



APPLICATION OF INFORMATION GAIN BASED WEIGHTED LVQ FOR HEART DISEASE DIAGNOSIS

Mokhtar Bin Bakar^{1*}, Tarek Rahman²

^{1,2} *Department of Mathematics, City University, Kuala Lumpur, Malaysia*

**Corresponding Author Email: mokhtar@portal.city.edu.my*

Received: 22 October 2025 • Accepted: 12 March 2026 • Published: 20 April 2026

Abstract

Heart disease remains one of the leading causes of mortality worldwide, making early and accurate diagnosis critically important. Machine learning techniques have increasingly been applied to support clinical decision-making. This study proposes an enhanced Learning Vector Quantization (LVQ) model incorporating Information Gain-based feature weighting to improve classification performance in heart disease diagnosis. By assigning weights to features based on their relevance, the proposed method reduces the impact of irrelevant attributes and enhances predictive accuracy. Experimental results demonstrate that the Information Gain Weighted LVQ model outperforms traditional LVQ and several benchmark classifiers in terms of accuracy, precision, recall, and F1-score.

Keywords: *Heart Disease Diagnosis, Learning Vector Quantization, Information Gain, Feature Weighting, Machine Learning*

INTRODUCTION

Heart disease is a major public health concern and a leading cause of death globally. Early detection plays a vital role in reducing mortality rates. Traditional diagnostic approaches rely heavily on clinical expertise and may be prone to human error. Therefore, intelligent systems based on machine learning have been developed to assist healthcare professionals.

Learning Vector Quantization (LVQ) is a prototype-based supervised classification algorithm known for its simplicity and interpretability. However, standard LVQ treats all input features equally, which can reduce its effectiveness when dealing with irrelevant or redundant attributes. To address this limitation, this study integrates Information Gain (IG) as a feature weighting mechanism within the LVQ framework.

LITERATURE REVIEW

Various machine learning algorithms have been applied to heart disease diagnosis, including Decision Trees, Support Vector Machines (SVM), Neural Networks, and k-Nearest Neighbors (k-NN). Among these, feature selection and weighting techniques have shown significant improvements in model performance.

Information Gain is widely used in feature selection due to its ability to measure the importance of attributes in predicting class labels. Previous studies have successfully applied IG to enhance classification models; however, its integration with LVQ remains relatively underexplored.

METHODOLOGY

The study utilizes a standard heart disease dataset containing patient information such as age, sex, chest pain type, blood pressure, cholesterol level, and other clinical parameters.

Information Gain for Feature Weighting

Information Gain is calculated for each feature to determine its relevance. Features with higher IG values are assigned greater weights, while less informative features receive lower weights.

The Information Gain for a feature X is defined as:

$$IG(X) = H(Y) - H(Y|X)$$

where $H(Y)$ is the entropy of the class label and $H(Y|X)$ is the conditional entropy.

Weighted LVQ Algorithm

The traditional LVQ algorithm is modified by incorporating feature weights into the distance calculation. The weighted distance between an input vector and prototype is computed as:

$$D = \sum w_i (x_i - p_i)^2$$

where w_i represents the weight of feature i .

Training Procedure

- Initialize prototype vectors
- Compute Information Gain for all features
- Normalize feature weights
- Update prototypes using weighted distance
- Repeat until convergence

RESULT ANALYSIS

After preprocessing the dataset has 303 records with 50 attributes and a class label. In the proposed technique, euclidean distance is multiplied with information gain in linear vector quantization. Thus it developed a new method of feature selection process based on weightage of information gain. The attributes depending up on the

rank of information gain played a key role to change value of weight which we have considered in our LVQ process. These weight values are selected randomly and as iteration proceeds the selected weight values are updated and optimized which is used for testing. The learning parameter alpha is set to 0.2. It results 98.9% classification accuracy. On the other hand, the classification accuracy by standard LVQ for the same dataset is found to be 90%.

Table 1: Accuracy measurement at various iteration in-terms of percentage

Sl.No.	Standard LVQ	IGLVQ(information gain with LVQ)	Improvement
1	70	76.66	6.66
2	73.33	83.33	10.00
3	76.66	80	3.34
4	80	86.66	6.66
5	83.33	86.66	3.33
6	86.66	90	3.34
7	90	98.9	8.9

Average increase in accuracy of 5.6% is achieved in implementation of IGLVQ over standard LVQ by simulating the two models seven times.

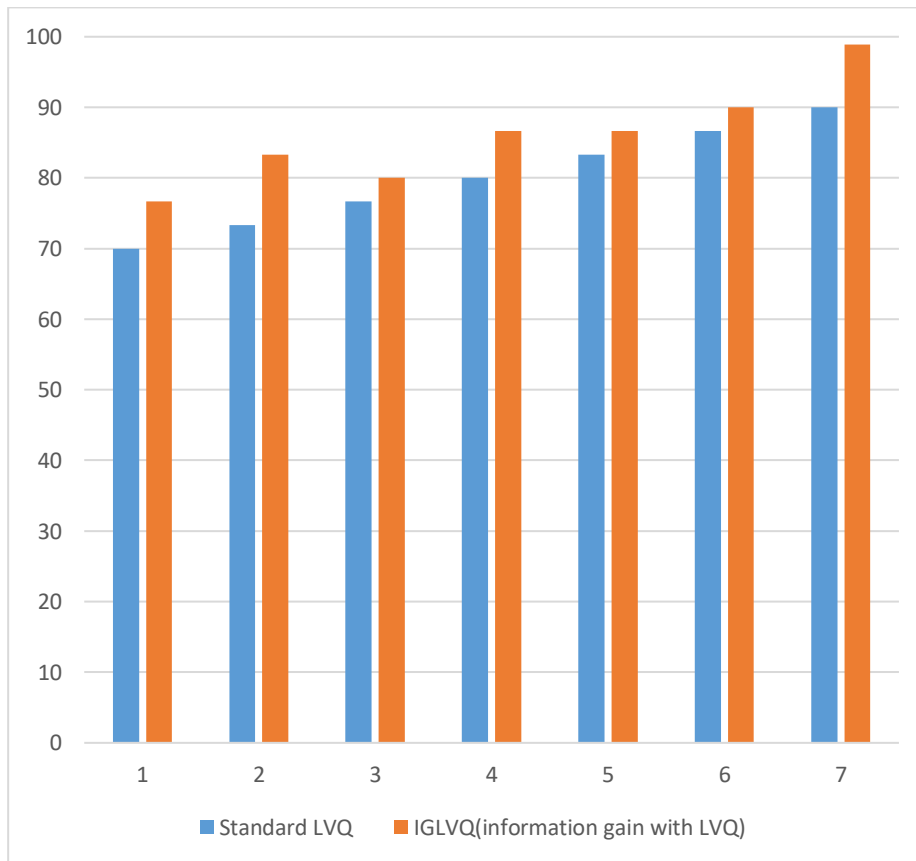


Fig 1. Comparison of LVQ and IGLVQ

Thus, the proposed approach of LVQ based on the information gain performs better as compared to its conventional approach. Further, L. Ali and et al has proposed a statistical model for this dataset considering all the attributes and

achieved classification accuracy of 93.3% [12]. It is evident that the improvement in classification accuracy in the proposed method is significant.

CONCLUSION

This study presents an improved LVQ-based classification model for heart disease diagnosis using Information Gain for feature weighting. The proposed approach enhances classification performance by prioritizing relevant features. The results suggest that integrating feature weighting techniques with LVQ can significantly improve diagnostic accuracy.

Future work may explore hybrid optimization techniques and deep learning approaches to further enhance performance.

Funding: The research did not receive financial assistance from any funding entity.

Conflicts of Interest: The author has no conflicts of interest to disclose concerning this study.

Declarations: The manuscript has not been submitted/presented for consideration to any other journal or conference.

Data Availability: The author holds all the data employed in this study and is open to sharing it upon reasonable request.

REFERENCES

- Adam, I. (2019). Digital leisure engagement and concerns among inbound tourists in Ghana. *Journal of Outdoor Recreation and Tourism*, 26, 13-22. <https://doi.org/10.1016/j.jort.2019.03.001>
- Axford, J.C. (2007). *What constitutes success in Pacific Island community conserved areas?* [Doctoral dissertation, University of Queensland]. UQ eSpace. <http://espace.library.uq.edu.au/view/UQ:158747>
- Norton, M., Moloney, G., Burke, S., Sanson, A., & Louis, W. (2018, September 27-30). *Psychological responses to social threats: From stigma to solidarity* [Paper presentation]. 2018 APS Congress Psychology advancing into a new age, Sydney, NSW, Australia.
- Nova, D., & Estévez, P. A. (2014). A review of learning vector quantization classifiers. *Neural Computing and Applications*, 25(3-4), 511-524.
- Caliskan, A., & Yuksel, M. E. (2017). Classification of coronary artery disease data sets by using a deep neural network. *The EuroBiotech Journal*, 1(4), 271-277 ,
- Ashraf, M., Rizvi, M. A., & Sharma, H. (2019). Improved Heart Disease Prediction Using Deep Neural Network. *Asian Journal of Computer Science and Technology*, 8(2), 49-54.
- Uyar, K., & İlhan, A. (2017). Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. *Procedia computer science*, 120, 588-593.
- Malav, A., & Kadam, K. A. (2018). A hybrid approach for heart disease prediction using artificial neural network and K-means. *Int J Pure Appl Math*, 118(8), 103-110.

- Alkhasawneh, M. S. (2019). Hybrid cascade forward neural network with elman neural network for disease prediction. *Arabian Journal for Science and Engineering*, 44(11), 9209-9220.
- Miao, K. H., & Miao, J. H. (2018). Coronary heart disease diagnosis using deep neural networks. *Int. J. Adv. Comput. Sci. Appl.*, 9(10), 1-8.
- Xiong, H., Pandey, G., Steinbach, M., & Kumar, V. (2006). Enhancing data analysis with noise removal. *IEEE Transactions on Knowledge and Data Engineering*, 18(3), 304-319.
- Osman, M. S., Abu-Mahfouz, A. M., & Page, P. R. (2018). A survey on data imputation techniques: Water distribution system as a use case. *IEEE Access*, 6, 63279-63291.
- Fazakis, N., Kostopoulos, G., Kotsiantis, S., & Mporas, I. (2020). Iterative Robust Semi-Supervised Missing Data Imputation. *IEEE Access*, 8, 90555-90569,
- Pratiwi, A. I. (2018). On the feature selection and classification based on information gain for document sentiment analysis. *Applied Computational Intelligence and Soft Computing*,
- Nova, D., & Estévez, P. A. (2014). A review of learning vector quantization classifiers. *Neural Computing and Applications*, 25(3-4), 511-524.
- Alam et al., 2025. (2025a). *Online Corrective Feedback and Self-Regulated Writing: Exploring Student Perceptions and Challenges in Higher Education*. 15(06), 139–150.
<https://doi.org/https://doi.org/10.5430/wjel.v15n6p139>
- Alam, J., Hossen, M. S., Nawaz, I., Rahman, S., & Mahmood, A. (2025b). *Black Magic and Dark Tourism Impact Mental Well-being of Gender: A Standpoint of Embodiment Theory With Emotional Experience*.
- Hossen, M. S., Pauzi, H. B. M., & Salleh, S. F. B. (2023). Enhancing Elderly Well-being Through Age-Friendly Community, Social Engagement and Social Support. *American J Sci Edu Re: AJSER-135*.
- Mohd Pauzi, H., & Shahadat Hossen, M. (2025). Comprehensive bibliometric integration of formal social support literature for elderly individuals. *Housing, Care and Support*, 1–17.
- Rahman, M. K., Hossain, M. A., Ismail, N. A., Hossen, M. S., & Sultana, M. (2025). Determinants of students' adoption of AI chatbots in higher education: the moderating role of tech readiness. *Interactive Technology and Smart Education*.
- Rashed, M., Jamadar, Y., Hossen, M. S., Islam, M. F., Thakur, O. A., & Uddin, M. K. (2025). Sustainability catalysts and green growth: Triangulating evidence from EU countries using panel data, MMQR, and CCEMG. *Green Technologies and Sustainability*, 100305.



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium upon the work for non-commercial, provided the original work is properly cited.