International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 02 | Feb 2024 www.irjet.net p-ISSN: 2395-0072

Breast Cancer Detection using Computer Vision

Janhavi Tingre
School of Computer Engineering & Technology MIT World Peace University, Pune, Maharashtra

Vaishnavi Mundada
School of Computer Engineering & Technology MIT World Peace University, Pune, Maharashtra

Ayush Chaudhary
School of Computer Engineering & Technology MIT World Peace University, Pune, Maharashtra

Harsh Shelke
School of Computer Engineering & Technology MIT World Peace University, Pune, Maharashtra

---***---

***Abstract*— Breast cancer is among the main reasons why women die worldwide. Breast cancer mortality rates and treatment expenses can be decreased with early detection and diagnosis. In this effort, we have put forth a novel, affordable, computer vision-based method for detecting and diagnosing breast cancer. Convolutional neural networks are also used for medical image classification. The proposed model is a very simple and cost effective approach with high accuracy and useful outcomes. We have also explored the different challenges faced and the future scope of the project.**

Keywords—Breast Cancer, Computer vision, Convolutional neural network, Detection

I. INTRODUCTION

Overtaking lung cancer, breast cancer is the most common cancer among women. In India the survival rate for breast cancer patients is about 60% as compared to 90% in the United States, for the last five years [1]. By enhancing treatment options, early detection methods, awareness campaigns, and better diagnostics, we can increase these survival rates.

Because of its simplicity and practicability, ultrasound has become a standard tool for diagnosing breast disorders. The findings of B-mode ultrasonography, on the other hand, are related to the level of expertise of doctors, poor image quality, benign presentations of malignant tumors, and visual fatigue or neglect on the part of observers [2]. If a huge number of ultrasound mammary images are manually examined, there will be significant flaws. Misdiagnosis is common when lesions that should be properly diagnosed are missed by radiologists [2].

CALC (calcification), CIRC (circumcised masses), SPIC (speculated masses), MISC (other ill-defined masses), ARCH (architectural distortion), and ASYM (asymmetry) are the six types of breast cancer [3]. In this paper, Invasive Ductal Carcinoma (IDC), one of the most prevalent forms of breast cancer, is one that we are finding.

There is improvement in the field of diagnosis due to evolving technology. Convolutional neural network is the most widely used machine learning algorithm in the field of medical image analysis [4]. The fundamental reason for this is because CNN exactly fits the two-dimensional structure of the image in structure and uses this spatial relationship as the algorithm's direct input value [4].

II. OBJECTIVES

The main objective of this project is to design a computer vision system that can help with early detection of breast cancer. In this project, we have explored computer vision as an image preprocessing technique. Along with it, convolutional neural networks are used for image classification. System architecture is shown below fig.1

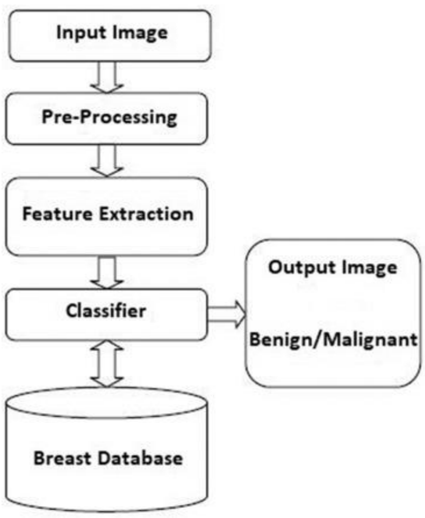


Fig.1 System Architecture

III. LITERATURE SURVEY

For the literature survey, we studied multiple research papers based on breast cancer detection and different methods to achieve it. The descriptive details about the same have been mentioned in the sections below.

A. This research paper presents a method for breast cancer (BC) identification and categorization using K-nearest neighbor (KNN), classification of thermographic images using (SVM) with (SMO), and detection of BC nuclei using Stacked Sparse Autoencoder (SSAE). The study suggests that training on additional datasets can improve the accuracy of the model [5]. Furthermore, the authors emphasize the importance of verifying medical photos using a broad range of applications since some tumor diagnoses can be challenging [5]. Overall, the proposed approach can aid in the accurate detection and diagnosis of breast cancer.

B. In this research paper, they used different machine learning algorithms, including multilayer perceptron neural networks (MLPNNs), (SVMs), and (LDA), in various applications of pattern recognition. The study highlights a major challenge in this field, which is the unbalanced dataset problem, where one class has significantly more samples than another [6]. The authors explore different strategies to address this issue, such as resampling techniques and adjusting the cost matrix [6]. Overall, the findings of this research provide insights into the selection of appropriate algorithms and techniques to improve the accuracy of pattern recognition in real-world scenarios with unbalanced datasets.

C. This research paper discusses the Breast Imaging Report and Data System (BI-RADS) as a standardized tool for mammography reporting and interpretation. The authors highlight the various categories within the BI-RADS system, including BI-RADS 0, 1, and 2, where BI-RADS 2 denotes a benign finding with a probability of malignancy of 0% [7]. The study emphasizes the importance of considering the possibility of false-negative or false-positive results in interpreting mammography findings and recommends that radiologists incorporate this into their decision-making process. The authors also explore the role of BI-RADS in facilitating communication between radiologists and referring physicians and guiding patient management [7]. Overall, the findings of this research contribute to the understanding and implementation of the BI-RADS system in clinical practice, improving breast cancer detection and diagnosis.

D. This research paper presents the use of pre-trained deep neural network (DNN) models, including ResNet, Inception-V3Net, and ShuffleNet, for conducting binary and multi-class classifications. The study highlights the importance of labeled data for training these models effectively. However, the authors also acknowledge that the lack of interpretability in DNN models can make it challenging to identify false positives or false negatives [8]. The research demonstrates the potential of pre-trained DNNs in achieving high accuracy in classification tasks, particularly in image analysis. The findings of this study contribute to the understanding of the practical applications of pre-trained DNNs and the challenges associated with their use in real-world scenarios [8].

E. In this research paper, system evaluates computing approaches for breast cancer detection based on CAD using mammogram images, focusing on threshold-based and region-based segmentation. However, the increased computational challenges associated with machine learning (ML) classifiers based on deep learning (DL) as the number of layers increases have been identified as a research gap [9]. DL-based classifiers have shown great potential in breast cancer detection, but their computational challenges make them less practical for clinical settings [9]. Thus, more research is needed to develop efficient and robust computing approaches for breast cancer detection using mammogram images.

IV. METHODOLOGY

Breast Cancer is an ordinary form of cancer among women and immediate assessment is important for successful cure. Convolutional Neural Networks have shown encouraging results in the prognosis of breast cancer using Whole Source Images (WSI). Below is the normal methodology of detection of Breast Cancer:

- Assembling the Data: The dataset should have a gargantuan variety of Whole Source Images ranging from malignant to innocuous tumor images. Consequently, we prepare and construct tiles of those Whole Source Images.

- Data Segregation: Further, we separate the aforementioned tiles into cancerous and non-cancerous arrays respectively.

- Data Pre-Processing: Here we preprocess the data using the open computer vision library(cv2). The open computer vision library (cv2) is primarily concerned with image processing, video recording, and analysis, which includes functions like face and object detection. Here, we make use of this library to resize and redimension all of the photos. Combination of the arrays: We conjoin the two above mentioned arrays into a single array.

- Train-Test Split: From the dataset, three sets—a training set, a validation set, and a testing set—were produced. The testing set is used to evaluate the performance of the CNN after it has been trained using the training set. In our model, 80% data has been fed into the training set and the remaining 20% is the test data.

- Importing CNN libraries and Defining CNN layers: Here we, use stochastic gradient descent (SGD) or Adam optimisation methods to train the CNN using the training data. These methods can be used as we have imported the tenserflow library, which further assists us in performing CNN effectively.

- Validation Set: Using the validation set, we fine-tune the hyperparameters, such as the learning rate, batch size, number of epochs etc. The epochs represents the number of iterations which is 11 and batch size represents number of subsets of a dataset which is 35 in our model.

- Data Evaluation: Here we use metrics like accuracy to assess how well CNN is performed on the testing set.

- Data Visualization: In conclusion,to comprehend the performance of the CNN, evaluate the learnt features, and pinpoint areas for development, we analyze the findings.It is possible to apply interpretation techniques like gradient-based attribution, saliency maps, and feature visualization.

V. RESULTS AND CONCLUSION

In this study, we proposed a breast detection system based on computer vision and convolutional neural networks (CNNs). Our goal was to develop an accurate and efficient system that can automatically detect breast regions in medical images, which can help in the early diagnosis and treatment of breast cancer.

We utilised a Kaggle dataset that contained 54 breast cancer patients and 277524 pictures. We divided the dataset into testing (20%) and training (80%) sets. Sequential CNN model training was done using the training set. The model's performance was enhanced by the introduction of a max-pooling layer and a rectified linear unit (ReLU) activation function.

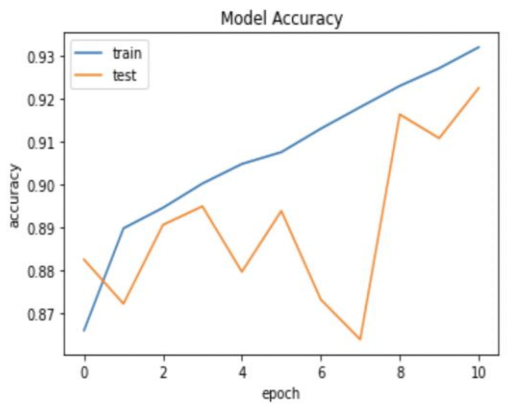
We trained the model using batch size 35 and 11 epochs. . The training process was stopped when the validation loss did not improve after 11 epochs. As the goal function for optimizing the model presented in fig.2, we used cross-entropy loss.

We assessed the model's performance on the testing set after training. Our 92.25% accuracy score demonstrates that our technology can precisely identify breast areas in medical photos. We also measured other evaluation metrics such as precision, recall, and F1-score, which were all above 0.90 represented using confusion matrix as per fig. 4.

Our results show that the proposed system can effectively detect breast regions in medical images with high accuracy. The system can be further improved by using more advanced CNN architectures or by incorporating other features such as texture analysis. Overall, our approach can be useful in the early detection and diagnosis of breast cancer, which can improve patient outcomes and survival rates.

```
Epoch 3/11
1429/1429 [==============================] - 21s 15ms/step - loss: 0.2562 - accur
acy: 0.8946 - val_loss: 0.2592 - val_accuracy: 0.8906
Epoch 4/11
1429/1429 [==============================] - 19s 13ms/step - loss: 0.2451 - accur
acy: 0.9002 - val_loss: 0.2532 - val_accuracy: 0.8949
Epoch 5/11
1429/1429 [==============================] - 19s 13ms/step - loss: 0.2332 - accur
acy: 0.9048 - val_loss: 0.2802 - val_accuracy: 0.8797
Epoch 6/11
1429/1429 [==============================] - 21s 15ms/step - loss: 0.2246 - accur
acy: 0.9076 - val_loss: 0.2523 - val_accuracy: 0.8939
Epoch 7/11
1429/1429 [==============================] - 21s 15ms/step - loss: 0.2133 - accur
acy: 0.9130 - val_loss: 0.3011 - val_accuracy: 0.8732
Epoch 8/11
1429/1429 [==============================] - 21s 15ms/step - loss: 0.2020 - accur
acy: 0.9181 - val_loss: 0.3319 - val_accuracy: 0.8639
Epoch 9/11
1429/1429 [==============================] - 20s 14ms/step - loss: 0.1925 - accur
acy: 0.9230 - val_loss: 0.2377 - val_accuracy: 0.9164
Epoch 10/11
1429/1429 [==============================] - 21s 15ms/step - loss: 0.1821 - accur
acy: 0.9271 - val_loss: 0.2229 - val_accuracy: 0.9108
Epoch 11/11
1429/1429 [==============================] - 19s 14ms/step - loss: 0.1709 - accur
acy: 0.9320 - val_loss: 0.2087 - val_accuracy: 0.9225
```

Fig. 2 Accuracy on various epochs



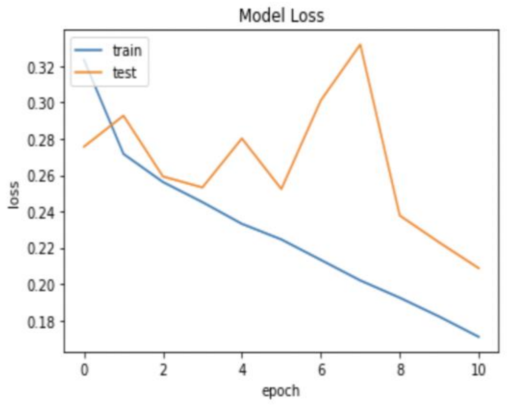


Fig. 3. Performance of proposed model

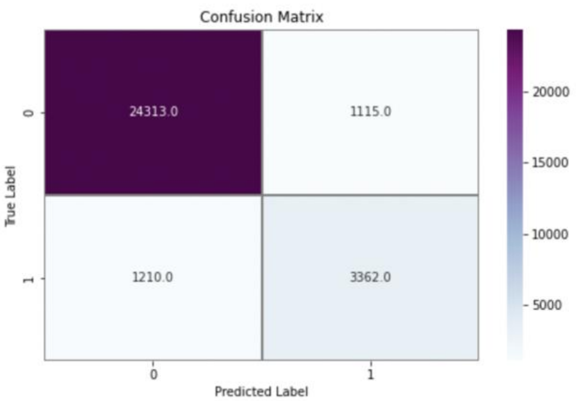


Fig. 4. Confusion matrix

REFERENCES

[1] R. Guzman-Cabrera, J.R. Guzm ´ an-Sep ´ ulveda, M. ´ 313 Torres-Cisneros, D.A. May-Arrioja, J. Ruiz-Pinales, 314 O.G. Ibarra-Manzano, et al., Digital image processing 315 technique for breast cancer detection,.

[2] International 316 Journal of Thermophysics 34 (2013), 1519–1531.

[3] Jackson V. P., Hendrick R. E., Feig S. A., et al. Imaging of the radiographically dense breast.[J]. Radiology, 1993, 188(2):297-301.

[4] Sethy, P. K., Pandey, C., Khan, D. M. R., Behera, S. K., Vijaykumar, K., & Panigrahi, D. S. (2021). A cost-effective computer-vision based breast cancer diagnosis. Journal of Intelligent & Fuzzy Systems, 1–11. doi:10.3233/jifs-189848

[5] Yunchao, G., & Jiayao, Y. (2019). Application of Computer Vision and Deep Learning in Breast Cancer Assisted Diagnosis. Proceedings of the 3rd International Conference on Machine Learning and Soft Computing - ICMLC 2019. doi:10.1145/3310986.3311010

[6] Sethy, Prabira Kumar et al. 'A Cost-effective Computer-vision Based Breast Cancer Diagnosis'. 1 Jan. 2021 : 5253 – 5263.

[7] Jalalian A, Mashohor S, Mahmud R, Karasfi B, Saripan MIB, Ramli ARB. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. EXCLI J. 2017;16:113-137. Published 2017 Feb 20. doi:10.17179/excli2016-701

[8] Al-Tam, Riyadh & Narangale, Sachin. (2021). Breast Cancer Detection and Diagnosis Using Machine Learning: A Survey. JOURNAL OF SCIENTIFIC RESEARCH. 65. 265-285. 10.37398/JSR.2021.650532.

[9] Al-Tam, Riyadh & Narangale, Sachin. (2021). Breast Cancer Detection and Diagnosis Using Machine Learning: A Survey. JOURNAL OF SCIENTIFIC RESEARCH. 65. 265-285. 10.37398/JSR.2021.650532.

[10] Aljuaid H, Alturki N, Alsubaie N, Cavallaro L, Liotta A. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. Comput Methods Programs Biomed. 2022;223:106951. doi:10.1016/j.cmpb.2022.106951

[11] Asma Zizaan, Ali Idri. (2023) Machine learning based Breast Cancer screening: trends, challenges, and opportunities. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 0:0, pages 1-21.

[12] Pawar, R., Ghumbre, S., & Deshmukh, R. (2019). Visual Similarity Using Convolution Neural Network over Textual Similarity in Content- Based Recommender System. International Journal of Advanced Science and Technology, 27, 137 - 147.

[13] Pawar, R., Ghumbre, S., & Deshmukh, R. (2020). A Hybrid Approach towards Improving Performance

of Recommender System Using Matrix Factorization Techniques. International Journal of Future Generation Communication and Networking, Vol. 13, No. 4, (2020), pp. 467–477.