Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230208

Blockchain and classification of mammograms and histopathology images in breast cancer lesions

Archana Barthwal[1], Kapil Joshi[2], Adarsh Kumar[3], Minakshi Memoria[2], Rajiv Kumar[2], Shubham Mahajan[4,5] and Laith Abualigah[6,7,8,9]

[1]Research Scholar, UIT,Uttaranchal University, Dehradun, India
[2]UIT, Uttaranchal University, Dehradun, India
[3]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttrakhand, India
[4]School of Engineering, Ajeenkya D Y Patil University, (iNurture Education Solutions Pvt. Ltd., Bangalore), Pune, India
[5]University Center for Research & Development (UCRD), Chandigarh University, Mohali, India
[6]Computer Science Department, Prince Hussein Bin Abdullah Faculty for Information Technology, Al al-Bayt University, Mafraq 25113, Jordan.
[7]MEU Research Unit, Middle East University, Amman, Jordan
[8]Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman 19328, Jordan.

[9]Aligah.2020@gmail.com

Abstract. This work provides an overview of the research that has been done on using Computer-Aided Detection (CAD) for the diagnosis of breast cancer. The focus is on mammographic and histopathology images and the different techniques used for image pre-processing, segmentation, and classification. The accuracy of the different algorithms was evaluated on different datasets, including MIAS, IRMA, DDSM, and CBIS-DDSM, and the results showed that deep learning models such as Convolutional Neural Networks (CNNs) and Random Forest, along with Multi-Layer Perception and Naïve Bayes, were effective in detecting and classifying breast cancer. The results of these studies show the potential of CAD in making the diagnosis of breast cancer easier and more accurate. In the segmentation stage, the U-Net model has been modified to perform better in terms of accuracy. The DDSM database has been found to have a higher accuracy percentage compared to the MIAS and CBIS-DDSM databases. This indicates that the modified U-Net model has performed well on the DDSM database in terms of accurately segmenting the images. The use of Computer Aided Diagnosis (CAD) has improved the accuracy of breast cancer diagnosis in mammography and histopathology images. The four-step process of preprocessing, segmentation, feature extraction, and classification has proven to be effective in detecting malignant and benign tumors. Different algorithms like Naive Bayes, Multilayer Perception, Random Forest, and Convolutional Neural Network (CNN) have been used with varying accuracy levels. The best results were obtained by using the modified U-Net model for segmentation and the Random Forest algorithm for classification, as they showed higher accuracy compared to other methods and databases. In this study it has been analyzed that the use of Computer-Aided Diagnosis (CAD) in mammography and histopathology images has shown promising results in the diagnosis of breast cancer. Different algorithms and techniques have been used to improve the accuracy of the classification of malignant and benign tumors. The MIAS database has shown the best results when using a Convolutional Neural Network approach. However, histopathology images have some limitations in comparison to radiologist images, but the use of CAD is still an important tool in cancer diagnosis.

Keywords: mammogram images, classification, CAD, histopathology images.

1. Introduction

Breast cancer is a common type of cancer that affects mostly women, especially in developed countries. It is caused by an abnormal increase in cells and spreads in body tissues, forming a tumor which can be benign or malignant. Mammography images are used for early detection of breast cancer, and if diagnosed early, the patient has a higher chance of successful recovery [1][2][3]. However, manual diagnosis is time-consuming, and there is a need for automatic diagnosis. Computer-aided diagnosis (CAD) systems can be helpful in medical imaging to provide quick and easy detection of diseases by doctors. In the world, 14.1 million people suffer from breast cancer each year, and 8.2 million people die from it, with 70% of new cases occurring in developing countries. It is estimated that this number will increase to 19.3 million by 2025 due to various factors such as unawareness, population, genetics, and family history [4][5][6].

2. Literature review

The pre-processing of mammographic images involves the use of filters such as Spatial and Wavelet filters, smoothing techniques, and data augmentation to increase the size and quality of the training dataset. Data augmentation is then performed by converting images from MIAS database to smaller dimensions for faster processing [7][8]. The classification of tumors is done using AlexNet architecture with an accuracy of 95.70% with and without data augmentation [9]. A novel deep learning framework using transfer learning is also used to achieve an accuracy of 97.52%.The classification of breast cancer is then performed using deep learning frameworks such as AlexNet, Modified U-Net, Inception V3, DenseNet 121, ResNet 50, VGG16, and MobileNet V7, with transfer learning being used to improve accuracy[10][11][12]. ANNs are also used for the detection of breast cancer, with a two-step process of pretreatment and extraction of images. These techniques have achieved high accuracy rates, with the best result being 98.87% achieved using the modified U-Net and Inception V3 with data augmentation on the DDSM dataset [13][14] as seen in Table 1.

Table 1. Category data on the category wise while phase are learning and testing.

Category	Learning phase	Testing Phase
Normal	150	58
Abnormal	72	42

The detection of breast cancer using neural networks is a two-part process. The first part involves the extraction of characteristics from mammographic images using the standard library CV. The second part involves using multilayer perception to classify images into normal and abnormal categories using C++ language and data from MIAS database. The results are good based on category and learning phase. Deep learning-based neural networks have improved biomedical images and various methods have been used to enhance the images [15]. Many deep learning models are used to describe the relationship between mammography and histopathology phenotypes. Challenges remain in using such systems for clinical decision making and treatment management. Mass classification is important for the benign-in-malignant detection in the region of interest, and texture is an important factor in mass classification. ANN methodology is also used in pharmaceutical science for interpretation of analytical data, pattern recognition, and modeling analytical data in pharmacokinetics [16]. Various techniques have been used by researchers for the classification of masses in mammographic images. Different classifiers such as Linear discriminants analysis (LDA), Artificial Neural Network (ANN), binary decision tree, support

DOI: 10.54254/2755-2721/8/20230208

vector machine (SVM), and Bayesian network have been used. LDA is a traditional method of classification that has been applied in the field. These classifiers aim to accurately differentiate between normal and abnormal masses in mammograms [17].ANN (Artificial Neural Network) modeling is a basic concept in pharmaceutical research that uses unsupervised learning. ANNs gather knowledge by detecting patterns and relationships in data, making it a useful tool for pattern recognition. ANN models are designed to mimic the way the human brain processes information, as it learns through experience instead of programming. In this way, ANNs are similar to the human brain and can be used to analyze complex data [18]. Classification in mammography is considered a problem of texture description and representation. The difficulty lies in detecting lesions masked behind dense tissue during mammography, particularly in the case of small tumors and microclassifications. The risk of breast cancer is increased with an increase in breast tissue density. A system has been designed using segmentation tissue (SI) and a fixed size and Region of Interest (ROI) based approach using an Extreme Learning Machine (ELM). The MIAS database has been used for classification, and an ensemble of neural network classifiers has been used to predict breast density [19]. Evaluating the risk of breast cancer is a decision-making problem. The Gail Model, which is well-measured in populations, has poor performance when applied to individuals. However, incorporating mammographic breast density into the Gail Model can improve its predictive accuracy. A study found that during 5.1 years, 9.5% of women suffered from invasive breast cancer, but the Gail Model's predictive accuracy was 95% [20].The use of Convolutional Neural Networks (CNNs) for the analysis of digital mammography images has shown promising results in terms of speed and accuracy in both image segmentation and tissue classification. The use of a median-sized database has also been found to improve the performance of the CNN.

3. Different datasets for mammographic image

Breast Imaging Report and Data System (BIRDS) of ACR (American College of Radiology) has become widely adopted as a standardized method of reporting in breast imaging, and helps to reduce the variability between radiologists in creating the reports. This system is used for mammography, ultrasonography, and MRI reports, and various datasets and learning models are used to improve its accuracy and efficiency in order to scale so here values are mentioned in Table 2 Table 3 and Table 4.

Table 2. Different databases with the method, Accuracy, and AUC (Area-Under Curve) [21].

Database	Method	Accuracy	AUC
MIAS	LeNet	0.5761	-
	CNN+SVM	0.6035	-
	AlexNet	0.6213	-
	CNN(Conv5+Fe7)+BN+SVM	0.6222	-
	CNN(Conv5 +fe7)+SVM	0.6128	-
	ZFNet	0.6295	-
	VGG16	0.6204	-
	VGG19	0.6273	-
	ResNet101	0.6062	-
	DenseNet 121	0.5871	-
	Inception V4	0.5948	-
	Mobile Net V2	0.6223	-
	Shuffle Net V2	0.6179	-
	MVCNN	0.6305	-
	MVMDCNN	0.6305	-
	MVMDCNN-LOSS	0.6305	-
IRMA	SVM	100%	1.0
	KNN K=1,2,3	100%	1.0
	BINARY DECISION TREE	100%	1.0

Table 3. Segmentation results based on modified U-Net model for the MIAS, DDSM, and CBIS-DDSM databases based on MLO view.

Data Set	Deep Learning Model	ACCURACY of IOC	ACCURACY OF DC
DDSM	U-Net Model	91.89%	91.89%
MIAS	U-Net Model	90.78%	89.99%
CBIS-DDSM	U-Net Model	87.96%	88.65%

Table 4. Determining the classification algorithm accuracy using various algorithms on different datasets.

Classification Algorithm	Inbuilt Dataset	WBC	WDBC	WPBC
Naive Bayes	74.8%	96.13%	94.20%	78.78%
Multilayer Perception	86.36%	99.28%	97.89%	81.81%
Random Forest	85.31%	99.4%	99.5%	99.6%

4. Computer-aided diagnosis

The use of computer-aided diagnosis (CAD) systems has greatly improved the accuracy and efficiency of breast cancer detection in various imaging modalities such as mammography, ultrasound, and MRI. Many techniques have been developed for the classification of masses and tissues, including linear discriminant analysis, artificial neural networks, decision trees, support vector machines, and Bayesian networks. The use of deep learning-based neural networks and data augmentation has shown promising results in detecting and classifying breast cancer. The use of Breast Imaging Report and Data System (BIRDS) of the American College of Radiology has also helped reduce variability in radiologists' reports. These efforts aim to improve the accuracy of breast cancer screening and support clinical decision making [21].

Breast cancer is a common type of cancer that occurs mostly in females diagnosing dissimilar types of imaging procedures are used for collecting the sample like mammography, magnetic resonance imaging (MRI) biopsy. The complete steps include: (i) Image acquisition: In this step, the images are collected using different imaging techniques like mammography, MRI, biopsy, etc. (ii) Image pre-processing: In this step, the images undergo several techniques to improve the quality of the images like smoothing, denoising, etc. (iii) Feature extraction: This step involves extracting relevant features from the pre-processed images to be used for classification.)iv) Classification: In this step, different classification algorithms like SVM, ANN, Bayesian network, etc. are used to classify the images into normal and abnormal categories. In this step, the extracted features from the previous step are used to classify the image into normal or abnormal (malignant) categories. The performance of the CAD system is evaluated based on different metrics such as sensitivity, specificity, accuracy, precision (Positive Predictive Value), Negative Predictive Value, and Matthew's Correlation Coefficient (MCC). These metrics help in determining the effectiveness of the system in detecting breast cancer [22] accurately.

Segmentation-In image segmentation, various techniques such as threshold clustering and active contour models are used to segment the regions of interest (ROI) that contain abnormalities from the normal breast tissue. The features extracted from the ROI are then used in the classification stage, where wavelet detection and other methods are used for detection. The results of the classification stage often show a high accuracy, as evidenced by the ROC curve being bounded by more than 90%, indicating a low rate of false positive outputs.

Feature Extraction-In this step, the extracted features from the mammography images are used to develop a breast cancer CAD system. A range of morphological, textual, fractal, topological, and intensity-based features are used to classify the new images. In one experiment, 80 mammography images were taken for feature extraction and the results showed that a breast cancer CAD system can be

DOI: 10.54254/2755-2721/8/20230208

developed with high accuracy. 23 significant features were identified which contributed to a better understanding of mammography images and the calculation of the percentage ratio.

Classification- In the classification step of the CAD system for breast cancer diagnosis, various machine learning algorithms such as decision tree, SVM (Support Vector Machines), and ANN (Artificial Neural Network) can be used. A study was conducted using the KIMA Path 960 dataset, which consisted of 960 histopathology images, and it was found that SVM achieved the highest accuracy of 90.52%, while deep learning obtained an accuracy of 81.14%. Computer-Aided Diagnosis (CAD) systems have been developed for the early detection of breast cancer. They use various imaging techniques such as mammography, ultrasonography, MRI, and histopathology. The process involves four steps: focal area identification and noise reduction, image segmentation, feature extraction, and classification. The CAD system has been tested using different algorithms such as decision tree, SVM, and ANN and has shown good accuracy in detecting and diagnosing breast cancer. The use of digital histopathology in CAD systems has made the detection and diagnosis of breast cancer easier and more accurate, leading to better outcomes for patients. The use of histopathology images in the CAD system for breast cancer diagnosis has some drawbacks compared to radiology images [23]. There is a significant parallelism between benign and malignant tumors, which makes the classification difficult. However, representation learning based on unsupervised domain adaptation can improve the classification accuracy. The results show an average increase of 5.1% compared to the basic method and an increase of 1.25% compared to advanced techniques. In the histopathology images for cancer diagnosis, the CAD system involves several steps including pre-processing, image enhancement, segmentation, and feature extraction. In pre-processing, filtering, histogram-based, and color-based techniques are used to enhance the image. Segmentation involves the use of thresholding techniques such as Gaussian and median. After segmentation, the next step is to split touching cells and extract features for further analysis.

5. Conclusion

This work describes the use of different techniques and algorithms in mammographic and histopathology images for the detection, diagnosis and prediction of breast cancer. The use of CAD (Computer Aided Diagnosis) systems is described in detail, including steps like image pre-processing, segmentation, feature extraction, and classification. The performance of different machine learning algorithms, such as SVM, ANN, and CNN, is also discussed. The accuracy of these techniques is measured using various performance metrics, such as sensitivity, specificity, precision, and recall. The paper concludes by highlighting the advantages and limitations of using CAD systems in the diagnosis of breast cancer. It has been observed that various datasets were used to evaluate the performance of different classification techniques on mammographic and histopathology images in literature. After literature analysis, this work discussed the results obtained using Convolutional Neural Network (CNN) on the MIAS database. The MVMDCNN-LOSS and MVM-DCNN methods showed an accuracy of 0.9%, which was slightly lower than the accuracy of the MVCNN method. On the other hand, the IRMA dataset was able to achieve 100% accuracy with a high Area Under the Curve (AUC) value. These results highlight the potential of using CNN and other deep learning techniques in the diagnosis of breast cancer based on mammographic and histopathology images. This work discussed the existing datasets and techniques to analyze mammographic and histopathology images for diagnosis of breast cancer. The results show that the Convolutional Neural Network (CNN) approach on the MIAS database gives a high accuracy of 0.9%, while the IRMA dataset gives 100% accuracy with an AUC score. The U-Net model is used for segmentation and is shown to be more accurate in MIAS and DDSM compared to CBIS-DDSM. In terms of classification, Random Forest and Multi-Layer Perception (MLP) algorithms are shown to be more effective than Naive Bayes. This work also briefly discussed the use of CAD in histopathology images, where preprocessing is followed by segmentation and classification.

DOI: 10.54254/2755-2721/8/20230208

References

[1] Eurostat, “Health statistics: Atlas on mortality in the European Union,” Office for Official Publications of the European Union, 2002.
[2] K. K. Saira Charan, Muhammed Jaleed Khan, “Breast Cancer Detection in Mammograms using Convolution Neural Network,” IEEE Int. Conf. Math. Eng. Technol., vol. vol 1, pp. 2–6, 2018.
[3] M. P. J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, “Competing Risks Analysis of African American Breast Cancer patients,” Adv. Breast Cancer Res., vol. vol 1.1, 2012.
[4] F. Jedy-Agba, Curado, Ogunbiyi, "Cancer Incidence in Nigeria a report from population-based cancer registries," West African J. Radiol., vol. 36, pp. 271–279, 2012.
[5] Acha, B., Rangayyan, R.M., Desautels, J.E.L.: Detection of Microcalcifications in Mammograms. In: Suri, J.S., Rangayyan, R.M. (eds.) Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer. SPIE, Bellingham (2006)
[6] Globocan IAC. Globocan 2018_ India factsheet - India Against Cancer [Internet]. 2018 [cited 2020 May 26]. http://cancerindia. org.in/globocan-2018-india-factsheet/.
[7] PBCR. Trends of Breast Cancer in India [Internet]. 2020[cited2020May26]. https://www.breastcancerindia.net/statistics/trends. HTML.
[8] Sathishkumar, V. E., Ramu, A. G., & Cho, J. (2023). Machine learning algorithms to predict the catalytic reduction performance of eco-toxic nitrophenols and azo dyes contaminants (Invited Article). Alexandria Engineering Journal, 72, 673-693.
[9] Shanmugavadivel, K., Sathishkumar, V. E., Kumar, M. S., Maheshwari, V., Prabhu, J., & Allayear, S. M. (2022). Investigation of Applying Machine Learning and Hyperparameter Tuned Deep Learning Approaches for Arrhythmia Detection in ECG Images. Computational & Mathematical Methods in Medicine.
[10] VE, S., & Cho, Y. (2023). MRMR-EHO-Based Feature Selection Algorithm for Regression Modelling. Tehnički vjesnik, 30(2), 574-583.
[11] S. Targ, D. Almeida, K. Lyman, Resnet in Resnet: Generalizingresidual architectures, In International Conference on LearningRepresentations, ICLR 2016, San Juan, Puerto Rico, 2-4 May, pp. 1–4, 2016.
[12] X. Zhang, J. Zou, K. He, J. Sun, Accelerating very deep convolutional networks for classification and detection, IEEETrans. Pattern Anal. Mach. Intell. 38 (10) (2015) 1943–1955
[13] Easwarmoorthy, S., Sophia, F., & Karrothu, A. (2016, March). An efficient key management infrastructure for personal health records in cloud. In 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 1651-1657). IEEE.
[14] Sathishkumar, V. E., & Cho, Y. (2019, December). Cardiovascular disease analysis and risk assessment using correlation based intelligent system. In Basic & clinical pharmacology & toxicology (Vol. 125, pp. 61-61). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.
[15] Arafi, Ayoub, et al. "Classification of mammographic images using artificial neural networks." *Applied Mathematical Sciences* 7.89 (2013): 4415-4423.
[16] Pavithra, E., Janakiramaiah, B., Narasimha Prasad, L. V., Deepa, D., Jayapandian, N., & Sathishkumar, V. E. (2022). Visiting Indian Hospitals Before, During and After COVID. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.
[17] Cheng, Heng-Da, et al. "Approaches for automated detection and classification of masses in mammograms." *Pattern recognition* 39.4 (2006): 646-668.
[18] Bektaş, Burcu, et al. "Classification of mammography images by machine learning techniques." *2018 3rd International Conference on Computer Science and Engineering (UBMK)*. IEEE, 2018.
[19] Tice, Jeffrey A., et al. "Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population." *Breast cancer research and treatment* 94.2 (2005): 115-122.
[20] Zhongshan, C., Xinning, F., Manickam, A., & Sathishkumar, V. E. (2021). Facial landmark

DOI: 10.54254/2755-2721/8/20230208

detection using artificial intelligence techniques. Annals of Operations Research, 1-19.
[21] Sun, Lilei, et al. "Multi-view convolutional neural networks for mammographic image classification." *IEEE Access* 7 (2019): 126273-126282.
[22] Krishnamoorthy, N., Prasad, L. N., Kumar, C. P., Subedi, B., Abraha, H. B., & Sathishkumar, V. E. (2021). Rice leaf diseases prediction using deep neural networks with transfer learning. Environmental Research, 198, 111275.
[23] Subramanian, M., Shanmuga Vadivel, K., Hatamleh, W. A., Alnuaim, A. A., Abdelhady, M., & VE, S. (2022). The role of contemporary digital tools and technologies in Covid‐19 crisis: An exploratory analysis. Expert systems, 39(6), e12834.