An improved Computer aided diagnosis of breast cancer detection using mammogram and machine learning algorithm

Ms.J. Sinjini[1], Ms.Nivethitha T[2], Ms. Kavitha N[3], Dr.A. Athiraja[4*]
Assistant Professor[1,2,3,4]
Malla Reddy Engineering College, Hyderabad, Telangana[1,]
Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu[2,3]
Bannari Amman Institute of Technology, Erode, Tamilnadu[4]
Corresponding author:Dr.A. Athiraja

ABSTRACT

The apparent structure is examined using computer-aided diagnostic (CAD) technologies. The majority of existing diagnostic approaches rely on mammography mass features, however the suggested strategy relies on microcalcification (MC) characteristics. Cancer cells are detected utilising mammography images, pre-processing methods, segmentation, classification, fuzzy logic, and Receiver Operating Characteristics (ROC) curve analysis in this research. The National Institute of Cancer provides mammogram scans. Normalization and median filtering were used as part of the pre-processing procedure. Segmentation is used to detect the location of microcalcification present cells using the local threshold approach. The Perceptron algorithm and the Case Based Reasoning (CBR) approach are two classification techniques. The microcalcification present cells and microcalcification missing cells are classified using the perceptron technique. CBR is used to divide microcalcification-affected cells into the following categories: initial, very small, small, medium, high, and very high. Fuzzy logic is utilised in decision-making. ROC curve analysis was used to assess system performance.

Key Words: CAD, MC, ROC, CBR, Perceptron, Mammogram

INTRODUCTION

Breast cancer is a cancer that develops in the breast tissue, most usually in the inner lining of milk ducts or the lobules that feed milk to the ducts [1]. Cancers that start in ducts are called ductal carcinomas, whereas cancers that start in lobules are called lobular carcinomas. Humans and other species both get breast cancer. While female breast cancer accounts for the vast majority of human occurrences, male breast cancer can also develop [2].

This study uses microcalcifications in mammography images to diagnose cancer cells. Using the Perceptron machine learning technique, categorise MC present cells and MC absence cells. Following that, using Case Based Reasoning, the MC present cell is divided into the following classes: initial, very small, small, medium, high, and very high (CBR). Fuzzy logic is utilised to make decisions and ROC curve analysis is used to assess system performance.

DESIGN AND METHODOLOGY

The various modules ofComputer aided diagnosis of cancer is detailed in figure 1

- Mammogram image
- Normalization
- Median filter
- Segmentation
- Perceptron algorithm
- Case Based Reasoning
- ROC Curve analysis

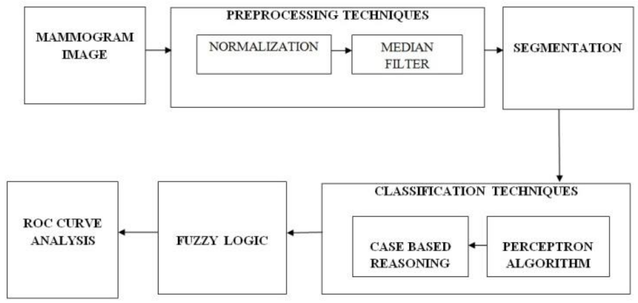

Figure 1 Block diagram for cancer diagnosis

MAMMOGRAPHY IMAGE

Mammography is a diagnostic and screening procedure that uses low-energy X-rays (typically approximately 30 kVp) to inspect the human breast [17]. Mammography is used to diagnose breast cancer in its early stages, usually by detecting distinctive lumps and/or microcalcifications. The picture below is an example of a mammography image (2)

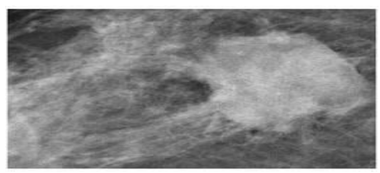

Figure 2 Mammography Image

PREPROCESSING

The majority of mammography pictures are big and high resolution, necessitating the need of specialist computing facilities for effective processing [19]. Image compression methods are commonly used to make it easier to send these pictures over computer networks.A grayscale image has simply the intensity information, whereas a colour image contains both colour and intensity information. Knowing the intensity of each colour in a pixel may be used to determine the overall intensity of the pixel [20]. To accomplish so, the RGB values of each pixel in the picture must first be retrieved. Then their average is calculated, yielding the pixel's overall intensity. This procedure is done for each of the image's pixels. A gray-level mammography picture is created from an RGB mammography image. The RGB picture size is (192 262), whereas the grey level image size is (192 262), as illustrated in the accompanying figure (3).

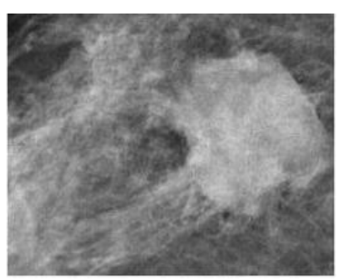

Fig 3 Normalization output

The median filter is a type of nonlinear digital filter that is commonly used to eliminate noise. A typical pre-processing procedure to improve the outcomes of later processing is noise reduction (for example, edge detection on an image) [21]. Median filtering is commonly employed in digital image processing because it retains edges while reducing noise under specific conditions.

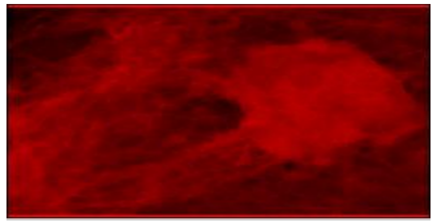

Fig 4 Median Filter Output

The median filter's window size is [22] in this case. This [23] represents the median filter's height and breadth; utilising this window size does not modify the mammography picture; it only removes the noise. After using the median filter, the image looks like this figure (4)

PERCEPTRON ALGORITHM

The Perceptron is a supervised classification technique for an input into one of two potential outputs in computational geometry. It's a linear classifier, which means it generates predictions based on a linear predictor function that combines a set of weights with the feature vector characterising a specific input [25]. The perceptron learning algorithm is an online algorithm, meaning it processes components in the training set one by one [26].

CASE BASED REASONING:

CBR, in its broadest sense, is the process of addressing new issues using the answers to comparable previous problems [29]. A common type of analogy building is case-based reasoning. For the sake of computer reasoning, case-based reasoning has been codified as a four-step process:**Retrieve, Reuse, Revise, Retain**.

The CBR can be classify the MC present cells into following classes initial, very small, small, medium, high and very high cells based on the following table(4.2)

Pixel value	**Class label**
180-190	Initial
191-200	Very small
201-220	Small
221-230	Medium
231-240	High
241-255	Very high

Table 2 – CBR classification table.

FUZZY LOGIC:

Fuzzy logic is a type of many-valued or probabilistic logic that deals with approximate rather than fixed and accurate reasoning [30]. Unlike typical binary sets (where variables can have true or false values), fuzzy logic variables can have a truth value that varies from 0 to 1. The idea of partial truth has been expanded to fuzzy logic, where the truth value can vary from totally true to completely false.

Fuzzy operators on fuzzy sets are defined by fuzzy set theory. The difficulty with this is that the necessary fuzzy operator could not be available [31]. As a result, fuzzy logic frequently employs IF-THEN rules or analogous constructions, such as fuzzy associative matrices.

ROC CURVE

A receiver operating characteristic (ROC) curve, or simply ROC curve, is a graphical representation of a binary classifier system's performance while its discrimination threshold is modified. It's made by graphing the proportion of true positives out of positives (TPR = true positive rate) vs. the fraction of false positives out of negatives (FPR = false positive rate) at different thresholds [32]. TPR stands for sensitivity, while FPR stands for one less than the specificity, or true negative rate.

PERFORMANCE ANALYSIS

Table 1 shows how the proposed system compares to existing ANN (Artificial Neural Network) classification methods in terms of performance. Table 5 shows the results of the performance study.

Database	ANN accuracy	CBR accuracy
Database 1	83%	97%
Database 2	85%	96%
Database 3	79%	93%
Database 4	86%	94%
Database 5	81%	97%
Database 6	75%	92%
Database 7	88%	98%

Table 5 performance analysis

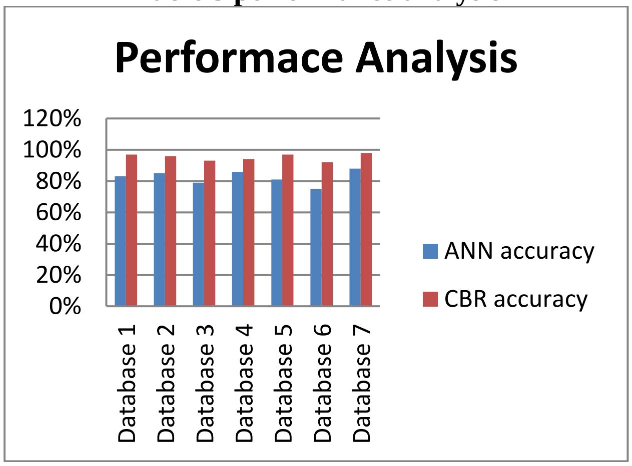

Figure 8 performance analyses by accuracy

Table 7 Performance analysis of Classification Method

Classification Techniques	**SEN %**	**SPEC**	**F1 Score**	**Acc %**
ID3	74	75	78	73
ANN	78	76	75	75
SVM	81	82	85	84
Perceptron	91	90	92	91
CBR	95	94	96	95

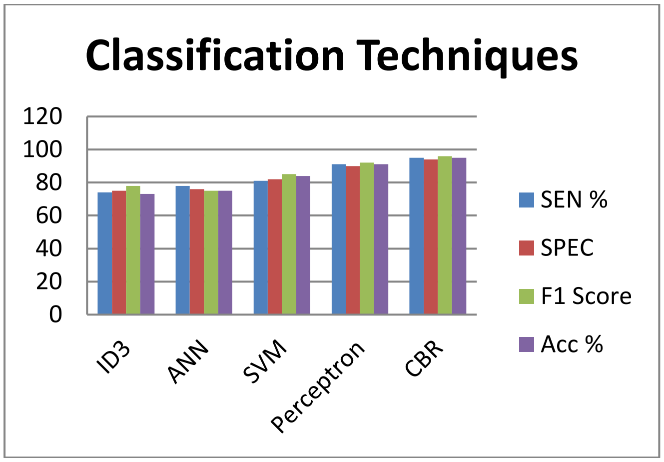

Figure 10 Performance analysis of Classification Method

The following characteristics were used to evaluate the classification method's performance: sensitivity, specificity, F1 score, and accuracy. We deduced from table 7 and figure 10 that CBR has greater sensitivity, specificity, F1 score, and accuracy values of 95 percent.

Table 8 Performance analysis based on Execution Time

Data set	Execution Time (seconds)				
	ID3	ANN	SVM	Perceptron	CBR
Data set 1	190	185	178	176	112
Data set 2	192	184	182	165	132
Data set 3	192	186	179	163	124
Data set 4	191	183	185	162	115
Data set 5	196	185	176	168	116
Data set 6	193	186	175	167	118
Data set 7	192	183	179	165	113

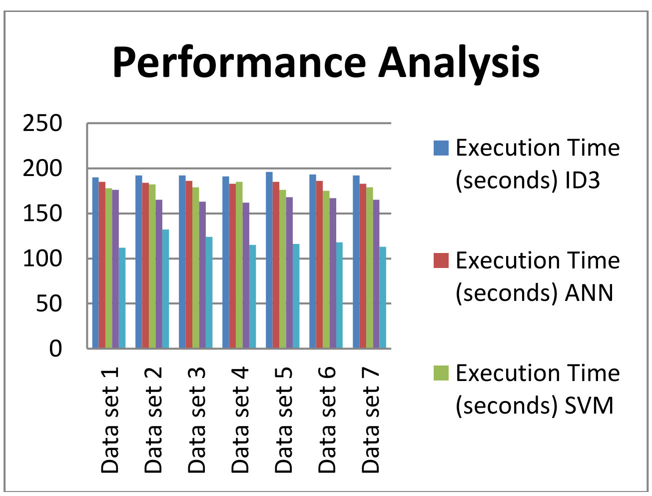

Figure 11 Performance analysis based on Execution Time

In terms of execution time, Table 8 and Figure 11 show the overall performance of many classification algorithms used for face recognition classification. When compared to other categorization algorithms, CBR required less time to perform.

Table 9 Performance analysis based on average number of iteration

Data set	Average Number of iteration				
	ID3	ANN	SVM	Perceptron	CBR
Data set 1	602	421	324	152	56
Data set 2	630	452	362	165	26
Data set 3	625	468	358	195	35
Data set 4	678	495	321	184	95
Data set 5	624	476	389	175	76
Data set 6	638	421	374	173	42
Data set 7	621	435	329	123	46

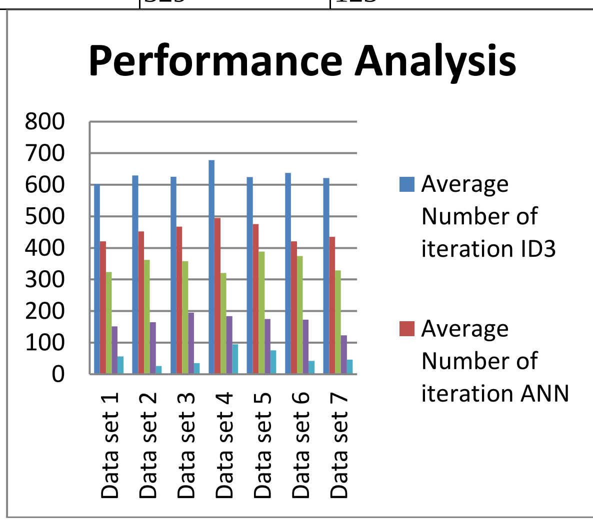

Figure 12 Performance analysis based on average number of iteration

Table 9 and Figure 12 illustrate the average number of iteration performance measurements used. According to a comparison of categorization approaches, CBR had the fewest repetitions of all the systems.

Table 12 Performance analysis based on error rate

Data set	Error Rate				
	ID3	ANN	SVM	Perceptron	CBR
Data set 1	0.321	0.213	0.152	0.052	0.016
Data set 2	0.345	0.265	0.164	0.058	0.012
Data set 3	0.328	0.248	0.174	0.056	0.017
Data set 4	0.324	0.218	0.191	0.057	0.011
Data set 5	0.389	0.298	0.182	0.053	0.015
Data set 6	0.347	0.276	0.172	0.057	0.013
Data set 7	0.365	0.256	0.163	0.051	0.014

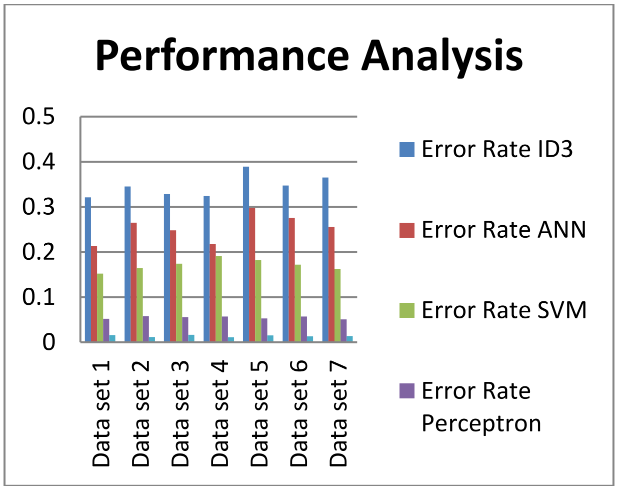

Figure 15 Performance analysis based on error rate

The overall effectiveness of several classification techniques used for categorization of data sets is shown in Table 12 and Figure 15. In a comparison of classification algorithms, CBR had the lowest error rate of all the systems tested.

CONCLUSION

Our technique for detecting microcalcification in mammograms. Using the Perceptron learning technique, locate microcalcification cells and microcalcification absence cells. This approach makes it simple to distinguish between microcalcification-positive and microcalcification-negative cells. The perceptron learning algorithm is a supervised learning classification approach in which the output value is anticipated, eliminating the possibility of misclassifying the microcalcification present and missing cells. Because the Perceptron learning process is a supervised machine learning approach, the actual and expected outputs were always the same. Based on the microcalcification analysis, if the microcalcification was found in the mammography, the patient was diagnosed with cancer. CBR was used to divide MC present cells into the following categories: beginning, small, medium, high, and very high. It is impossible to anticipate whether or not a patient will be afflicted by cancer. If a patient is diagnosed with cancer, antibiotics or surgery may be recommended. The fuzzy set strategy is used to make decisions. In comparison to the prior diagnosing procedure, this method performs better.

REFERENCES

1. H. Zhang et al., "Multi-Source Transfer Learning Via Multi-Kernel Support Vector Machine Plus for B-Mode Ultrasound-Based Computer-Aided Diagnosis of Liver Cancers," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 10, pp. 3874-3885, Oct. 2021, doi: 10.1109/JBHI.2021.3073812..
2. C. Chen, Y. Wang, J. Niu, X. Liu, Q. Li and X. Gong, "Domain Knowledge Powered Deep Learning for Breast Cancer Diagnosis Based on Contrast-Enhanced Ultrasound Videos," in IEEE Transactions on Medical Imaging, vol. 40, no. 9, pp. 2439-2451, Sept. 2021, doi: 10.1109/TMI.2021.3078370..
3. Y. Wang, Z. Ma, K. -C. Wong and X. Li, "Evolving Multiobjective Cancer Subtype Diagnosis From Cancer Gene Expression Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 6, pp. 2431-2444, 1 Nov.-Dec. 2021, doi: 10.1109/TCBB.2020.2974953..
4. X. Tang, L. Cai, Y. Meng, C. Gu, J. Yang and J. Yang, "A Novel Hybrid Feature Selection and Ensemble Learning Framework for Unbalanced Cancer Data Diagnosis With Transcriptome and

Functional Proteomic," in IEEE Access, vol. 9, pp. 51659-51668, 2021, doi: 10.1109/ACCESS.2021.3070428..
5. M. Li et al., "Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images," in IEEE Access, vol. 9, pp. 53687-53707, 2021, doi: 10.1109/ACCESS.2021.3071057.
6. Y. -H. Zhang et al., "Screening Dys-Methylation Genes and Rules for Cancer Diagnosis by Using the Pan-Cancer Study," in IEEE Access, vol. 8, pp. 489-501, 2020, doi: 10.1109/ACCESS.2019.2961402.
7. Z. Li et al., "Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide Histopathology Images—The ACDC@LungHP Challenge 2019," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 2, pp. 429-440, Feb. 2021, doi: 10.1109/JBHI.2020.3039741.
8. Z. Huang and D. Chen, "A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm," in IEEE Access, vol. 10, pp. 3284-3293, 2022, doi: 10.1109/ACCESS.2021.3139595.
9. Y. Li, F. Zhang and C. Xing, "Screening of Pathogenic Genes for Colorectal Cancer and Deep Learning in the Diagnosis of Colorectal Cancer," in IEEE Access, vol. 8, pp. 114916-114929, 2020, doi: 10.1109/ACCESS.2020.3003999.
10. X. Wang et al., "Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis," in IEEE Transactions on Cybernetics, vol. 50, no. 9, pp. 3950-3962, Sept. 2020, doi: 10.1109/TCYB.2019.2935141.
11. Y. Li et al., "Computer-Aided Cervical Cancer Diagnosis Using Time-Lapsed Colposcopic Images," in IEEE Transactions on Medical Imaging, vol. 39, no. 11, pp. 3403-3415, Nov. 2020, doi: 10.1109/TMI.2020.2994778.
12. S. M. S. Hasan, D. Rivera, X. -C. Wu, E. B. Durbin, J. B. Christian and G. Tourassi, "Knowledge Graph-Enabled Cancer Data Analytics," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, pp. 1952-1967, July 2020, doi: 10.1109/JBHI.2020.2990797.
13. D. Florensa et al., "The Use of Multiple Correspondence Analysis to Explore Associations Between Categories of Qualitative Variables and Cancer Incidence," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3659-3667, Sept. 2021, doi: 10.1109/JBHI.2021.3073605.
14. G. Li, C. Li, G. Wu, D. Ji and H. Zhang, "Multi-View Attention-Guided Multiple Instance Detection Network for Interpretable Breast Cancer Histopathological Image Diagnosis," in IEEE Access, vol. 9, pp. 79671-79684, 2021, doi: 10.1109/ACCESS.2021.3084360.
15. Q. Zhou, B. Yong, Q. Lv, J. Shen and X. Wang, "Deep Autoencoder for Mass Spectrometry Feature Learning and Cancer Detection," in IEEE Access, vol. 8, pp. 45156-45166, 2020, doi: 10.1109/ACCESS.2020.2977680.
16. P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong and H. Zheng, "Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 1, pp. 148-160, Jan. 2021, doi: 10.1109/TBME.2020.2993278.
17. G. Huang et al., "Channel-Attention U-Net: Channel Attention Mechanism for Semantic Segmentation of Esophagus and Esophageal Cancer," in IEEE Access, vol. 8, pp. 122798-122810, 2020, doi: 10.1109/ACCESS.2020.3007719.
18. G. Alexandrou et al., "Detection of Multiple Breast Cancer ESR1 Mutations on an ISFET Based Lab-on-Chip Platform," in IEEE Transactions on Biomedical Circuits and Systems, vol. 15, no. 3, pp. 380-389, June 2021, doi: 10.1109/TBCAS.2021.3094464.
19. L. Duran-Lopez, J. P. Dominguez-Morales, A. F. Conde-Martin, S. Vicente-Diaz and A. Linares-Barranco, "PROMETEO: A CNN-Based Computer-Aided Diagnosis System for WSI Prostate Cancer Detection," in IEEE Access, vol. 8, pp. 128613-128628, 2020, doi: 10.1109/ACCESS.2020.3008868.

20. A. Masood et al., "Cloud-Based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 8, pp. 1-13, 2020, Art no. 4300113, doi: 10.1109/JTEHM.2019.2955458..