JPPIPA 10(11) (2024)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Heart Disease Prediction Using Optimized Weighted K-Nearest Neighbor (WKNN)

Faiq Madani[1*], Kusworo[1], Farikhin[1]

[1] Magister Sistem Informasi, Universitas Diponegoro, Semarang, Indonesia

Received: September 26, 2024
Revised: November 10, 2024
Accepted: November 25, 2024
Published: November 30, 2024

Corresponding Author:
Faiq Madani
madanifaiq11@gmail.com

DOI: 10.29303/jppipa.v10i11.9257

Abstract: Heart disease remains a significant challenge in the medical field, particularly in predictive diagnostics. This research aims to present a comprehensive investigation into the development and evaluation of a novel approach for heart disease detection using a Weighted k-Nearest Neighbors (WKNN) method. The method employs Euclidean distance metrics and Gaussian kernel weighting for optimal classification results. The research dataset consists of 200 data points, each with 10 key indicators such as age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, and ST depression relative to rest. Through rigorous experimentation, it is identified that the optimal value of *K* for classification is 11, with a *sigma* value of 1.5 for the Gaussian kernel weighting. During the training and evaluation phase, the proposed WKNN method achieved impressive performance metrics, with an accuracy of 91.8%, precision of 93%, and recall of 91%. These findings underscore the potential of the WKNN model as a reliable tool for heart disease detection, showing great promise for practical application in clinical settings. The results emphasize that the proposed method can contribute significantly to improving diagnostic accuracy for heart disease patients.

Keywords: Euclidean; Gaussian Kernel; Heart Disease; Optimized Weighted K-Nearest Neighbor

Introduction

Heart disease, a chronic and life-threatening condition, remains one of the leading causes of morbidity and mortality globally (Schneider, et al., 2024). During 2015, the World Health Organization has predicted that 17 million of life has been lost in the world due to heart-related ailments (Natarajan, et al., 2024). Heart disease, which is another name for cardiovascular illness, includes a variety of disorders that affect the heart and blood arteries, such as arrhythmias, heart failure, and coronary artery disease (R., et al., 2024). Timely detection and accurate diagnosis are crucial for effective management and prevention of severe outcomes among affected individuals. Heart disease detection presents unique challenges due to the complex interplay of various risk factors and clinical indicators. These factors include age, sex, cholesterol levels, and other physiological measures that reflect the overall health of the cardiovascular system. As a result, an accurate and reliable detection system requires an approach that can effectively capture and differentiate these intricate features.

Technology is growing faster, like social media it can be advantage or disanvantage depends on its user (L. Minocha, 2022). With the development of technology and information that is increasingly fast, it can help in difficult things to predict better for predictions one of them in the health sector (Alfyani & Muljono, 2020). The availability of big healthcare datasets and the development of machine learning techniques present a chance to improve the forecast accuracy of cardiac disease (R., et al., 2024). The classification of medical health big data is of great significance for the intelligentization of medical information (Xing & Bei, 2019). The use of Machine Learning (ML) techniques has made huge contributions to biomedical signals

How to Cite:
Madani, F., Kusworo, K., & Farikhin, F. (2024). Heart Disease Prediction Using Optimized Weighted K-Nearest Neighbor (WKNN). *Jurnal Penelitian Pendidikan IPA*, *10*(11), 8847–8854. https://doi.org/10.29303/jppipa.v10i11.9257

classification (Briouza, Gritli, Khraief, Belghith, & Singh, 2022). The development of advanced data analysis techniques and machine learning models has opened new possibilities for enhancing diagnostic accuracy and efficiency. Among these methods, there is method that is well known for several years, that is KNN.

The K-Nearest Neighbors (KNN) algorithm operates as a non-parametric, instance based learning method, commonly employed in supervised learning tasks, including classification and regression (Halder, Uddin, Uddin, Aryal, & Khraisat, 2024). The KNN method has experienced developments, one of which is WKNN. Weighted K-Nearest Neighbors (WKNN) approach, combined with Euclidean distance metrics and Gaussian kernel weighting. Euclidean Distance Matrix Analysis (EDMA) is a relative and effective method for morphologic research, which is used to analyze subject forms by special landmarks determined by the anatomical prominences (Zhai & Evison, 2011). Meanwhile, gaussian kernel is a function used in machine learning and statistical models that applies a gaussian (normal) distribution to assign weights to data points based on their distance from a central point, with closer points receiving higher weights and farther points receiving lower weights, thereby smoothing the influence of distant points (Lai & Xu, 2020). WKNN usage has shown promise in capturing the subtle patterns indicative of heart disease. Unlike traditional screening methods that rely heavily on subjective assessments, WKNN offers the potential for objective, standardized evaluations, minimizing the risk of inter-observer variability and improving diagnostic reliability.

The primary objective of this research is to introduce WKNN model specifically designed for heart disease detection using a dataset of 200 data points with 10 key variables, including age, sex, cholesterol levels, blood pressure, and exercise-induced angina. The proposed method is tailored to account for the complex relationships between these variables and heart disease outcomes. It utilizes Gaussian kernel weighting to emphasize the relevance of closer neighbors while incorporating Euclidean distance metrics for determining proximity.

Method

This research focuses on the development of the K-Nearest Neighbors (KNN) algorithm. The enhancement of KNN involves adding weighted calculations using a Gaussian Kernel, resulting in the Weighted K-Nearest Neighbor (WKNN). The addition of these weights aims to account for the contribution of each neighbor, ultimately improving the effectiveness and accuracy of the results. The following are the steps of the research process.

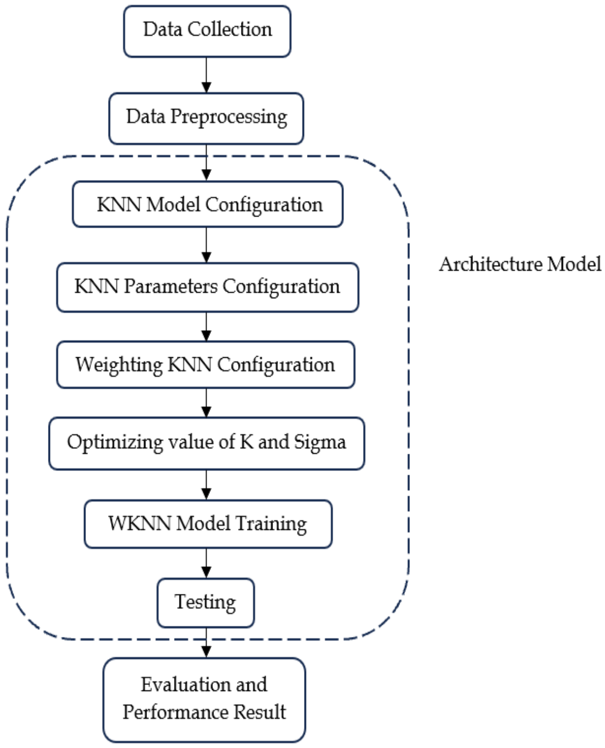

Figure 1. The steps in developing heart disease prediction using optimized WKNN

As shown in the diagram, the research process begins with data collection, where in this study, a heart disease prediction dataset is used. Following this, data preprocessing is performed. The next step is to process the model architecture, where WKNN will be employed using Euclidean as the distance metric and Gaussian Kernel for weighting calculations. The optimization of WKNN involves finding the appropriate combination of K values and sigma to achieve the highest possible accuracy.

Data Collection

Data collection is the process of gathering relevant information and datasets that are essential for the analysis and development of the research model. This step gathers required data about the patients and makes a patient simulation profile with related attributes and instances (Barwal & Raheja, 2022). In this study, data collection involves retrieving a heart disease dataset from a reliable source, which provides key indicators and symptoms required for accurate diagnosis. To conduct this research, a dataset containing symptoms and indicators related to heart disease is required. The dataset was obtained from the Kaggle website at https://www.kaggle.com/datasets/krishujeniya/heart-diseae/data. The total dataset consists of 305 entries, each containing the necessary symptom indicators to identify the presence of heart disease.

Table 1. The attribute of the heart disease dataset

No	Attributes	Symbol
1	Age of the patient	age
2	Sex of the patient	sex
3	Chest pain type	cp
4	Resting blood pressure	trestbps
5	Serum cholesterol	chol
6	Fasting blood sugar	fbs
7	Resting electrocardiographic results	restecg
8	Maximum heart rate achieved	thalach
9	Exercise-induced angina	exang
10	ST depression induced by exercise relative to rest	oldpeak

This dataset contains medical data used for predicting heart disease. The data includes various attributes such as age, sex, chest pain type (cp), resting blood pressure (trestbps), cholesterol (chol), fasting blood sugar (fbs), resting electrocardiographic results (restecg), maximum heart rate achieved (thalach), exercise-induced angina (exang), and ST depression induced by exercise relative to rest (oldpeak).

Data Preprocessing

To enhance the quality and suitability of the data, preprocessing steps is need to be performed (Islami, Sumijan, & Defit, 2024). This process involves handling missing values, removing duplicates, normalizing or scaling numerical features, and encoding categorical variables to ensure consistency. In this study, data preprocessing was essential for standardizing the heart disease dataset, allowing the WKNN model to effectively analyze and interpret the key indicators and symptoms. Proper preprocessing enhances the model’s performance by reducing noise, improving data quality, and enabling more accurate and reliable predictions.

Architecture Model

The K-Nearest Neighbors (KNN) model is a simple yet powerful machine learning algorithm widely used for classification tasks. The KNN algorithm makes predictions based on the most frequently occurring feedback from K data points nearest the test point (Duan, 2024). KNN algorithm has many advantages , such as being easy to implement and good generalization ability (Yunneng, 2020). There are three important factors in the KNN algorithm, including K value, distance measurement and decision rules (Lei, Zhu, Fang, Li, & Liu, 2020). The configuration of the KNN model involves selecting distance metrics and the number of neighbors (K) that will be considered when making predictions. The distance metrics that use for this study is euclidean distance. The Euclidean distance formula, widely used in the K-Nearest Neighbors (KNN) and Weighted K-Nearest Neighbors (WKNN) methods, is mathematically expressed as:

$$d(xi, xj) = \sqrt{\sum_{k=1}^{n}(xik - xjk)^2} \tag{1}$$

In this equation, $d(xi, xj)$ represents the Euclidean distance between two points xi and xj, where xik and xjk are the values of the k-th feature in the respective points, and n is the number of features in the dataset. A well-configured KNN model can effectively handle both small and large datasets, and its simplicity allows for easy implementation. However, the model's performance is highly dependent on the appropriate selection of parameters, such as K and the distance metric used.

To achieve optimal performance, the KNN algorithm requires careful tuning of its parameters. One of the key parameters is the number of neighbors (K), which dictates how many data points will be considered when classifying a new instance. A small value of K may lead to overfitting, as the model might become too sensitive to noise in the data, while a large value may cause underfitting by smoothing over important local patterns. Another crucial parameter is the distance metric, which determines how the proximity between points is calculated. In this research, the Euclidean distance metric is chosen for its effectiveness in measuring straight-line distances between points in multi-dimensional space.

In addition to configuring K and the distance metric, this research enhances the traditional KNN algorithm by incorporating a weighting mechanism using Gaussian Kernel. The Gaussian kernel formula, which is used for weighting in the Weighted K-Nearest Neighbors (WKNN) method, is mathematically expressed as:

$$K(xi, xj) = \exp\left(-\frac{\|xi - xj\|^2}{2\sigma^2}\right) \tag{2}$$

In this equation, $K(xi, xj)$ represents the weight assigned to the distance between two points xi and xj, where $\|xi - xj\|^2$ is the squared Euclidean distance between the two points. In Gaussian kernel function, parameter σ controls the local effect sphere of kernel function (Siyu, Chongnan, Mingjuan, & Linna, 2021).

Weighted KNN (WKNN) assigns varying levels of influence to the nearest neighbors based on their distance from the target point, giving closer neighbors more importance. The Gaussian Kernel is applied to calculate these weights, ensuring that points further away have a smaller impact on the classification decision. This configuration improves the KNN model by refining its sensitivity to local patterns in the dataset, ultimately leading to more accurate and reliable predictions, particularly in cases where the data is not uniformly distributed.

Evaluation and Performance Result

The evaluation of the Weighted K-Nearest Neighbors (WKNN) model involves assessing its performance based on various metrics that indicate its effectiveness in classifying liver disease accurately. To thoroughly evaluate the model, we employed a validation approach that included cross-validation techniques to ensure the reliability of the results. This method allowed us to analyze how well the model generalizes to unseen data by measuring its performance on different subsets of the dataset. The evaluation focused on key performance indicators such as accuracy, precision, and recall, providing a comprehensive view of the model's strengths and weaknesses. The equation used to calculate accuracy, precision and recall are as follow :

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$precision = \frac{TP}{TP+FP} \quad (4)$$

$$recall = \frac{TP}{TP+FN} \quad (5)$$

Accuracy, precision, and recall are essential metrics used to evaluate the performance of classification models. Accuracy measures the overall correctness of the model by calculating the proportion of true positive and true negative predictions among all instances. Precision, on the other hand, focuses specifically on the model's ability to identify positive cases correctly, representing the ratio of true positive predictions to the total predicted positives, which helps assess how many of the identified cases are actual positives. Recall, also known as sensitivity, evaluates the model's capability to capture all relevant instances, defined as the ratio of true positive predictions to the total actual positives, thus indicating how effectively the model identifies positive cases. Together, these metrics provide a comprehensive understanding of the model's performance, particularly in contexts where the cost of false positives and false negatives varies significantly.

Result and Discussion

The K-Nearest Neighbors method in machine learning is an algorithm that utilizes instance-based learning, also known as lazy learning (Kalcheva, Todorova, & Penev, 2023). This distance is a fundamental distance metric frequently utilized in the K-Nearest Neighbors (KNN) algorithm (Song & Pei, 2023). The preprocessed dataset will next undergo classification using the Weighted K-Nearest Neighbor method, with Euclidean distance as the distance calculation and Gaussian kernel for the weighting. Euclidean distance metric, as well as Manhattan distance metric, are the widely used distance metrics because of their good clustering accuracy and simple circuit implementation (Guo, Guo, Zeinolabedin, & Mayr, 2022). The Euclidean distance is the most frequently used definition of distance and represents the genuine distance (Lei, Yan, & Zhou, 2023). The optimization process in the WKNN method involves determining the value of *K* for the number of neighbors and the value of *sigma* to define the weight distribution in the Gaussian kernel calculation. For every neighbor in the neighborhood, a coefficient, called weight, is calculated to classify the unknown element more accurately (Mladenova & Valova, 2022). The classification process is conducted using the Python programming language and utilizes PyCharm as the compiler/IDE.

Optimizing K Value

K values, the category with the largest number of these K values can represent the category of the sample point (Hu, Li, Shuting, & Ratcliffe, 2018). Optimizing the value of *K* in the Weighted K-Nearest Neighbors (WKNN) algorithm is essential for enhancing model performance, and in this study, will be compared odd values of *K*, specifically 3, 5, 7, 9, 11, and 13. The selection of odd numbers is intentional, as it ensures a definitive majority vote when determining the class label during classification (Syriopoulos, Kalampalikis, Kotsiantis, & Vrahatis, 2023). The distance between the data is computed and the nearest distance is selected to represent the number of K (Samad, Ali, & Tajuddin, 2017). If *K* were even, there is a risk of ties, where two or more classes could receive an equal number of votes, complicating the decision-making process (Saini & Maheshkar, 2022). By evaluating these specific odd values, we aim to identify the optimal *K* that yields the highest accuracy in classifying heart disease, thereby improving the model's effectiveness and reliability. This systematic approach allows for a clearer interpretation of results and supports the goal of achieving robust classification outcomes. The result can be obtained in Figure 2.

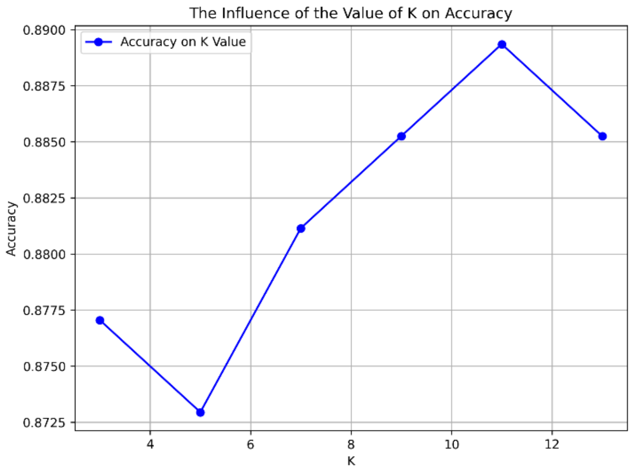

Figure 2. the influence of the k value on accuracy

Optimizing Sigma Value

Optimizing the value of *sigma* in the Weighted K-Nearest Neighbors (WKNN) algorithm is crucial for effective weighting of neighboring points, and in this study, it will examine a range of *sigma* values: 0.5, 1.0, 1.5, and 2.0. This range is selected to provide a balance between sensitivity and smoothness in the Gaussian kernel weighting (Wadhwa, Kumar, Gupta, & Kukreja, 2022). A smaller *sigma* value can lead to a sharper decay in weights, making the model highly sensitive to local variations but also prone to overfitting, as it may react excessively to noise in the data (Keylabs, 2024). Conversely, a larger *sigma* value results in a smoother weighting curve, which can enhance generalization and reduce the influence of outliers; however, it may also obscure important local patterns, leading to underfitting. The graph is depicted in Figure 3.

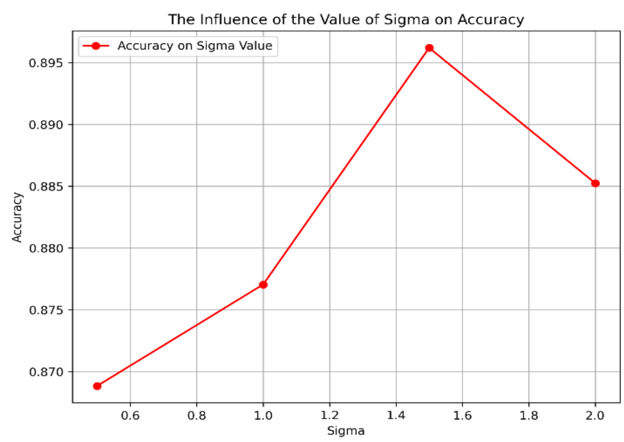

Figure 3. the influence of the sigma value on accuracy

The WKNN method is one of the machine learning (supervised) methodology that used to solve problems in classification, prediction stage (Akhiladevi, Anitha, Amrutha, & Chandanashree, 2022). After determining the optimal values for *K* and *sigma*, the next step is to combine both parameters to achieve the highest possible accuracy in the Weighted K-Nearest Neighbors (WKNN) model. In this case, the optimal *K* value is found to be 11, while the most effective *sigma* for the Gaussian kernel weighting is 1.5. The present study attempts to contribute to address two specific gaps on the current literature on immersion and distancing (Ana, et al., 2016). By integrating these values, the model can effectively balance the influence of neighboring data points and the smoothness of the weight distribution (Sih-Huei Chen, et al., 2016). The chosen *K* ensures a sufficient number of neighbors are considered without overfitting, and the selected *sigma* provides an appropriate weight distribution, capturing local patterns while minimizing noise. This combination is expected to maximize the accuracy of the model in classifying liver disease cases, providing robust and reliable predictions. The performance can be seen in Figure 4.

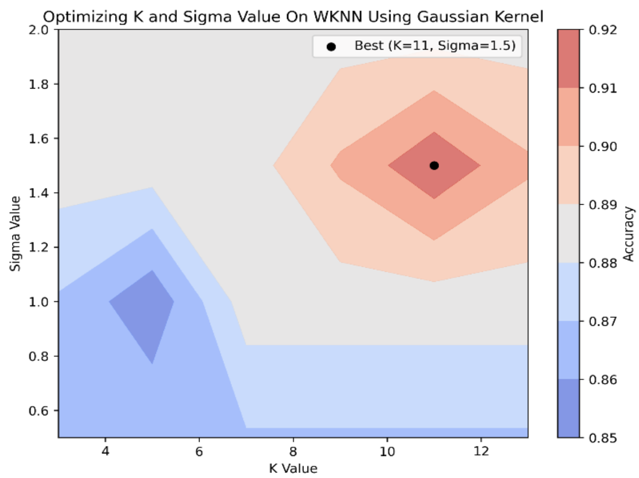

Figure 4. Optimizing k and sigma value on WKNN

Evaluating performance metrics is crucial for assessing the effectiveness of a classification model, as it provides insights into its predictive capabilities and reliability. In the context of the Weighted K-Nearest Neighbors (WKNN) model, several key metrics are used, including accuracy, precision, and recall. The comparison between accuracy, precision and recall will be shown on Figure 5.

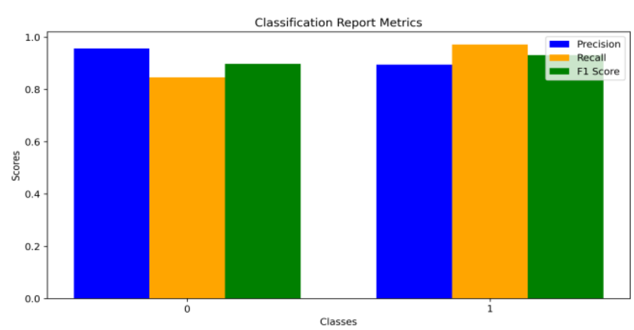

Figure 5. Performance metrics result

Based on the image above, the accuracy, precision, and recall values for label 0, indicating the absence of heart disease, are 90%, 96%, and 85%, respectively. Conversely, for label 1, which represents individuals with heart disease, the model achieves an accuracy of 93%, a precision of 89%, and a recall of 97%. These metrics highlight the model's effectiveness in distinguishing between the two labels, with particularly strong performance in identifying cases of heart disease. When averaging the results across both labels, the overall accuracy reaches 91.8%, while the precision stands at 93% and the recall at 91%. This balanced performance suggests that the model not only accurately classifies the presence or absence of heart disease but also minimizes false positives and captures a high proportion of actual positive cases, indicating its reliability and potential for practical applications in medical diagnostics. Such results underscore the importance of using multiple evaluation metrics to gain a comprehensive understanding of a model's capabilities and limitations (Kiyak, Ghasemkhani, & Birant, 2023).

The success of this research is evidenced by the high performance achieved using the Weighted K-Nearest Neighbors (WKNN) method, with Euclidean metrics and Gaussian kernel weighting, resulting in an impressive accuracy of 91.8%. This strong performance is complemented by a precision of 93% and a recall of 91%, indicating that the model effectively distinguishes between patients with and without heart disease. These results highlight the significant potential of WKNN in the healthcare field, particularly in predicting heart disease. The ability to achieve such high accuracy and reliable metrics suggests that this approach can serve as a valuable tool for medical practitioners, enabling early detection and intervention, which are crucial for improving patient outcomes. Overall, the findings affirm that WKNN, when appropriately configured, can contribute meaningfully to advancements in predictive analytics within the realm of healthcare.

Conclusion

The aim of this research was to develop a model capable of predicting the presence of heart disease using the Weighted K-Nearest Neighbors (WKNN) method, with Euclidean distance metrics and Gaussian kernel weighting. By leveraging a dataset of heart disease indicators, the study sought to optimize the classification process and improve the accuracy of predictions. The results demonstrated that the model achieved a high level of performance, with an overall accuracy of 91.8%, precision of 93%, and recall of 91%. These metrics confirm the model's ability to accurately distinguish between individuals with and without heart disease, reducing the likelihood of false positives while effectively identifying true positive cases. The findings underscore the potential of the WKNN method as a reliable tool for heart disease prediction, offering significant value to healthcare professionals in early diagnosis and intervention. The success of this model highlights its practical applicability in medical diagnostics and its potential to enhance patient outcomes through timely and accurate prediction of heart disease. In conclusion, this study not only achieved its goal of improving heart disease prediction but also demonstrated the effectiveness of WKNN in healthcare-related predictive modeling, providing a foundation for further exploration and application of this method in medical research.

Acknowledgments

Thanks to all parties who have supported the implementation of this research. I hope this research can be useful.

Author Contributions

Conceptualization, F.M.; methodology, F.M.; software, F.M.; validation, K and F.; formal analysis, F.M.; investigation, K and F.; resources, F.M.; data curation, F.M.; writing—original draft preparation, F.M.; writing—review and editing, K and F.; visualization, F.M.; supervision, K and F.; project administration, F.M.; funding acquisition, F.M. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding

Conflicts of Interest

The authors declare no conflict of interest

References

Akhiladevi, Anitha, Amrutha, & Chandanashree. (2022). Accident Prediction Using KNN Algorithm. *Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*.

Alfyani, R., & Muljono. (2020). Comparison of Naïve Bayes and KNN Algorithms to understand Hepatitis. *International Seminar on Application for Technology of Information and Communication (iSemantic)*, https://doi.org/10.1109/iSemantic50169.2020.9234299.

Ana, B. C., Barbosa, E., Silva, S., Bento, T., Teixeira, A. S., Salgado, J., & Cunha, C. (2016). Client's immersed and distanced speech and therapist's interventions in emotion-focused therapy for depression: an intensive analysis of a case study. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 136-149.

Barwal, R. K., & Raheja, N. (2022). A Classification System for Breast Cancer Prediction using SVOF-KNN method. *International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, https://doi.org/10.1109/ICAISS55157.2022.10010736.

Briouza, S., Gritli, H., Khraief, N., Belghith, S., & Singh, D. (2022). EMG Signal Classification for Human Hand Rehabilitation via Two Machine Learning Techniques: KNN and SVM. *5th International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, https://doi.org/10.1109/IC_ASET53395.2022.9765856.

Duan, M. (2024). Innovative compressive strength prediction for recycled aggregate/concrete using K-nearest neighbors and meta-heuristic optimization approaches. *Journal of Engineering and Applied Science*, 71:15. https://doi.org/10.1186/s44147-023-00348-9.

Guo, S., Guo, L., Zeinolabedin, S., & Mayr, C. (2022). Various Distance Metrics Evaluation on Neural Spike Classification. *Biomedical Circuits and Systems Conference (BioCAS)*.

Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, 113-168. https://doi.org/10.1186/s40537-024-00973-y.

Hu, B., Li, X., Shuting, S., & Ratcliffe, M. (2018). Attention Recognition in EEG-Based Affective Learning Research Using CFS+KNN Algorithm. *IEEE/ACM Transactions on Computational Biology & Bioinformatics*, 38-45. https://doi.org/10.1109/TCBB.2016.2616395.

Islami, F., Sumijan, & Defit, S. (2024). Customized Convolutional Neural Network for Glaucoma Detection in Retinal Fundus Images. *Journal of Research in Science Education*, 10(8), 4606-4613. https://doi.org/10.29303/jppipa.v10i8.7614.

Kalcheva, N., Todorova, M., & Penev, I. (2023). Study of the K-Nearest Neighbors Method with Various Features for Text Classification in Machine Learning. *International Conference AUTOMATICS AND INFORMATICS*.

Keylabs. (2024, September 13). *Keylabs.ai LTD*. Retrieved from https://keylabs.ai/blog/k-nearest-neighbors-knn-real-world-applications/

Kiyak, E. O., Ghasemkhani, B., & Birant, D. (2023). High-Level K-Nearest Neighbors (HLKNN): A Supervised Machine Learning Model for Classification Analysis. *Electronics*, 12(18):3828.

L. Minocha, P. J. (2022). Social Media's Impact on Business and Society: A Study. *8th International Conference on Advanced Computing and Communication Systems (ICACCS)*.

Lai, H., & Xu, W. (2020). Robust Gaussian kernel based signal detection in the presence of non-Gaussian noise. *International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*, https://doi.org/10.1109/BMEI.2011.6098736.

Lei, X., Yan, S., & Zhou, L. (2023). Extensions of local-mean K-nearest neighbor classifier with various distance metrics. *International Conference on Power, Intelligent Computing and Systems (ICPICS)*.

Lei, Z., Zhu, L., Fang, Y., Li, X., & Liu, B. (2020). Anomaly detection of bridge health monitoring data based on KNN algorithm. *Journal of Intelligent & Fuzzy Systems 39*, 5243–5252. https://doi.org/10.3233/JIFS-189009.

Mladenova, T., & Valova, I. (2022). Comparative analysis between the traditional KNearest Neighbor and Modifications with Weight- Calculation. *International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*.

Natarajan, K., Kumar, V. V., Mahesh, T. R., Abbas, M., Kathamuthu, N., Mohan, E., & Annand, J. R. (2024). Efficient Heart Disease Classification Through Stacked Ensemble. *International Journal of Computational Intelligence Systems*, 170-174. https://doi.org/10.1007/s44196-024-00538-0.

R., G., S. K. B, S., Mathivanan, S. K., Rajadurai, H., Malar, B. A., Mallik, S., & Qin, H. (2024). Enhancing heart disease prediction with reinforcement learning and. *Systems and Soft Computing 6*, https://doi.org/10.1016/j.sasc.2024.200129.

Saini, I., & Maheshkar, V. (2022). Effects of Distance Metrics on Color Recognition using Color Histogram and KNN Algorithm. *Delhi Section Conference (DELCON)*.

Samad, M. M., Ali, N. M., & Tajuddin, S. N. (2017). Analysis of Distance Metric Variations in KNN for Agarwood Oil Compounds Differentiation. *Conference on Systems, Process and Control*.

Schneider, K., Loizaga, S. d., Beck, A., Morales, D., Seo, J., & Divanovic, A. (2024). Socioeconomic Influences on Outcomes Following Congenital Heart. *Pediatric Cardiology*, 1072–1078. https://doi.org/10.1007/s00246-024-03451-7.

Sih-Huei Chen, Jia-Ching Wang, Wen-Chi Hsieh, Yu-Hao Chin, Chin-Wen Ho, & Chung-Hsien Wu. (2016). Speech Emotion Classification Using Multiple Kernel Gaussian Process. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*.

Siyu, Z., Chongnan, H., Mingjuan, S., & Linna, L. (2021). Two-parameter KNN algorithm and its application in recognition of brand rice. *Journal of Intelligent &*

Fuzzy Systems, https://doi.org/10.3233/JIFS-210584.

Song, Y., & Pei, F. (2023). An Improved Local Mean-Based Distance Weighted K-Nearest Neighbor with Distance Metrics. *IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE)*.

Syriopoulos, P., Kalampalikis, N., Kotsiantis, S., & Vrahatis, M. (2023). KNN Classification: a review. *Ann Math Artif Intell*, https://doi.org/10.1007/s10472-023-09882-x.

Wadhwa, S., Kumar, D., Gupta, S., & Kukreja, V. (2022). Dissected Urdu Dots Recognition Using Image Compression and KNN Classifier. *International Conference on Data Analytics for Business and Industry (ICDABI)*.

Xing, W., & Bei, Y. (2019). Medical Health Big Data Classification Based on KNN Classification Algorithm. *IEEE Access*, https://doi.org/10.1109/ACCESS.2019.2955754.

Yunneng, Q. (2020). A new stock price prediction model based on improved KNN. *7th International Conference on Information Science and Control Engineering (ICISCE)*, https://doi.org/10.1109/ICISCE50968.2020.00026.

Zhai, G., & Evison, M. (2011). Facial shape analysis based on Euclidean Distance Matrix Analysis. *4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 1896-1900. https://doi.org/10.1109/ICCSS52145.2020.9336819.