International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Development of Machine Learning Algorithm (LVQ) for the detection of Breast Tumor using MATLAB

Mr. Jit Modak[1], Mr. Sukdeb Saha[2], Mr. Chintan Roy[3], Mrs. Poulmi Banerjee[4], Koheli Adhikary[5], Ayan De[6], Aniket Giri[7], Subham Ghosh[8]

[1,2,3,4]Assistant Professor, ECE, Regent Education and Research Foundation Group of Institutions
[5,6,7,8]Student, ECE, Regent Education and Research Foundation Group of Institutions

Abstract
Breast cancer is recognized as the most prevalent malignancy among females and is the second primary cause of cancer-related mortality, following lung cancer. While various genetic codes play a role for the formation of breast cancer. Many techniques are employed to identify breast cancer. This paper presents a MATLAB-based Artificial Neural Network through Learning Vector Quantization algorithm for detecting breast cancer using mammography images. Learning Vector Quantization one of the most pattern recognize method through it, the output unit represent specific class. The weight vector of the output refer as a codebook vector of the particular category. During training process the position should be adjusted with their weight through the supervised learning method. We assumed a set of training pattern with some known data for the purpose of classification of the malignant and benign tumor, where 70% data are uses as training and 30% data are uses as testing purpose("Prof. Dr. R¸Diger Schulz-Wendtland Original owners of database" which specifies that the BIRADS evaluation of mammography's").Some unknown data are collected from the website to find the efficiency of the network.

Keywords: Supervised learning, Learning Vector Quantization, Pattern, Codebook vector.

1. Introduction

Cancer, or carcinoma, is widespread in today's world. As explained by Dr. Mary Ling, a breast and general surgeon, the milk-producing part of the breast is divided into roughly 15 to 20 small sections known as lobes. The ducts are like a bunch of tiny tubes inside the breast that carry milk. These tubes join together and finally bring the milk out through the nipple on the skin. The areola is the dark ring of skin surrounding the nipple, The breast keeps its shape because of connective tissues and ligaments which act like support structures holding it firmly together. The nerves in the breast help you feel touch all kinds of sensations. According to Ann Pietrangelo from the Healthline website, there are some early signs of breast cancer. These include a new lump in the breast or under the arm, swelling in any part of the breast, rashes or irritation around the nipple, pain in the breast, changes in its shape or size, and sometimes fluid coming out of the nipple that is not milk.A lump in the breast can happen when cells st-

art growing in a strange way. There are mainly two types of breast lumps or tumors:
Benign tumors: These are lumps that do not spread to other parts of the body. According to the National Breast Cancer Foundation, doctors sometimes leave them as they are instead of removing them. These tumors are usually not very dangerous, but they can still grow and cause problems. If they start causing pain or trouble, doctors may remove them to make the person feel better.Malignant tumors: These are are cancerous lumps and can be very dangerous because they spread and harm nearby parts of the body. According to the National Breast Cancer Foundation, when doctors find this kind of tumor, they suggest doing a biopsy to check how serious it is. In the body, cancer can start when certain normal genes change into harmful ones, causing cells to grow in a bad way. This can happen for many reasons like family history, gene changes, or pollution. The use of LVQ in breast cancer detection involves training models on datasets like the Wisconsin Breast Cancer Dataset, which contain features derived from digitized images of breast tissue, including radius, texture, and smoothness. LVQ learns prototype patterns for each class and classifies new samples based on similarity, making it effective for medical data analysis. It is valued for its simplicity, interpretability, and low computational cost, though its performance relies on proper parameter tuning and data quality. This study evaluates LVQ's accuracy, strengths, and limitations while comparing it with other machine learning approaches.

2. Architecture

First Layer

Prototype learning: Prototype vectors are acquired by the first layer neurons through the calculation of distances between weight and input vectors, instead of using inner product calculations.
No Normalization Required: Vector normalization is deemed unnecessary for this network.
Equivalent Response: The same output is generated by the network in both situations.
The negative sign is utilized in the distance calculation due to the mechanics of the competitive activation function. Because the winning neuron is intended to be the one in closest proximity, the negative of the distance is calculated.
Consequently, the highest response (1) is assigned to the neuron whose weights most closely match the input vector, and a null response (0) is given by all other neurons
A subclass, not a class, is indicated by the winning neuron. Each class may be constituted by several different neurons or subclasses[2].

Second Layer

Subclasses are merged into a single class within the second layer, utilizing the W^2matrix. Subclasses are represented by the columns, while classes are represented by the rows. A single 1 is contained in each column, with all other elements set to 0, and the specific class of a subclass is indicated by the row in which the 1 occurs. It should be noted that W^2 is treated as a non-trainable parameter; therefore, it cannot be modified during training. The definition of W^2 is determined by the specific problem statement. Furthermore, the total number of neurons is determined by the quantity of classes selected by the user[2].

3. Literature Review

- According to (Qadri Rahmadani.et.al 2025) early detection of breast cancer, a prevalent disease in women, is vital for improving treatment outcomes. This study develops a classification system utilizing MATLAB and Learning Vector Quantization (LVQ) to analyze mammography images sourced from Kaggle. Preprocessing includes grayscale conversion and normalization, followed by texture feature extraction using the Gray Level Co-occurrence Matrix (GLCM) and LVQ model training. The model achieved an accuracy of 80.45%, precision of 78.92%, 100% recall, and an 88.30% F1-score. A MATLAB-based Graphical User Interface (GUI) enables direct image classification. While cancer detection is positive, errors remain in classifying normal images. Future work will focus on data balancing and enhancing model performance for clinical screening applications.

- According to (Lulu Wang.et.al 2024) although X-ray mammography is regarded as the golden standard for breast cancer screening, limitations in sensitivity and specificity are observed. Recent advances in deep learning allow for personalized, improved risk assessment and prognosis. This review examines deep learning-based mammography for improved cancer detection. While the potential for enhanced screening accuracy is evident, clinical implementation challenges exist. Future focus is required on refining algorithms, ensuring data privacy, and enhancing model generalizability to integrate this technology into routine screening, ultimately aiding clinicians in achieving more effective breast imaging diagnostics.
- According to (Ayush Dogra.et.al 2019) breast cancer has become a primary cause of mortality in women worldwide, making early detection vital. Computer-aided diagnosis (CAD) systems have emerged as critical assistive tools that enhance patient care by decreasing false positives and accelerating diagnosis. The automatic detection of breast cancer is facilitated by rapid advancements in sophisticated deep learning-based image processing techniques.This paper presents a brief overview of recent trends in deep learning-based methods for breast cancer detection. In addition to outlining a general framework for a CAD tool, various preprocessing and segmentation techniques are discussed.
- According to (Reem Jalloul.et.al 2023) breast cancer, a prevalent, often fatal malignancy stemming from uncontrolled cell proliferation, is best detected early through medical imaging analysis. This paper reviews the evolution of this field, examining how machine learning and deep learning techniques are applied for detection. Various medical image types and classification systems for tumor, non-tumor, and dense masses are thoroughly detailed using diverse study datasets. Finally, classification challenges and optimal approaches for diagnosing breast cancer are addressed.
- According to (Denis Enachescu.et.al 2005) owing to its low cost and ease of use, electrical impedance spectroscopy, a minimally invasive technique, is favored for characterization of living tissues. Application of this method to breast cancer detection and tissue classification is described in this work. Several classes of excised breast tissue are discriminated using a Learning Vector Quantization (LVQ) network trained on features derived from the spectra. An overall classification efficiency ranging from 77% to 100%, based on parameters, was achieved from a 106-case dataset.
- According to (Chun Huang.et.al 2024) precise cancer medicine is crucially supported by DNA nanotechnology. Tumor-specific biomarkers are detected and therapeutic drug release is controlled through the current application of molecular logic circuits. However, self-learning capabilities for intelligent diagnostics in biological samples are lacking in these systems, and their data processing capabilities are considered limited. A molecular learning vector quantization neural network (LVQNN) model based on DNA strand displacement (DSD) technology for breast tumor diagnosis has been developed here. The molecular LVQNN is recognized to boast powerful computing abilities, handling high-dimensional data for intelligent cancer diagnosis, when compared to previous work. Two distinct typical datasets were selected to verify the feasibility and versatility of the

network: one from a single source with cell morphology data from 569 cases, and a more extensive one spanning different populations and ages, with mRNA gene expression data from 1881 cases. Diagnostic experiments were conducted on 50 and 120 public individuals from these two datasets using the molecular LVQNN, achieving accuracy rates of 94% and 97.5%. The LVQNN model is shown by this study to possess significant advantages in breast cancer diagnosis, as well as to enhance diagnostic accuracy, while new approaches for intelligent cancer diagnosis are introduced, which are anticipated to bring significant breakthroughs and application prospects to precise cancer medicine.

4. Proposed Work

This study utilizes a database originally provided by Rüdiger Schulz-Wendtland, containing BIRADS-based evaluations of mammographic images. The dataset consists of 961 instances, including 516 benign and 445 malignant cases, described using 6 attributes. This dataset serves as the primary reference for training and evaluating the LVQ model in breast cancer classification[16].

Initially, raw data is collected and preprocessed by transforming it into a numerical or binary format, while eliminating inconsistent or irrelevant entries. The cleaned dataset is then stored in Excel and imported into MATLAB for analysis. An LVQ model is developed, where the data is split into training and testing sets. During training, prototype vectors are initialized and continuously updated by moving them closer to correctly classified samples and away from incorrect ones. Model performance is measured using classification accuracy. LVQ is advantageous due to its simplicity, interpretability, and reliable performance in handling diverse input patterns.

For output measurement:-
0000= Benign Tumor
0001=Malignant Tumor
For Input measurement:-
Input1) The outcomes and observations of breast imaging on mammography, ultrasound, and MRI are described by the Breast Imaging Reporting and Data System (BI-RADS
BI-RADS 0 (incomplete): Recommend additional imaging — mammogram or targeted ultrasound
BI-RADS 1 (negative): Routine breast MR screening if cumulative lifetime risk ≥ 20%
BI-RADS 2 (benign): Routine breast MR screening if cumulative lifetime risk ≥ 20%
BI-RADS 3 (probably benign): Short-interval (6-month) follow-up
BI-RADS 4 (suspicious): Tissue diagnosis
BI-RADS 5 (highly suggestive of malignancy): Tissue diagnosis
BI-RADS 6 (known biopsy-proven malignancy): Surgical excision when clinically appropriate.[17]
BI-RADS 0= 0000
BI-RADS 1= 0001
BI-RADS 2= 0010
BI-RADS 3= 0011
BI-RADS 4= 0100
BI-RADS 5= 0101
BI-RADS 6= 0110
Input2) Shape (Mass shape, categorical):
round=0001

Oval=0011
irregular oval=0010
lobular =0100
Input3) Margin: mass margin (categorical):
circumscribed=0001
microlobulated=0010
obscured=0011
ill-defined=0100
spiculated=0101
Input4) Density: mass density (ordinal)
High=0001
Iso=0010
Low=0011
Fat-containing=0100[17]

5. Equations

$$n_i^1 = -\left\| {}_i\mathbf{w}^1 - \mathbf{p} \right\| ,$$

n^1 is the input to the activation function
w^1 is the weight matrix of layer 1
p is the input to the network
The subscript 'i' tells the particular neuron

$${}_{i^*}\mathbf{w}^1(q) = {}_{i^*}\mathbf{w}^1(q-1) + \alpha(\mathbf{p}(q) - {}_{i^*}\mathbf{w}^1(q-1)) \text{, if } a^2_{k^*} = t_{k^*} = 1$$

$${}_{i^*}\mathbf{w}^1(q) = {}_{i^*}\mathbf{w}^1(q-1) - \alpha(\mathbf{p}(q) - {}_{i^*}\mathbf{w}^1(q-1)) \text{, if } a^2_{k^*} = 1 \neq t_{k^*} = 0$$

Kohonen Rule

'*' here represents the winning neuron
't' is the target vector for training[2].

6. Result

Training and Evaluation Results: The LVQ network was trained using a preprocessed dataset. During the training phase, prototype vectors were continuously updated according to input patterns. These vectors moved closer to correctly classified samples and shifted away from misclassified ones, enabling the model to gradually learn the underlying data structure. After several iterations, the network successfully developed a clear distinction between the two target classes: benign and malignant tumors. The training process resulted in stable prototype vectors that effectively represent each class. For performance evaluation, a separate test dataset was used. The model's predicted outputs were compared with the actual tumor labels to measure classification accuracy. The LVQ network achieved an accuracy ~94% demonstrating its strong capability in correctly identifying breast cancer cases. This result indicates that the model performs reliably in distinguishing between benign and malignant tumors and is effective for medical classification tasks based on pattern recognition.

7. Conclusion

The LVQ model effectively classified tumors as benign or malignant using selected imaging features, demonstrating its suitability for medical diagnostic applications. Its high classification accuracy highlights the potential of LVQ in supporting clinical decision-making processes. The results indicate that prototype-based learning can successfully capture patterns in breast cancer data. Future work can further improve performance by using larger and more diverse datasets, incorporating additional relevant features, or comparing LVQ with advanced machine learning and neural network models to enhance diagnostic accuracy and reliability.

8. Figures

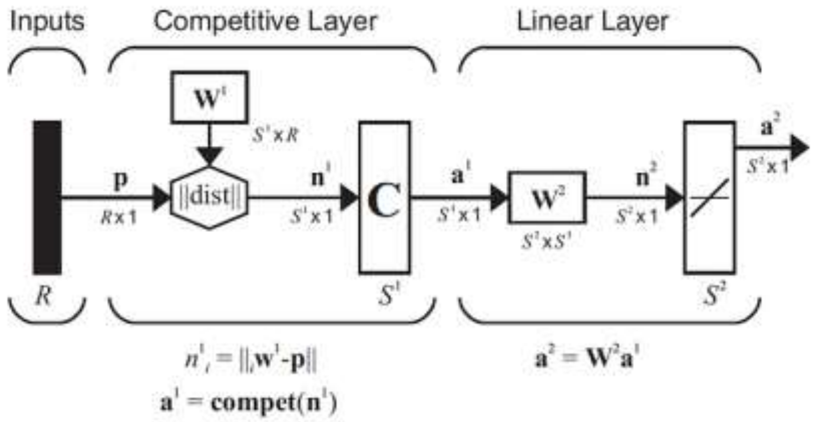

Fig.1 LVQ Network[2]

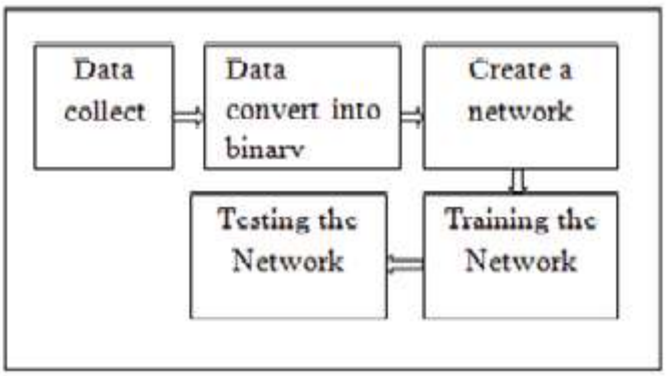

Fig.2 Block Diagram of the work[16}

9. Authors' Biography

We would like to sincerely thank all those who contributed to the successful completion of this research. Our special gratitude goes to the faculty members of the Department of Electronics and Communication Engineering, Regent Education and Research Foundation Group of Institutions, for their constant guidance, encouragement, and valuable suggestions throughout this study. We also acknowledge their support in providing access to essential computational facilities and MATLAB resources, which were crucial for developing and implementing the neural network model. Lastly, we express our appreciation to the researchers whose previous work in neural networks and medical diagnosis served as the foundation for this study.

.

10. References

1. S.Vijaya , Dr D. Gladis. (International Journal of Pure and Applied Mathematics(2018)Volume 119 No. 16 2018, 365-381) Breast Cancer classification using ART1 neural network with Improved F-Score feature reduction.
2. https://medium.com/@udbhavkush4/demystifying-learning-vector-quantization-a-step-by-step-guide-with-code-implementation-from-ea3c4ab5330e
3. Lilian C. Wang , Megan Sullivan, Hongyan Du, Marina I. Feldman, Ellen B. Mendelson(2013). US Appearance of Ductal Carcinoma in Situ. RadioGraphics, Vol. 33,No. 1.
4. Choudhary, T., et al. (2019). Breast mass classification using machine learning techniques. International Journal of Computer Applications, 182(46).
5. Elmore, J. G., et al. (2015). Variability in interpretive performance of screening mammography and radiologists' use of BI-RADS. Annals of Internal Medicine, 162(10), 673–681.
6. Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. E. (2000). Gradient and texture analysis for the classification of mammographic masses. IEEE Transactions on Medical Imaging, 19(10), 1032–1043.
7. Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991). ART 2-A: An adaptive resonance algorithm for rapid category learning. Neural Networks, 4(4), 493–504.
8. Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. Expert Systems with Applications, 46, 139–144.
9. Carpenter, G. A., & Grossberg, S. (1987). ART 1: Self-organization and pattern recognition with a stability–plasticity tradeoff. Biological Cybernetics, 56(2–3), 77–93.
10. Jiang, Y., et al. (2007). Breast cancer diagnosis with neural networks. Cancer Letters, 245(1–2), 90–98.
11. Jieun Koh, MD and Min Jung Kim, MD, PhD(2018); Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage.
12. Adam Mracko , Lucia Vanov canová and Ivan Cimrák (Journal of Imaging(2023)) .Mammography Datasets for Neural Networks—Survey
13. Mohammed Amine Naji , Sanaa El Filali , Kawtar Aarika , EL Habib Benlahmar , Rachida Ait Abdelouhahid , Olivier Debauche(Procedia Computer Science Volume 191, 2021, Pages 487-492). Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis
14. Yap, M. H., et al. (2018). Automated breast ultrasound lesions detection using convolutional neural networks. IEEE Journal of Biomedical and Health Informatics, 22(4), 1218–1226.
15. American cancer society(2019).Newer and Experimental Breast Imaging Tests https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_adaptive_resonance_theory.htm
16. https://radiology.ucsf.edu/blog/bi-rads-and-breast-mr-how-well-are-we-managing-patients-based-established-standards.

17. Poulmi Banerjee , Chintan Roy , Munmoon Chaki, Jit Modak(2026) Development of a Neural Network using MATLAB for the detection of Breast Cancer Using ART 1 Method, Advanced International Journal for Research (AIJFR).
18. https://www.kaggle.com/datasets/hayder17/breast-cancer-detection
19. Qadri Rahmadani, Romi Mulyadi, Image Analysis For Breast Cancer Classification Using Learning Vector Quantization (LVQ) Method(2025), Journal of Engineering Science and Technology Management
20. Lulu Wang, Mammography with deep learning for breast cancer detection(2024), Volume 14 - 2024 | https://doi.org/10.3389/fonc.2024.1281922
21. Ayush Dogra, Kaushik K, Bhawna Goyal, A Brief Review of Breast Cancer Detection via Computer Aided Deep Learning Methods(2019), IJERT, Volume 08, Issue 12 (December 2019).
22. Reem Jalloul, H. K. Chethan, Ramez Alkhatib, A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images (2023), Diagnostics, № 14, p. 2460.