNN (OWKNN) for different k values (3, 5, 7, 9, 11) and in this section, we selected the results for the k values that gave the best results for each version. The accuracy, precision and recall of the different KNN versions for the research datasets are shown in Table 2 , Table 3 and Table 4 respectively. The proposed OWKNN outperformed all other versions and gave the highest accuracy across all 5 datasets. OWKNN is clearly better than the others with an average accuracy of 91.79%.

As shown in Table 3, OWKNN also performed well in precision, best in 3 datasets and tied with the top performing methods in 1 dataset. It had an average precision of 93.16% across the datasets, that's more proof of its effectiveness. Moreover, the ranking of the proposed OWKNN was consistent across all datasets, that's proof of its robustness and reliability. These results show that OWKNN not only improves accuracy but also outperforms other KNN versions.

Table 2. Accuracy (%) comparison among KNN variants

Dataset ID	Classic KNN	Adaptive KNN	Locally adaptive KNN	Fuzzy KNN	K-means clustering-based KNN	Weight adjusted KNN	Hassanat KNN	Generalised mean distance KNN	Mutual KNN	Ensemble approach KNN	Proposed OWKNN
D1	76.35	73.64	69.59	73.65	39.86	73.65	85.14	69.59	71.62	77.03	**92.46**
D2	58.87	63.83	58.87	63.83	47.52	63.83	67.38	62.41	65.96	68.79	**88.00**
D3	75.25	75.00	75.51	74.24	68.18	74.24	76.26	74.24	74.24	79.29	**82.34**
D4	96.26	96.26	98.40	95.72	67.38	95.72	96.79	98.40	95.72	93.05	**100**
D5	90.38	92.10	90.03	91.07	89	91.07	91.07	91.41	90.72	90.38	**96.14**
Average	79.422	80.166	78.48	79.702	62.388	79.702	83.328	79.21	79.652	81.708	**91.79**

Table 3. Precision (%) comparison among KNN variants

Dataset ID	Classic KNN	Adaptive KNN	Locally adaptive KNN	Fuzzy KNN	K-means clustering-based KNN	Weight adjusted KNN	Hassanat KNN	Generalised mean distance KNN	Mutual KNN	Ensemble approach KNN	Proposed OWKNN
D1	80.90	77.42	77.11	76.84	51.52	76.84	86.96	77.11	73.53	78.57	**93.13**
D2	39.13	46.51	42.65	46.15	34.15	46.15	52.78	46.27	50	54	**91.34**
D3	62.16	60.32	60.61	61.17	49.12	61.17	63.48	58.91	60.75	74.71	**84.48**
D4	**100**	**100**	**100**	**100**	67.38	**100**	**100**	**100**	**100**	**100**	**100**
D5	98.98	98.06	97.52	**98.99**	97.49	**98.99**	**98.99**	97.12	**98.99**	98.98	96.84
Average	76.234	76.462	75.578	76.63	59.932	76.63	80.442	75.882	76.654	81.252	**93.16**

Table 4. Recall (%) comparison among KNN variants

Dataset ID	Classic KNN	Adaptive KNN	Locally adaptive KNN	Fuzzy KNN	K-means clustering-based KNN	Weight adjusted KNN	Hassanat KNN	Generalised mean distance KNN	Mutual KNN	Ensemble approach KNN	Proposed OWKNN
D1	80	80	71.11	81.11	18.89	81.11	88.89	71.11	83.33	85.56	**92.18**
D2	37.50	41.67	60.42	37.50	58.33	37.50	39.58	64.58	25	56.25	**80.89**
D3	55.20	60.80	64	50.40	22.40	50.40	58.40	60.80	52	52	**76.11**
D4	94.44	94.44	97.62	93.65	**100**	93.65	95.24	97.62	93.65	89.68	**100**
D5	88.24	91.40	89.14	89.14	87.78	89.14	89.14	91.40	88.69	88.24	**95.14**
Average	71.076	73.662	76.458	70.36	57.48	70.36	74.25	77.102	68.534	74.346	**88.86**

As shown in Table 4, the OWKNN method exhibited superior performance and guaranteed the top position with 4 best results and matched the performance of the best method in 1 dataset in terms of recall across the 5 datasets by an average of 88.86%. The provided results revealed the supremacy of the proposed method in case of recall. The experimental results explicitly demonstrate that the OWKNN method surpasses the conventional KNN and variants of the KNN method considered in this study across a variety of clinical datasets.

5 Conclusion

In this paper, we developed an Optimized Weighted K-Nearest Neighbor (OWKNN) model to improve the performance of KNN and its variants in terms of accuracy, precision and recall. The proposed OWKNN outperformed the traditional KNN and other variants with an average accuracy, precision and recall of 91.79%, 93.16%, and 88.86% respectively across the datasets. This is due to the efficient use of the optimized feature weighting and a Bayesian framework in OWKNN which makes its predictions more accurate and interpretable in clinical scenarios. The feature importance derived through the optimization process tells the most important features based on their contribution to each diagnosis and helps the clinicians understand the underlying factors behind the model's decisions. Also, the Bayesian framework enhances prediction power by incorporating probabilistic insights, enabling clinicians to better understand the likelihood of each classification outcome and make more informed and reliable critical medical decisions. Together these two elements make OWKNN more explainable and align with the high standards of transparency and trust required for clinical applications. Future work could focus on extending the OWKNN to handle multi-class classification scenarios more effectively, addressing overlapping class boundaries, and integrating the model with deep learning techniques to further enhance classification performance and feature representation.

References

1) Javaid M, Haleem A, Singh RP, Suman R, Rab S. Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*. 2022;3:58–73. Available from: https://doi.org/10.1016/j.ijin.2022.05.002.
2) Sheikhi S, Kheirabadi MT, Bazzazi A. A Novel Scheme for Improving Accuracy of KNN Classification Algorithm Based on the New Weighting Technique and Stepwise Feature Selection. *Journal of Information Technology Management (JITM)*. 2020;12(4):90–104. Available from: https://doi.org/10.22059/jitm.2020.296305.2455.
3) Syaliman KU, Labellapansa A, Yulianti A. Improving the Accuracy of Features Weighted k-Nearest Neighbor using Distance Weight. *Proceedings of the Second International Conference on Science, Engineering and Technology (ICoSET 2019)*;p. 326–330. Available from: https://www.scitepress.org/Papers/2019/93909/93909.pdf.
4) Vutakuri N, Maheswari AU. Breast cancer diagnosis using a Minkowski distance method based on mutual information and genetic algorithm. *International Journal of Advanced Intelligence Paradigms*. 2020;16(3-4):414–433. Available from: https://doi.org/10.1504/IJAIP.2020.107537.
5) Salem H, Shams MY, Elzeki OM, Elfattah MA, Al-Amri JF, Elnazer S. Fine-tuning fuzzy KNN classifier based on uncertainty membership for the medical diagnosis of diabetes. *Applied Sciences*. 2022;12(3). Available from: https://doi.org/10.3390/app12030950.
6) Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*. 2022;12. Available from: https://doi.org/10.1038/s41598-022-10358-x.
7) Vahedifar MA, Akhtarshenas A, Sabbaghian M, Rafatpanah MM, Toosi R. Information Modified K-Nearest Neighbor. *arXiv*. 2023. Available from: https://doi.org/10.48550/arXiv.2312.01991.
8) Chakravarthy SRS, Bharanidharan N, Rajaguru H. Deep learning-based metaheuristic weighted k-nearest neighbor algorithm for the severity classification of breast cancer. *IRBM*. 2023;44(3). Available from: https://doi.org/10.1016/j.irbm.2022.100749.
9) Dashdondov K, Lee S, Erdenebat MU. Enhancing Diabetes Prediction and Prevention through Mahalanobis Distance and Machine Learning Integration. *Applied Sciences*. 2024;14(17). Available from: https://doi.org/10.3390/app14177480.
10) Moon K, Jetawat A. Predicting Lung Cancer with K-Nearest Neighbors (KNN): A Computational Approach. *Indian Journal of Science and Technology*. 2024;17(21):2199–2206. Available from: https://indjst.org/articles/predicting-lung-cancer-with-k-nearest-neighbors-knn-a-computational-approach#:

~:text=Result%20and%20Discussion%3A%20The%20effectively,dataset%20comprising%2010%2C000%20patient%20records.
11) Sachdeva RK, Bathla P, Rani P, Lamba R, Ghantasala GSP, Nassar IF. A novel K-nearest neighbor classifier for lung cancer disease diagnosis. *Neural Computing and Applications*. 2024;36:22403–22416. Available from: https://link.springer.com/article/10.1007/s00521-024-10235-w.
12) Zhang J, Bian Z, Wang S. Bayes-Decisive Linear KNN with Adaptive Nearest Neighbors. *International Journal of Intelligent Systems*. 2024;2024:1–24. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1155/2024/6664942.
13) Barry KA, Manzali Y, Lamrini M, Rachid F, Elfar M. Heart Disease Prediction Using Weighted K-Nearest Neighbor Algorithm. *Operations Research Forum*. 2024;5. Available from: https://link.springer.com/article/10.1007/s43069-024-00356-2.
14) Bhat N. Health care: Heart attack possibility. 2020. Available from: https://www.kaggle.com/nareshbhat/health-care-data-set-on-heart-attack-possibility.
15) Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*. 2020;20:1–6. Available from: https://doi.org/10.1186/s12911-020-1023-5.
16) Mahgoub A. Diabetes prediction system with KNN algorithm. 2021. Available from: https://www.kaggle.com/abdallamahgoub/diabetes.
17) Soundarapandian P. Chronic Kidney Disease Data Set. 2015. Available from: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease.
18) Suwal MS. Breast Cancer Prediction Dataset. 2018. Available from: https://www.kaggle.com/datasets/merishnasuwal/breast-cancer-prediction-dataset.
19) Lopez-Bernal D, Balderas D, Ponce P, Molina A. Education 4.0: Teaching the basics of KNN, LDA and simple perceptron algorithms for binary classification problems. *Future Internet*. 2021;13(8). Available from: https://doi.org/10.3390/fi13080193.
20) Sun S, Huang R. An adaptive k-nearest neighbor algorithm. In: 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 10-12 August 2010, Yantai, China. 2010;p. 91–94. Available from: https://ieeexplore.ieee.org/document/5569740.
21) Pan Z, Wang Y, Pan Y. A new locally adaptive k -nearest neighbor algorithm based on discrimination class. *Knowledge-Based Systems*. 2020;204. Available from: https://dx.doi.org/10.1016/j.knosys.2020.106185.
22) Keller JM, Gray MR, Givens JA. A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics*. 1985;15(4):580–585. Available from: https://ieeexplore.ieee.org/document/6313426.
23) Cherif W. Optimization of K-NN algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis. *Procedia Computer Science*. 2018;127:293–299. Available from: https://dx.doi.org/10.1016/j.procs.2018.01.125.
24) Han EH, Karypis G, Kumar V. Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification. In: Advances in Knowledge Discovery and Data Mining. Berlin Heidelberg. Springer. 2001;p. 53–65. Available from: https://link.springer.com/chapter/10.1007/3-540-45357-1_9.
25) Alkasassbeh M, Altarawneh GA, Hassanat ABA. On enhancing the performance of nearest neighbour classifiers using hassanat distance metric. *arXiv*. 2015. Available from: https://doi.org/10.48550/arXiv.1501.00687.
26) Gou J, Ma H, Ou W, Zeng S, Rao Y, Yang H. A generalized mean distance-based k-nearest neighbor classifier. *Expert Systems with Applications*. 2019;115:356–372. Available from: https://dx.doi.org/10.1016/j.eswa.2018.08.021.
27) Dhar J, Shukla A, Kumar M, Gupta P. A weighted mutual k-nearest neighbour for classification mining. *arxiv*;p. 1–5. Available from: https://arxiv.org/pdf/2005.08640#:~:text=This%20Weighted%20%F0%9D%91%98%F0%9D%91%81%F0%9D%91%81%20(i.e.%2C%20%F0%9D%91%8A%F0%9D%91%98%F0%9D%91%81%F0%9D%91%81,consideration%20for%20class%20label%20prediction.&text=This%20definition%20says%20that%20given,the%20nearest%20neighbours%20of%20%F0%9D%91%A5%F0%9D%91%96.
28) Hassanat AB, Abbadi MA, Altarawneh GA, Alhasanat AA. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. *arXiv*. 2014. Available from: https://doi.org/10.48550/arXiv.1409.0919.