

Evaluating the Weighted K-Nearest Neighbor Algorithm: Potential and Challenges in Heart Disease Prediction

Quyen Dang

2024-12-12

Introduction

Imagine sitting in a doctor’s office, waiting for test results that could change your life. For many facing heart disease, this is an all-too-familiar reality. Globally, cardiovascular diseases are the leading cause of death, accounting for 32% of all global deaths in 2019 (World Health Organization, 2020). Yet, for most patients, diagnosis often comes too late to prevent severe consequences. However, what if a simple algorithm could predict heart disease early on, even before the first symptoms appeared? Advances in machine learning hold the promise of making this a reality, offering doctors the ability to detect and treat heart disease more effectively. One study that stands out in this field is “Heart Disease Prediction Using Weighted K-Nearest Neighbor Algorithm” by Barry et al., which introduces a novel approach to heart disease prediction using a weighted adaptation of the K-Nearest Neighbor (KNN) algorithm. The researchers claim remarkable predictive accuracy in determining heart disease risk, exemplifying the growing reliance on computational methods in healthcare decision-making. However, their work raises critical questions about data bias, patient privacy, and the potential over-reliance on algorithmic outputs. Thus, this paper will critically evaluate the statistical methodology that Barry et al. employed while diving deeper into their approach’s normative implications by considering utilitarianism. By balancing these perspectives, I hope to provide a comprehensive critique highlighting both the potential and the challenges of leveraging machine learning to advance public health.

Analysis of Methods

The research paper’s authors employed a quite comprehensive methodology to develop and validate their **weighted K-Nearest Neighbor (KNN)** algorithm for heart disease prediction. They began by aggregating five well-known heart disease datasets (Sid, 2024) and preprocessing them to handle missing values, remove outliers, and ensure data balance. Key features were grouped using the **K-means clustering** algorithm, and the **Relief** method was applied to select the most representative feature from each cluster, which was then assigned weights based on their relevance. To benchmark its performance, the weighted KNN algorithm was compared to multiple machine learning models, including *Naïve Bayes*, *Logistic Regression*, *Multilayer Perceptron (MLP)*, *Support Vector Machines (SVM)*, and the *traditional KNN*. These models were evaluated using **accuracy**, **precision**, **recall**, **F1-score**, and **AUC-ROC** metrics. This comparative framework allowed them to demonstrate the superiority of the weighted KNN algorithm, which achieved higher performance metrics than the other models.

My analysis will focus on recreating the author’s comparisons between the traditional KNN and the weighted KNN models. By narrowing the scope, I aim to validate the authors’ claim that incorporating feature weighting improves the predictive performance of the KNN algorithm. This targeted comparison allows for a better understanding of the core innovation in the study—the integration of feature selection and weighting. In contrast, if I were to spend time learning and implementing all four other machine learning models (such as *Naïve Bayes*, *Logistic Regression*, *SVM*, and *MLP*), the analysis would risk becoming too broad and taking away the focus on the weighted KNN model. This approach ensures a detailed evaluation of the

methodology’s impact without the additional complexity of analyzing other machine learning models.

Novel Analysis

To validate the authors’ methodology as described in their research paper, I followed the multi-stage process of data preprocessing, feature selection, and model evaluation as closely as possible. First, I began with data preprocessing using the publicly available heart disease dataset described in the research paper. This dataset was created by integrating five well-known cardiovascular disease datasets previously accessible separately: *Cleveland, Hungary, Switzerland, VA Long Beach, and Statlog* (Sid, 2024). In the dataset, 11 features such as **age**, **sex**, **chest pain type**, and **resting blood pressure** were described for nearly 1200 data patients (Sid, 2024). First, the dataset was checked for missing values. Consistent with the authors’ description, I verified that no missing values were present in the dataset. In addition to this, outliers were removed using the Interquartile Range (IQR) method to ensure a clean dataset. The target variable was identified, indicating whether a patient had heart disease (1) or not (0), and numeric features were identified, leading to a clean version of the dataset for further analysis.

Next, I implemented feature clustering to group similar features, as outlined in the authors’ methodology. A correlation matrix was computed to measure relationships between numeric features, and distances were calculated based on their correlations. These distances were then reduced to two dimensions using multidimensional scaling, and **K-means clustering** was applied to group features into clusters. However, the authors failed to specify the critical parameter of the number of clusters (k) used for K-means. Thus, I had to make a choice of eight clusters, which may or may not align with their implementation. This ambiguity likely influenced the subsequent feature selection process, as the clusters formed in my analysis could differ significantly from theirs.

Following feature clustering, I performed feature selection using the **Relief** algorithm. This method assigns importance weights to features based on their ability to distinguish between classes. The algorithm evaluated each feature by analyzing its interaction with neighboring data points, increasing the weights of features that contributed significantly to accurate classification. Relief has several versions, such as *ReliefF* and *RReliefF*, each optimized for different use cases (Cui et al., 2022). Without knowledge of the implementation of **Relief**, I implemented a basic version of **Relief**, but the lack of clarity in the paper left me wondering if this was the correct approach. Comparing it to the weights provided in the paper, I noticed several inconsistencies. For example, some features assigned substantial weights in the paper received significantly lower weights in my implementation. Despite this, these calculated weights were normalized and later integrated into the weighted KNN model.

With the processed data and feature weights, I implemented and tuned the weighted KNN model. As described in the research paper, the weighting mechanism modified the traditional **Euclidean distance** metric by prioritizing the most relevant features during classification. The authors described modifying the **Euclidean distance** metric by incorporating feature weights but provided no mathematical details or examples of how this was done. Therefore, I assumed to multiply the weights by the **Euclidean distance** as seen in Figure 1.

Figure 1 Weighted Euclidean Distance Formula

$$d_i = \sqrt{\sum_{j=1}^m w_j \cdot (x_{\text{train},ij} - x_{\text{test},j})^2}, \quad \forall i \in \{1, 2, \dots, n\}$$

To optimize the model, I mimicked the authors. I performed **tenfold cross-validation** to determine the best number of neighbors (k) for both the traditional KNN and the weighted KNN models. This ensured that each model was configured for optimal performance based on the dataset. While the paper mentioned optimal k values of 17 and 9 for the standard and weighted KNN models, it provided no real explanation of how these values were derived outside the mention of tenfold cross-validation. Therefore, as the paper suggested, I used tenfold cross-validation to determine k, but my optimal values differed from theirs.

Finally, I conducted a comparative evaluation of the two models as seen in Figure 2. Both the traditional KNN and weighted KNN were trained on an 80/20 train-test split of the dataset. Predictions were generated for the test set using the optimal k values identified during cross-validation. Each model’s performance metrics, including **accuracy, precision, recall, F1-score, and AUC-ROC**, were calculated to assess their effectiveness. While my weighted KNN model was supposed to outperform the traditional KNN across most metrics, the improvements were not present as reported in the paper. For example, the paper claimed an **accuracy** improvement of 5.04% for the weighted KNN over the standard KNN (Figure 3), but my analysis showed the opposite (Figure 2). In addition to the challenges outlined above, my analysis results differed significantly from those reported in the paper, with performance metrics deviating by nearly 40% in some cases. Moreover, even the simple implementation of a non-weighted KNN model had an **accuracy** that also deviated by over 10%. This suggests that the discrepancies in results are not solely due to the complexities introduced by the weighted KNN but may also stem from broader inconsistencies or omissions in the methodology described in the paper.

Figure 2 Standard KNN vs Weighted KNN Results using calculated parameters (Dang)

Metric	Standard KNN	Weighted KNN
Accuracy	0.742616	0.5949367
Precision	0.76	0.624
Recall	0.7539683	0.6141732
F1 Score	0.7569721	0.6190476
AUC Score	0.741849	0.5934503

Figure 3 Standard KNN vs Weighted KNN Results (Barry et al.)

Metric	Standard KNN	Weighted KNN
Accuracy	0.8824	0.9328
Precision	0.8872	0.9528
Recall	0.9008	0.9237
F1 Score	0.8939	0.9380
AUC Score	0.9190	0.9720

Since my results for the feature weights and overall model performance were so different from those reported in the paper, I decided to hard-code the weights and optimal k values exactly as described by the authors. I used the specific feature weights provided in their tables and set the **optimal k values** for the standard and weighted KNN models to **17** and **9**, respectively, as they mentioned. To my surprise, even with these parameters directly implemented, the performance metrics of my models still did not align with those reported in the paper as seen in Figures 4 and 5. The weighted KNN model’s **accuracy, precision, recall, and F1-score** remained significantly lower—off by nearly 20%—compared to the authors’ claims. This discrepancy suggests that there may be additional, undocumented steps or assumptions in the authors’ methodology that were critical to achieving their results.

Figure 4 Standard KNN vs Weighted KNN Results using parameters from paper (Dang)

Metric	Standard KNN	Weighted KNN
Accuracy	0.742616	0.7679325
Precision	0.76	0.864
Recall	0.7539683	0.739726
F1 Score	0.7569721	0.797048
AUC Score	0.741849	0.7764564

Figure 5 Standard KNN vs Weighted KNN Results (Barry et al.)

Metric	Standard KNN	Weighted KNN
Accuracy	0.8824	0.9328
Precision	0.8872	0.9528
Recall	0.9008	0.9237
F1 Score	0.8939	0.9380
AUC Score	0.9190	0.9720

Critique of Methodology

The methodology presented in the research paper “Heart Disease Prediction Using Weighted K-Nearest Neighbor Algorithm” by Barry et al. suffers from a significant lack of detail, which undermines its reproducibility and calls into question the validity of its results. While the authors broadly summarize their methodological approach, they fail to provide the necessary depth and clarity in critical areas, leaving individuals attempting to reproduce their results, like me, guessing at crucial implementation steps.

One of the most glaring issues lies in the description of the Relief feature selection algorithm. The paper mentions that Relief was used to assign feature weights based on their relevance. However, it does not specify which version or any specifics as to how the algorithm they utilized was implemented. Relief has numerous variants, such as ReliefF and RReliefF, each with its own implementation (Cui et al., 2022). Without this information, I was forced to implement a basic version of Relief, uncertain if it matched the authors’ approach. This ambiguity resulted in discrepancies between my feature weights and those reported in the paper, which could have been avoided with a clear explanation of the algorithm version and implementation details.

Similarly, the use of K-means clustering needs to be more detailed. While the authors describe clustering features to group similar attributes, they fail to mention critical parameters such as the number of clusters (k) used in their analysis. K-means performance and results are very sensitive to the choice of k. Without guidance, I experimented with different values, adding unnecessary uncertainty and variation to the analysis. This omission reflects a missed opportunity to guide future researchers in replicating their work.

Another significant oversight is the lack of detail about the core algorithm proposed in the paper—the weighted KNN. Ironically, the authors spend more time explaining the machine learning algorithms they compared (e.g., Naïve Bayes, Logistic Regression, and Support Vector Machines) than detailing how their weighted KNN model works. While there is a brief mention of using feature weights and optimal k values, the explanation is superficial and scattered across sections of the paper, such as the methodology, results, and even the conclusion. For an algorithm-centric study, this lack of a coherent, step-by-step description is a major flaw.

The disorganization of the paper exacerbates these issues. Key methodological steps, such as hyperparameter tuning, are scattered across various sections, including the results and conclusion, rather than being consolidated in the methodology. This makes it difficult to follow the authors’ workflow and identify the sequence of operations. Additionally, the paper lacks formulas and equations for several critical steps, which is particularly problematic for a study rooted in statistical and machine-learning techniques. For instance, while the KNN algorithm and Euclidean distance formula are briefly mentioned, the mathematical basis for the weighting scheme is completely absent.

This lack of detail and disorganization creates a barrier to reproducibility, which is a cornerstone of scientific research. My inability to replicate the authors’ results, despite following their described methodology as closely as possible, underscores the importance of thorough and transparent reporting. When critical steps are omitted or ambiguously described, it raises concerns about whether the results are correct or if the study was rushed for publication.

This issue is not unique to this paper but is indicative of a broader trend in research, particularly in fast-evolving fields like machine learning. Papers are often published with an emphasis on achieving high citation counts rather than ensuring their findings can be validated and extended by others. This undermines the

credibility of the research and slows scientific progress. To address this, journals and reviewers must enforce stricter standards for methodological transparency, requiring authors to include detailed descriptions, code repositories, and clear step-by-step workflows.

In conclusion, while the weighted KNN model proposed in this paper shows promise, the lack of methodological detail and coherence significantly detracts from its impact. Future research should prioritize clarity, transparency, and reproducibility to ensure that innovative methods like weighted KNN can be appropriately evaluated and built upon by the scientific community.

Analysis of Normative Consideration

Integrating machine learning in healthcare, particularly predictive diagnostics like heart disease detection represents a groundbreaking advancement in modern medicine. With its impressive predictive accuracy, the weighted KNN model highlights the transformative potential of machine learning to revolutionize patient care. Rooted in utilitarian philosophy, which emphasizes maximizing benefits and minimizing harm, machine learning emerges as a largely beneficial innovation when implemented thoughtfully. While it raises ethical concerns, its capacity to improve diagnostic accuracy, enhance patient outcomes, and tackle critical healthcare challenges must be addressed. By providing more benefits than drawbacks, the integration of machine learning in healthcare aligns with utilitarian ideals, reinforcing its moderated and responsible use to advance human well-being.

Medical data is inherently sensitive, containing personal information about individuals' physical and mental health. The datasets used to train machine learning algorithms like the weighted KNN for heart disease prediction typically include variables such as age, cholesterol levels, and chest pain history. While these variables are essential for improving model accuracy, they also expose patients to risks of privacy breaches. Cyberattacks on healthcare institutions have become increasingly common, and a violation of such datasets could result in identity theft or emotional distress from the public exposure of sensitive health information. Beyond the immediate consequences, even a single breach could destroy public trust in healthcare technology. Furthermore, exchanging and processing medical data across hospitals, research labs, and other entities raises critical questions about consent. Patients often need a complete understanding of how their data is used or the potential implications of contributing to these datasets. This lack of clarity, compounded by insufficient regulatory frameworks, emphasizes the urgent need for measures to protect patient data.

While data privacy remains a central concern, the consequences of predictive errors in machine learning systems are equally critical. Errors such as false positives and negatives can have profound implications. A false positive occurs when a model incorrectly predicts the presence of a disease. For heart disease, this can lead to unnecessary stress, invasive procedures, and treatments that could have emotional, physical, and financial burdens on patients and healthcare systems. On the other hand, false negatives, where the model fails to identify an actual case of heart disease, are even more risky. Misdiagnosing a patient as healthy can delay essential interventions, potentially leading to preventable deaths or irreversible complications. The stakes are exceptionally high for conditions like heart disease, where timely detection often determines outcomes. Furthermore, predictive errors can increase healthcare disparities when training datasets are not representative. For instance, a dataset skewed toward middle-aged men may lead to less accurate predictions for women or individuals from diverse ethnic backgrounds. These effects are compounded by excluding critical variables specific to minority populations, such as genetic predispositions or socioeconomic factors, reducing the model's relevance and accuracy for these groups.

Given the potential for errors, it is essential to consider the role of machine learning in medical decision-making. While algorithms like the weighted KNN offer valuable support in diagnostics, they must remain tools to aid, not replace, human judgment. Healthcare decisions are inherently complex, requiring consideration of factors beyond those captured in datasets. Algorithms can process large amounts of data efficiently, identifying patterns that might be overlooked by human detection. Still, they need a more nuanced understanding of the context and patient-specific factors that clinicians bring. Relying solely on machine learning predictions risks reducing patient care to a series of numbers, potentially missing atypical symptoms or unique circumstances. Integrating machine learning as a supportive tool allows healthcare providers to leverage their strengths while

maintaining the flexibility to adapt decisions to individual cases, ensuring a balanced approach to patient care.

Effectively addressing the ethical challenges of machine learning in healthcare requires implementing several key measures. First, robust data encryption is imperative to protect sensitive patient information during storage and transfer. Moreover, patients must be provided with clear and comprehensive information regarding the intended use of their data, the parties who will have access to it, and the associated risks and benefits, thereby enabling informed consent. Regular bias audits are essential to identify and mitigate systemic inequities embedded within machine learning models, ensuring equitable outcomes across diverse demographic groups. Furthermore, it is crucial to maintain human oversight in healthcare decision-making processes. Machine learning algorithms should function as supportive tools that enhance clinical practice rather than acting as definitive authorities. By positioning machine learning as a complementary asset, healthcare systems can uphold ethical standards while maximizing the transformative potential of technological advancements.

In conclusion, under a utilitarian framework, the integration of machine learning in healthcare represents a big step in maximizing societal benefits. Machine learning aligns with the utilitarian goal of increasing overall well-being by improving diagnostic accuracy, enabling earlier interventions, and enhancing healthcare efficiency. However, its ethical challenges underscore the need for careful implementation. When used responsibly as a supportive tool rather than a definitive decision-maker, and with measures in place to mitigate risks, machine learning holds immense potential to contribute positively to the human experience, fulfilling its promise to minimize harm and maximize good for the most significant number of people.

Conclusion and Impact

The paper “Heart Disease Prediction Using Weighted K-Nearest Neighbor Algorithm” by Barry et al. underscores the transformative potential of machine learning in healthcare diagnostics. Through its innovative weighted KNN algorithm, the study demonstrates significant improvements in predictive accuracy for heart disease risk compared to traditional methods. By integrating feature selection and weighting into the KNN framework, the authors highlight a practical application of computational techniques in advancing early disease detection. This contribution is particularly impactful in addressing the global burden of cardiovascular diseases, where timely and accurate diagnosis is critical for improving patient outcomes. However, the research also reveals methodological gaps that must be addressed to ensure reproducibility and scalability, paving the way for future refinement and broader implementation in clinical settings.

Heart disease is the leading global cause of death, and for many patients, diagnosis often comes too late to prevent severe consequences. Barry et al.’s study directly addresses this issue by proposing a weighted KNN algorithm that could predict heart disease early—even before the first symptoms appear. The proposed algorithm exemplifies the promise of machine learning in predicting heart disease early, potentially transforming patient care. The research draws attention to the urgency of integrating computational methods into healthcare, emphasizing their potential to improve accuracy and timeliness in diagnosis. While my findings affirm the algorithm’s potential, issues with data quality, unclear methods, and ethical concerns highlight the need for careful, responsible integration of machine learning into medical practice.

References

Barry, K. A., Manzali, Y., Lamrini, M., Rachid, F., & Elfar, M. (2024). Heart Disease Prediction Using Weighted K-Nearest Neighbor Algorithm. *Operations Research Forum*. <https://doi.org/10.1007/s43069-024-00356-2>

Cui, X., Li, Y., Fan, J. et al. A novel filter feature selection algorithm based on relief. *Appl Intell* 52, 5063–5081 (2022). <https://doi.org/10.1007/s10489-021-02659-x>

Sid, S. (2024). Heart Disease Dataset: Statlog, Cleveland, and Hungary Combined [Data set]. Kaggle. <https://www.kaggle.com/datasets/sid321axn/heart-statlog-cleveland-hungary-final/data>

World Health Organization. (2020, December 9). Cardiovascular diseases (CVDs) fact sheet. World Health Organization. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))