International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET

Roi-Centric Texture Feature Extraction and Neural Network Classification for Mammography-Based Cancer Detection

[#1]Keerthana. V, [#2]Reikha C

[#1]PG Scholar, [#2]Associate Professor

[#1, #2]Department of Computer Science and Engineering, Krishnasamy College of Engineering and Technology, Cuddalore, Tamil Nadu, India

ARTICLEINFO

Article History :

Accepted: 01 Feb 2024
Published: 05 Feb 2024

Publication Issue :
Volume 11, Issue 1
January-February-2024
Page Number :
165-170

ABSTRACT

Breast cancer stands as a leading cause of mortality among women, emphasizing the critical need for advanced diagnostic tools to enhance early detection. Despite extensive research efforts in image processing and classification techniques, the elusive nature of its causation renders prevention unattainable. Thus, early detection remains pivotal for effective intervention.

Computer-Aided Diagnosis (CAD) applied to mammographic images emerges as a promising and efficient approach for breast cancer diagnosis. By leveraging CAD techniques, accurate identification and subsequent reduction of mortality rates associated with breast cancer become achievable. Notably, masses and microcalcification clusters serve as crucial early indicators, offering insights into potential breast malignancies in their nascent stages. This study draws upon the DDSM Database (Digital Database for Screening Mammography), a globally utilized resource containing around 3000 cases, to analyze texture features for cancer detection. Focusing on Regions of Interest (ROI) within mammograms, the research quantitatively characterizes microcalcifications as benign, normal, or malignant. Employing Principal Component Analysis (PCA), the extracted features are further streamlined to enhance the identification of masses within the mammogram. The refined texture features undergo meticulous comparison and are processed through a Feed-Forward Algorithm employing Neural Networks. This methodology facilitates a deeper understanding of the cancerous patterns present within mammography images, aiding in precise and timely diagnosis.

Keywords : DDSM, Digital Database for Screening Mammography, Regions of Interest, Principal Component Analysis, Computer-Aided Diagnosis

I. INTRODUCTION

Breast cancer remains a profound health concern globally, posing a significant threat to women's lives. Despite considerable advancements in medical technology and research efforts, its prevalence and the challenges associated with early detection persist. The imperative to develop robust and efficient diagnostic methodologies that facilitate early-stage detection and subsequent intervention is paramount. In this pursuit, Computer-Aided Diagnosis (CAD) emerges as a promising avenue for enhancing the accuracy and efficiency of breast cancer detection.

Mammography, a cornerstone in breast cancer screening, offers unparalleled insights into the breast tissue's composition, allowing for the identification of anomalies that might signify malignancy. The quest for accurate diagnosis pivots on the extraction of relevant and discriminative features from mammographic images. Among the multitude of techniques, Region of Interest (ROI)-centric texture feature extraction, coupled with Neural Network classification, stands as a formidable approach to enhance the precision of cancer detection.

The efficacy of any breast cancer detection system relies heavily on the discriminatory power of the extracted features. Thus, in this study, we focus on ROIs within mammography images, recognizing the significance of these regions in identifying potential indicators of breast malignancies. Masses and microcalcifications, recognized as early symptoms of breast cancer, serve as crucial markers for early detection.

Our work utilizes the extensive and widely employed DDSM Database (Digital Database for Screening Mammography), a repository of approximately 3000 cases, revered in the realm of cancer research. Leveraging this resource, we endeavor to establish a comprehensive framework that not only captures these intricate textures but also endeavors to classify them into categories such as harmless, ordinary, and threatening.

The texture features extracted from ROIs hold immense potential in unraveling the subtle nuances of breast tissue anomalies, including the differentiation between benign and malignant masses, and the characterization of microcalcifications. However, the sheer volume and complexity of these features often pose challenges in deriving actionable insights. To address this, we employ Principal Component Analysis (PCA) as a means to distill and highlight the most discriminative features, streamlining the identification of masses within mammograms.

The subsequent step involves harnessing the power of Neural Networks, specifically employing a Feed-Forward Algorithm, to process and classify these refined texture features. This approach facilitates a deeper understanding of the underlying patterns associated with breast cancer within mammographic images, thereby enabling precise and early detection.

In the forthcoming sections, we delve into the methodology, detailing the techniques employed in texture feature extraction from ROIs, the nuances of PCA for feature reduction, and the intricacies of the Neural Network classification process. Through this rigorous exploration, we aim to present a holistic and effective approach that amalgamates image analysis and machine learning for enhanced mammography-based breast cancer detection.

II. RELATED WORKS

Basha et al., utilizes Morphological operators to isolate and enhance distinct features of masses and micro calcifications present in mammography images. These operators help in highlighting and delineating subtle variations in density and structure, aiding in the segmentation process. Following this, the Fuzzy c-means clustering algorithm is employed, leveraging intensity-based segmentation to further differentiate these identified regions from the background tissue. By combining Morphological operators and Fuzzy c-means clustering, this method achieves effective delineation and isolation of potential cancerous tumor masses, crucial for accurate diagnosis in mammography-based breast cancer detection.

Das et al., involves a comparative analysis of breast tissue types—normal ductal epithelial cells, ductal/lobular invasive carcinogenic cells—via image processing techniques. Features specific to invasive cancerous tissue are extracted and contrasted with normal tissue characteristics. Additionally, the paper proposes a breast cancer recognition technique utilizing image processing methodologies. Furthermore, it suggests potential preventive measures by targeting p53 gene mutation as a means to mitigate the risk of breast cancer development.

Kekre et al., introduces a vector quantization segmentation technique tailored for the identification of cancerous masses in mammogram images. As an alternative to conventional mammography, contrast-enhanced magnetic resonance imaging of the breast is highlighted. The paper focuses on enhancing radiologists' diagnostic capabilities by leveraging computer-aided diagnosis (CAD) schemes. Specifically, the method aims to improve the detection of primary indicators of breast cancer, namely masses and microcalcifications, through the proposed segmentation approach using vector quantization.

Palcic et al., introduces a non-invasive approach for imaging early-stage lung cancer, bypassing the need for investigative drugs. It relies on exploiting spectral variations in autofluorescence between normal and cancerous lung tissue. By capturing fluorescence images at distinct spectral bands and amplifying them through an image-intensified camera, a sensitive and detailed representation of the tissue is obtained. These images undergo digitization via mathematical transformation to generate a pseudo image. Displayed on a color video monitor, this pseudo image effectively highlights and delineates the cancerous area, offering a potential breakthrough in early lung cancer detection without the requirement for exogenous fluorescent markers.

Nawgaje et al., introduces a robust and automated segmentation approach for mammographic images using kernel-based fuzzy c-means (FCM) clustering. By leveraging regional features specific to various breast densities, the proposed technique efficiently captures these characteristics through appropriate kernel utilization. The methodology employs kernel-based FCM in a folded manner to process significant features concurrently, allowing for the identification of masses with reduced blocking effects. This approach outperforms other clustering-based segmentation methods by effectively handling uncertainties in mammograms, demonstrating superior performance both qualitatively and quantitatively. Moreover, the kernel-based strategy reduces data points, enhancing convergence speed compared to conventional algorithms, and illustrates the impact of image size on segmentation convergence rates.

Al-Tarawneh et al., focuses on medical image enhancement and assessment, crucial for early detection and treatment, especially in cancer diagnosis like lung and breast cancer. Employing Gabor filter within Gaussian rules enables low pre-processing, enhancing image quality. Subsequent segmentation produces an enhanced region of the object of interest, forming the basis for feature extraction. The research emphasizes normality comparison using key features: pixel percentage and mask labeling, facilitating accurate comparisons between images for improved diagnosis and assessment of abnormalities in medical images.

Kanungo et al., introduces GA-ACO-FCM, a novel hybrid segmentation technique designed to improve the accuracy of breast cancer mass detection in mammogram images. Leveraging genetic algorithm (GA) and ant colony optimization (ACO) with Fuzzy C-means (FCM), this method addresses issues like initialization problems and sensitivity to noise present in traditional FCM-based segmentation methods. By integrating optimization tools, GA-ACO-FCM enhances segmentation accuracy without modifying the core objective function of FCM. This innovative approach demonstrates superior performance, overcoming initialization challenges and achieving heightened accuracy in detecting breast cancer masses within mammogram images.

III. PROPOSED MODEL

The significance of early breast cancer detection necessitates effective screening methods, with X-ray mammography being the foremost diagnostic tool. However, interpreting mammograms remains challenging for radiologists, leading to inconsistencies in results. To address this, computer-aided diagnosis (CAD) schemes have emerged, aiming to augment the identification of primary indicators of breast cancer.

Digital mammography images, displayed on computer monitors, undergo image enhancement processes, including adjustments in brightness or darkness before printing on film. Image processing techniques play a pivotal role in medical settings, enhancing images for early detection and treatment of abnormalities in various cancer types like breast cancer and lung cancer. These techniques are vital due to the critical time factor involved in identifying anomalies within target images.

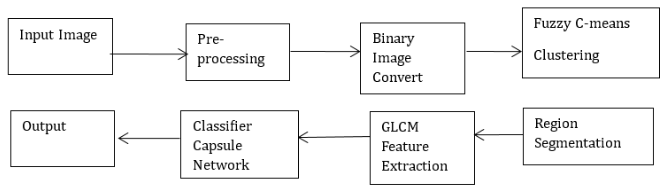


Figure 1: Overall Architecture

The paper introduces a novel method for classifying texture features utilizing neural networks, focusing on achieving highly accurate texture feature generation by delineating Regions of Interest (ROI) within mammography images. These selected ROIs are employed for extracting detailed texture features crucial for the classification process. The extracted features are then input into a neural network tasked with categorizing images as either cancerous or non-cancerous. Training the neural network involves utilizing a feed-forward algorithm for weight adjustments, optimizing its ability to accurately classify mammography images based on the identified texture features.

The data is taken from the dataset. Prepare the data to use the breast cancer dataset by loading it. Separate the information into features (X) and labels (y). For quantitative analysis of features, ensure each feature has the same effect on the algorithms by normalizing or standardizing them. Separate the data into a training set and a test set. Seventy percent for instruction and thirty percent for testing is an example of a frequent breakdown. For the evaluation and training of models, random forest (RF), decision tree (DT), k-nearest neighbors (KNN), logistic regression (LR) and linear support vector classifier (linear SVC) are used and the effectiveness of each algorithm is evaluated using the criteria provided. Select the most effective algorithm for spotting breast cancer. To determine which attributes are most important for algorithms such as RF and DT, perform a feature importance analysis. To better understand the algorithm's decision-making process, decision trees may be represented graphically.

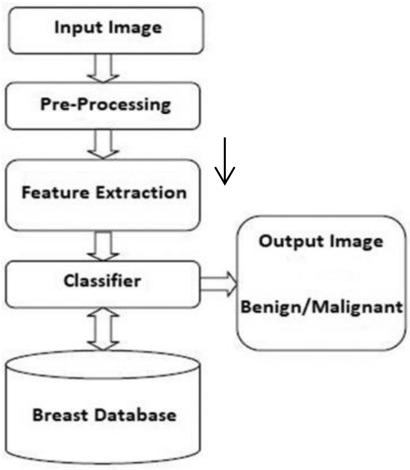


Figure 2: Flow of Proposed Model

Pre-processing: Pre-processing involves preparing the input data or images for further analysis. This can include operations like resizing, noise reduction, normalization, or enhancing the images to improve their quality and suitability for subsequent algorithms.

Binary Image Conversion: Binary image conversion is a process where an image is converted into a binary form, often representing objects as black and the background as white. This simplifies subsequent analysis, especially in object detection or segmentation tasks.

Image Clustering Technique: Image clustering involves grouping similar pixels or regions together based on their characteristics such as color, intensity, or texture. Clustering techniques like K-means, Fuzzy C-means, or hierarchical clustering can be applied to partition the image into distinct groups or clusters.

Segmentation: Image segmentation divides an image into meaningful regions or segments to simplify its representation and facilitate analysis. Segmentation can be based on various attributes like color, texture, intensity, or edges, separating objects of interest from the background.

Non-contextual Thresholding: Non-contextual thresholding is a method where a global threshold value is applied to an image to separate foreground objects from the background. Pixels above a certain intensity threshold are classified as one type (e.g., object) while those below are classified as another (e.g., background).

Simple Thresholding: Simple thresholding is a basic method that categorizes pixels in an image based on a fixed threshold value. Pixels are set to one value if their intensity is above the threshold and another value if below it.

Gray Level Co-occurrence Matrix (GLCM): GLCM is a statistical method used for texture analysis in images. It calculates the occurrences of pairs of pixel values at various spatial relationships, providing information about texture features like contrast, homogeneity, and entropy.

Feature Extraction: Feature extraction involves extracting relevant information or features from images that are essential for subsequent analysis or classification. These features could include texture, shape, color, or statistical measures derived from the image data.

Classifier (Feed Forward): The feed-forward classifier is a type of neural network where information moves in one direction—from input to output layer—without cycles or loops. It's commonly used in pattern recognition or classification tasks, taking extracted features as input and making predictions or classifications based on learned patterns.

Each of these modules plays a crucial role in image analysis and processing pipelines, contributing to tasks like segmentation, feature extraction, and classification in the context of image-based systems and applications. These processes are fundamental in tasks like medical imaging analysis, object detection, or pattern recognition.

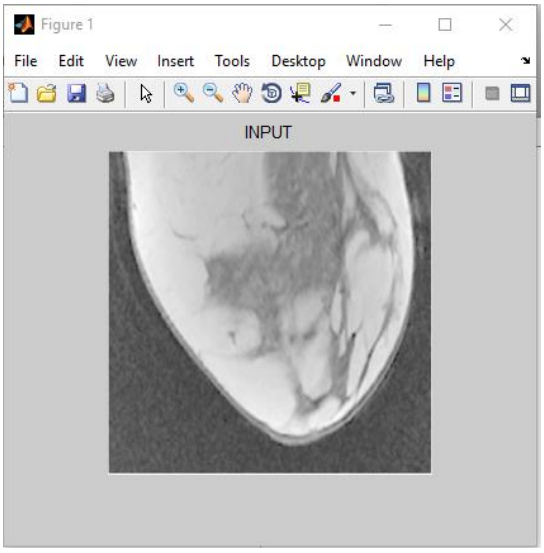
Pseudocode for Proposed Model

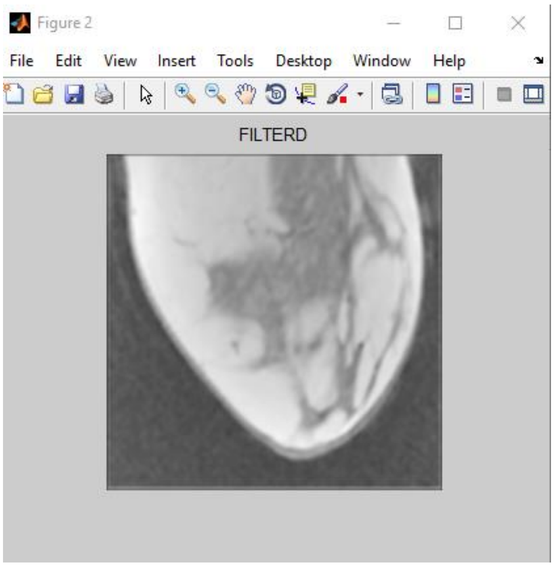
```
# Load breast cancer dataset
data = load_breast_cancer()
X = data.data  # Features
y = data.target  # Labels
# Normalize or standardize features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
# Split data into training and test sets (70% training, 30% testing)
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3, random_state=42)
# Initialize classifiers
rf = RandomForestClassifier(random_state=42)
dt = DecisionTreeClassifier(random_state=42)
knn = KNeighborsClassifier()
lr = LogisticRegression(max_iter=10000)
svc = LinearSVC(max_iter=10000)
# Training the models
rf.fit(X_train, y_train)
dt.fit(X_train, y_train)
knn.fit(X_train, y_train)
lr.fit(X_train, y_train)
svc.fit(X_train, y_train)
# Evaluate each model
rf_score = rf.score(X_test, y_test)
dt_score = dt.score(X_test, y_test)
knn_score = knn.score(X_test, y_test)
lr_score = lr.score(X_test, y_test)
svc_score = svc.score(X_test, y_test)
```

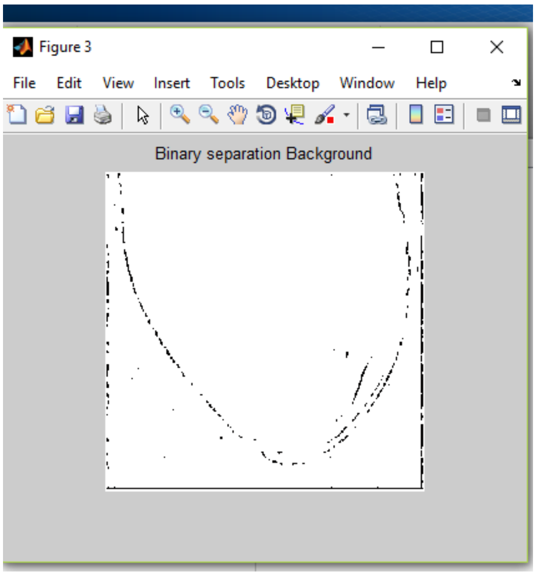
The code describes the following steps:

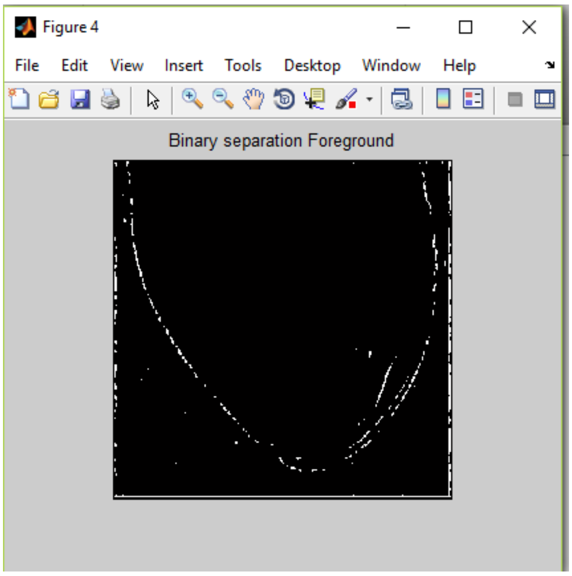
1. Loads the breast cancer dataset from scikit-learn.
2. Splits the dataset into features (X) and labels (y).
3. Normalizes or standardizes the features.
4. Splits the data into training and testing sets.
5. Initializes various classifiers: Random Forest, Decision Tree, K-Nearest Neighbors, Logistic Regression, and Linear Support Vector Classifier.
6. Trains each model on the training data.
7. Evaluates each model's accuracy on the test data.
8. Displays the accuracy scores of each model.
9. Performs feature importance analysis for Random Forest and Decision Tree.
10. Visualizes the Decision Tree.

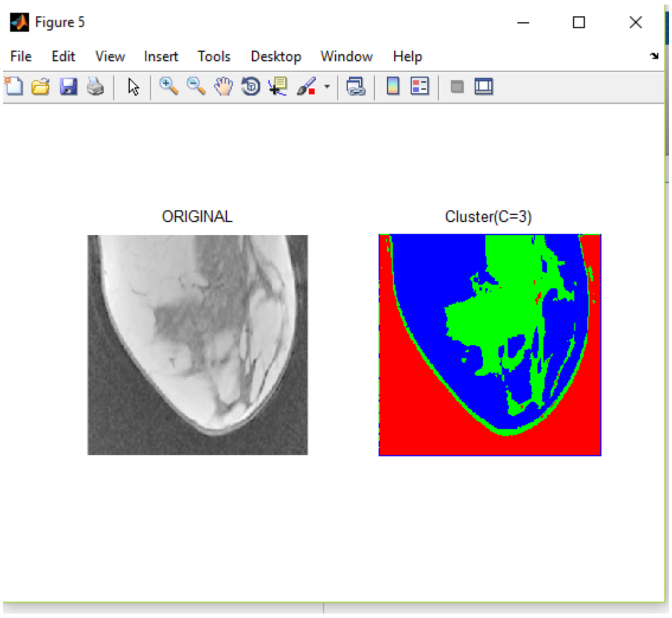
IV. EXPERIMENTAL ANALYSIS











V. CONCLUSION

The integration of ROI-centric texture feature extraction and neural network classification marks a significant advancement in the realm of mammography-based breast cancer detection. This research underscores the pivotal role of accurate texture feature extraction, derived from Regions of Interest (ROI) within mammography images, in facilitating precise cancer identification. By meticulously curating these ROIs, the method achieves enhanced and highly discriminative texture features critical for effective cancer classification. Leveraging a neural network for image categorization further amplifies the potential for accurate differentiation between cancerous and non-cancerous instances. This approach not only addresses the inherent complexities of mammography images but also contributes to mitigating the challenges faced by radiologists in interpreting and classifying these images consistently. The amalgamation of advanced feature extraction techniques with powerful classification models significantly augments the potential for early and accurate breast cancer detection. The successful utilization of this ROI-centric methodology coupled with neural network classification serves as a promising avenue for bolstering the efficiency and reliability of breast cancer diagnosis. As advancements continue in refining these methodologies and leveraging technological innovations, the potential for improved prognostic outcomes and personalized patient care in breast cancer management becomes increasingly promising.

VI. REFERENCES

[1]. Basha, S. S., & Prasad, K. S. (2009). AUTOMATIC DETECTION OF BREAST CANCER MASS IN MAMMOGRAMS USING MORPHOLOGICAL OPERATORS AND FUZZY C--MEANS CLUSTERING. Journal of Theoretical & Applied Information Technology, 5(6).

[2]. Das, P., Bhattacharyya, D., Bandyopadhyay, S. K., & Kim, T. H. (2009). Analysis and diagnosis of breast cancer. International Journal of u-and e-Service, Science and Technology, 2(3), 1-12.

[3]. Kekre, H. B., Sarode, T., Gharge, S., & Raut, K. (2010). Detection of cancer using vector quantization for segmentation. International Journal of Computers and Applications, 4(9), 14-19.

[4]. Palcic, B., Lam, S., Hung, J., & MacAulay, C. (1991). Detection and localization of early lung cancer by imaging techniques. Chest, 99(3), 742-743.

[5]. Nawgaje, D. D., & Rajendra, K. D. (2011). Implementation of fuzzy logic fo r detection of suspicious masses in mammograms using DSP TMS320C6711. International Journal of Advanced Engineering and Application.

[6]. Al-Tarawneh, M. S. (2012). Lung cancer detection using image processing techniques. Leonardo electronic journal of practices and technologies, 11(21), 147-58.

[7]. Kanungo, G. K., Singh, N., Dash, J., & Mishra, A. (2015). Mammogram image segmentation using hybridization of fuzzy clustering and optimization algorithms. In Intelligent Computing, Communication and Devices: Proceedings of ICCD 2014, Volume 2 (pp. 403-413). Springer India.

Cite this article as :

Keerthana. V, Reikha C, "Roi-Centric Texture Feature Extraction and Neural Network Classification for Mammography-Based Cancer Detection", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 11 Issue 1, pp. 165-170, January-February 2024.

Journal URL : https://ijsrset.com/IJSRSET2310657