IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# Cancernet Classifier for Breast Cancer Classification Using Deep Neural Networks and U-NET segmentation

**Anu Krishnan K R [1], Dr. N. Satyabalaji [2]**
[1]PG Student
Department of Computer Science and Engineering,
J.K.K. Munirajah College of Technology, T.N.Palayam
anukkr07@gmail.com
[2]Associate Professor
Department of Computer Science and Engineering,
J.K.K. Munirajah College of Technology, T.N,Palayam
hodcse@jkkmct.edu.in

**Abstract.** In the present situation accurate breast cancer detection using automated algorithms is one of the most discussing issue. Despite the fact that a lot of effort has been put into addressing this issue, an exact answer has that the majority of existing datasets are unbalanced, which means that the number of occurrences of one class vastly outnumbers those of the others. In this paper, we proposed a framework based on the concept of transfer learning and segmentation to address this issue and focus on histopathological and imbalanced image classification. To increase the overall performance of the system, we will employ the Convolutional Neural Network model with segmentation and supplement it with many state-of-the-art methodologies. The learnt knowledge was applied to the target domain of histopathology pictures using the ImageNet dataset as the source domain

**Keywords:** Breast cancer, histopathological images, transfer learning, CNN, VGG-19, UNet

## 1 Introduction

Breast Cancer (BC) is one of most prevalent form of cancer. The distinction between triple negative breast cancer, the most aggressive and fatal form of breast cancer, and non-triple negative breast cancer is a crucial unmet medical need. Timely identification of breast cancer improves the prognosis and chances of survival by allowing patients to receive timely clinical treatment. More precise benign tumour classification can help patients avoid needless therapies As a result, accurate BC diagnosis and classification of individuals into malignant or benign groups is a hot topic of research. Machine learning (ML) is generally considered as the approach of choice in breast cancer pattern classification and forecast modelling due to its unique benefits in detecting essential characteristics from complex breast cancer datasets. The outcomes that are envisioned in this work involves classifying the large-scale histopathological images using automated machine-driven procedures. The main problem of this task is that the dataset was highly imbalanced.

To address this problem, we will use transfer learning paradigm with segmentation and will train the existing CNN on more than a million ImageNet images before applying the learned knowledge to the dataset of histopathological images. The model will be further enhanced with various CNNs and segmentation methods.

The remainder of the paper is organized as follows. The related work is presented in Section 2. In Section 3, depict the system description with a brief discussion on overall system architecture. The experiments details are discussed in Section 4. The analysis of results is given in Section 5. Finally, conclusion is drawn in Section 6

## 2 Related Works

L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride and W. Sieh used deep learning to improve breast cancer detection on screening mammography [2]. In this work, they developed a deep learning algorithm that can accurately detect breast cancer on screening mammogram using an end-to-end training approach using Convolutional neural network that efficiently leverages training datasets. This study demonstrated that CNN models are trained in an end-to-end fashion can be highly accurate and potentially readily transferable across every platform. The main limitation of this work is huge memory requirement. So, the mammograms were downsized to fit available GPU which cause destruction to original image resolution. It requires larger image sizes datasets to train the models. So, it provides leverages to the training data.

In Second paper by J. Xie, R. Liu, I. Joseph Luttrell, et al., is a deep learning-based analysis of histopathological images [3] of breast cancer is conducted. In this work, they utilize the property of deep learning technique which can extract high-level abstract features from images automatically to analyze the histopathological images of breast cancer via unsupervised deep convolutional networks by K-means clustering approaches. This work has the advantage when it comes to extract expressive intricate features from histopathological images of breast cancer. But the clustering accuracies of the histopathological images of breast cancer are not as good as classification accuracy. Noise is another prevalent issue which result in significant effect on contrast resolution.

The third research work is a deep learning mammography-based model[4] for improved breast cancer risk prediction by Adam Yala, Constance Lehman, Tal Schuster, Tally Portnoi, and Regina Barzilay. In this work, they developed a deep learning model for improved breast cancer risk prediction. In this work, they developed a deep learning model that uses full field mammograms and traditional risk factor to assess a patients' future breast cancer risk. Rather than manually identifying discriminative image patterns they rely on the deep learning model to discover these patterns directly from the data. This model has an increased level of accuracy but it requires huge training on extended level that includes the rise of breast density as the proxy for the detailed information embedded on the mammogram

The fourth work is detection of breast abnormality from thermograms using curvelet transform based feature extraction[5] done by S. V. Francis, M. Sasikala and S. Saranya. This paper proposes a curvelet transform based feature extraction method for automatic detection of abnormality in breast thermograms. Breast thermograms is a image modality which represents the temperature variations of breast in its form of intensity variation on an image. Statistical and texture features are extracted from thermograms in curvelet domain to fed a Support Vector Machine for automatic classification. The classifier detects abnormal thermograms with better accuracy but the training is slow compared to Naïve Bayes.

The fifth work is a thermography-based breast cancer detection using texture features and minimum variance quantization [6] by M. Milosevic, D. Jankovic and A. Peulic. They offer a system for diagnosing aberrant patterns in breast thermograms using feature extraction techniques. Using the Nave Bayes classifier, the capacity of feature set to distinguish abnormal from normal tissue is explored. The image segmentation is done by minimum variance quantization technique. It requires small amount of data to estimate the necessary parameters. The classes are mutually exclusive thus the attributes and class frequencies affect the overall accuracy.

The sixth work by Y. Qiu, S. Yan, R. R. Gundreddy et al titled A new approach to develop computer aided diagnostic scheme for breast mass classification using deep learning technology [7] demonstrates the feasibility of applying a deep learning-based CAD scheme to classify between malignant and benign breast masses without a lesion segmentation, image feature computation and selection process. This is a mammogram-based method of classification with minimal error but it heavily depends on Region of Interest curve extracted from digital mammograms.

The final paper is the design of cognitive filter [8] for suppression of noise levels in medical images by S.Prabu, v. Balamurugan, k. Vengatesan. In this paper, a sensible and cognitive image filter which uses the principle of adaptive selection theorem in which filters are selected based on powerful deep feature fusion model for medical image processing by eliminating noises which in flip influences the accuracy of diagnosis. So, the main advantage is expanded degree of accuracy but major limitation is the availability of test subjects in context of medical images is often very limited.

From the reviews of all these papers, it is found that most of the existing works requires large memory consumption, huge processing time and performance degradation. Two approaches are used to classify breast cancer tumors such as machine learning algorithms, and deep learning. Using traditional machine learning algorithms, it requires initial pre-processing and feature selection, which take much time for computational consumption. As a result, deep learning has been used in recent studies since it can automatically extract relevant information.

## 3 Proposed Method

The basic architecture of the system is depicted in figure 1

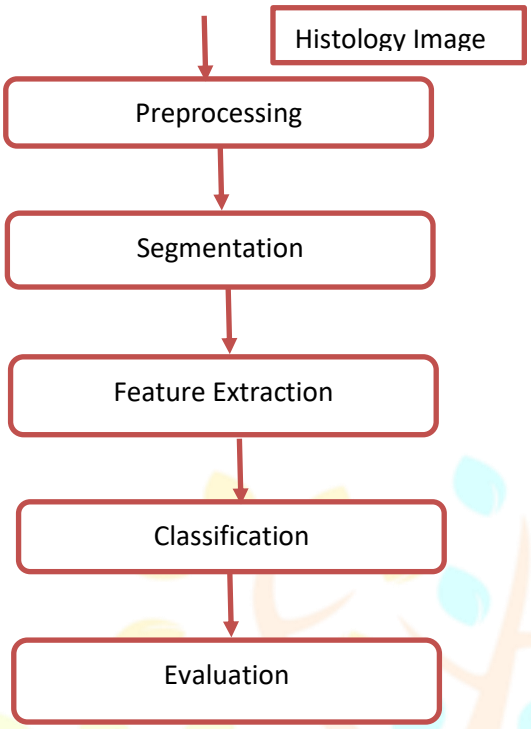


**Fig. 1.** System Architecture

The important phases of system are Preprocessing, Segmentation, Feature Extraction, Classification and Evaluation.

a. Preprocessing: The important steps in preprocessing are scaling and normalization. Scaling is the process in which the image is resized to 224 x 224 x3. In normalization, the pixel values are converted to values between 0 and 1.
b. Segmentation: Segmentation is the process of splitting the regions into multiple regions. U-Net architecture is used for segmentation. The architecture of U-Net is shown in figure 2.

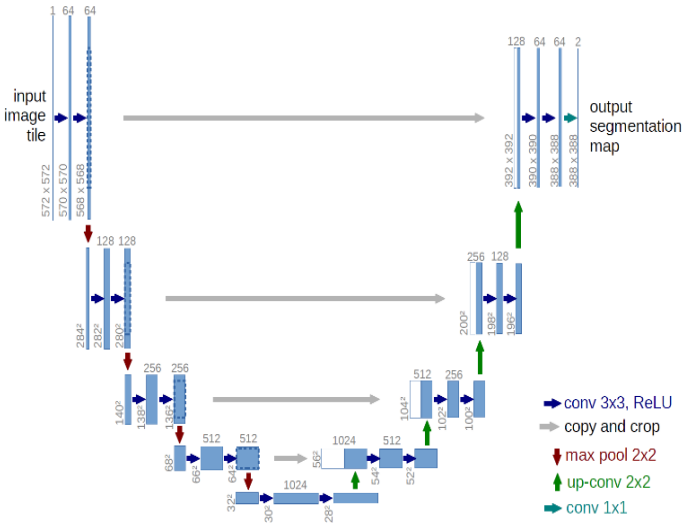


**Fig. 2.** U-Net Architecture

The network has a U-shaped architecture with a contracting path and an expansive path. The contracting path consists of repeated convolution operations, each followed by a rectified linear unit (ReLU) and a max pooling operation. During the contraction, the spatial content is decreased while feature information is increased. The expansive pathway merges the feature and spatial information through a sequence of up-convolutions and appends with high-resolution features from the contracting path.

c. Feature Extraction: The image features are extracted using Convolutional Neural Network (CNN). CNN contains three types of layers such as Convolutional layer, Pooling layer and fully connected layer. The pretrained model of CNN is utilized for this work. The model is trained using ImageNet dataset which has thousand classes. For breast cancer detections, the output of the convolution layer is taken as the input to the classification

d. Classification: The Softmax function is used for classification. This project has two class labels such as benign and malignant. So, the Softmax function predicts the probability values for two classes. The label with maximum probability is taken as the class label of the histopathology image.
e. Evaluation: The model is evaluated using Accuracy, loss, Confusion matrix, Precision, Recall and F1-score. During training, the accuracy and loss of the model is obtained. Loss is computed using Categorical Cross-Entropy. Confusion matrix is the two-dimensional representation showing the relationship between the actual and predicted output. It contains the number of True Positives (TP), False Positives(FP), True Negative(TN), False Negative(FN). Precision is the fraction of the retrieved actual positive result to the total retrieved positive results. Recall is the fraction of retrieved actual positive to the total actual positives. F1-score is the harmonic mean of Precision and recall.

## 4 Experiment

A. **Dataset**: The dataset used for this work is BreakHis dataset. BreakHis dataset is the Breast Cancer Image Classification dataset which is consisting of 9,109 images of 82 patients using different microscopic magnifying factors such as 40X, 100X, 200X and 400X. The dataset is divided into two classes such as Malignant and Benign. Benign group contains 2,480 images and malignant group consisting of 5429 images.
B. **Implementation parameters**: The optimizer used for this work is Adam Optimizer. The learning rate used here is 0.001. The model is optimized using loss. Loss is computed using categorical cross entropy. Categorical cross entropy is defined as

$$Loss = \sum_{i=1}^{N} y_i \, . \, log\hat{y}_i$$

Where $y_i$ is the actual output, $\hat{y}_i$ is the predicted output and N is the output size.

1. Training: The number of epochs for training is 25. The number of breast tissue images used for training is 5220 and that for validation is 2689.
2. Testing: The test set consists of 1000 images.

## 5 Results and Analysis

The system is experimented with various segmentation methods such as thresholding, watershed and UNET segmentation. UNET segmentation get best results compared to other methods. The various CNNs such as VGG, Resnet and Densenet are experimented for feature extraction.

In the training phase, the model is evaluated using accuracy. The graph for the accuracy is shown in figure 3.

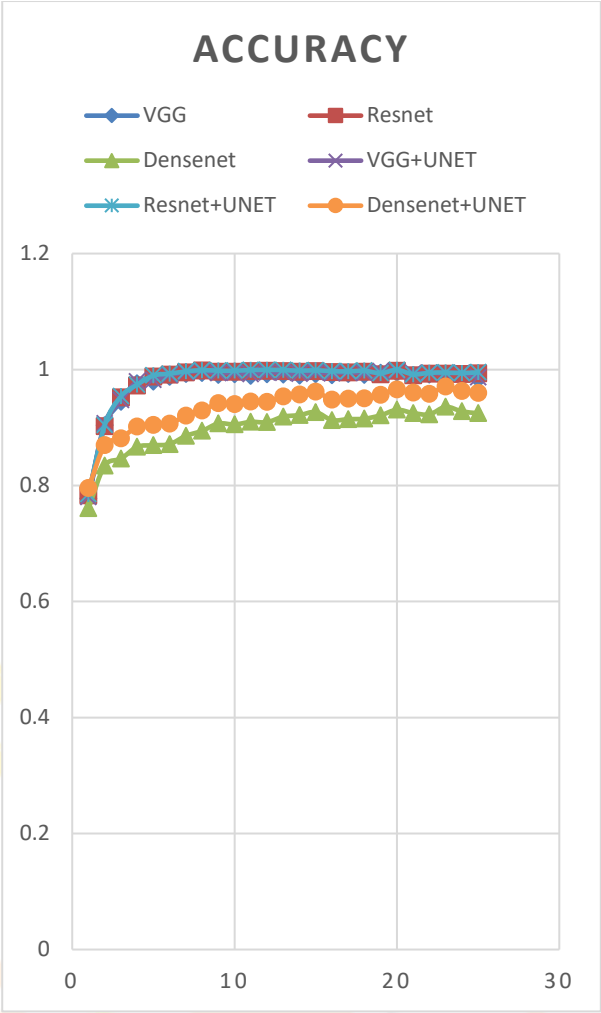


**Fig. 3.** Accuracy plot under various CNNs and Unet Segmentation

The model is optimized using loss. The plot for loss in various iterations are drawn in figure 4.

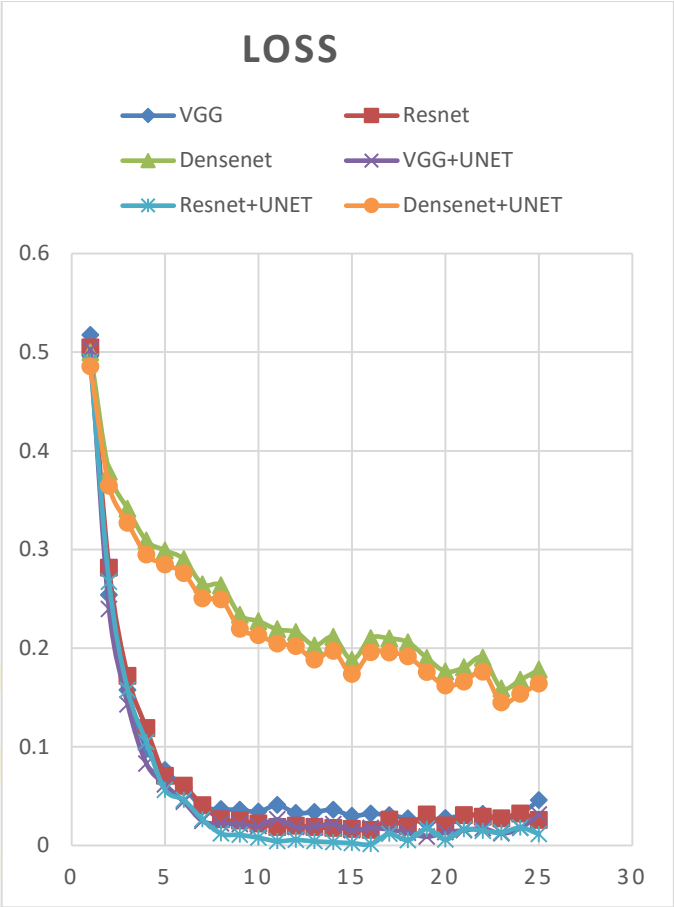


**Fig. 4.** Loss plot under various CNNs and Unet segmentation

After training, the trained model is saved in .h5 format. In testing, the images are resized to 224X224X3 and the pixel values are changed to values between 0 and 1.

The testing result is evaluated using Confusion matrix, Precision, Recall, F1-score. The table 1 shows the result of testing. From the table, VGG model with and without U-NET segmentation gets a precision score of 1(100%). Densenet model with and without U-NET segmentation gives a recall score of 1(100%). VGG model with U-NET segmentation achieves a maximum F1-score of 0.994(99.4%).

**Table 1:** Results of Testing

Some of the results obtained using the system is depicted in figure 5 and figure 6. Figure 5 shows the result of malignant image. The model tests the below image and generates a class label of malignant with a probability score of 97.6%.

| Model | Confusion Matrix | Precision | Recall | F1-score | # of layers | Parameters | Time |
|---|---|---|---|---|---|---|---|
| VGG19 | [[86 0]<br>[ 2 80]] | **1.0** | 0.9756 | **0.9877** | 19 | 22,533,886 | 700s |
| Resnet50 | [[76 10]<br>[ 2 80]] | 0.8889 | 0.9756 | 0.9302 | 50 | 33,623,614 | **675 s** |
| Densenet | [[73 13]<br>[ 0 82]] | 0.8632 | **1.0** | 0.9266 | 201 | 27,730,686 | 750s |
| VGG19+UNET | [[86 0]<br>[ 1 81]] | **1.0** | 0.9885 | **0.994** | 19 | - | 750s |
| Resnet+UNET | [[81 5]<br>[ 1 81]] | 0.9419 | 0.9878 | **0.9643** | 50 | - | 725s |
| Densenet+UNET | [[80 6]<br>[ 0 82]] | 0.9302 | **1.0** | 0.9638 | 201 | - | 800s |

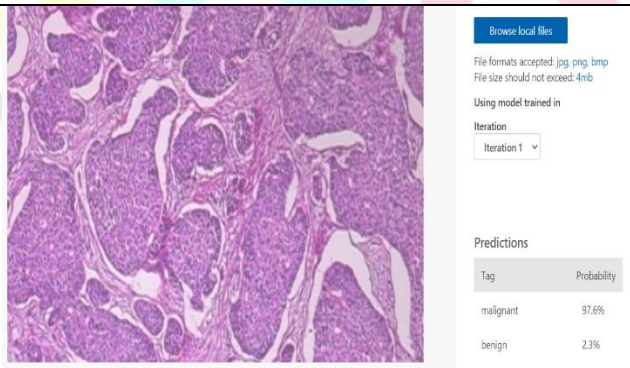


**Fig. 5.** True Positive Results

Figure 6 shows the result of Benign image. This image gets a probability score of 99.4% for benign class.

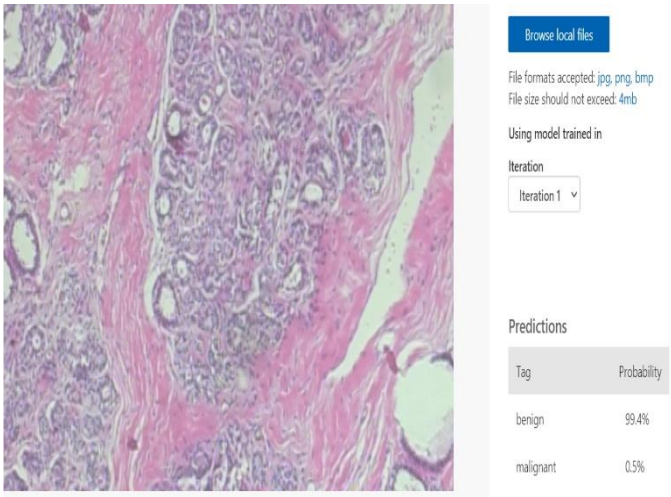


**Fig. 6.** True Negative Results

## 6 Conclusion

In this paper, Breast cancer detection is implemented from histopathology images. We employed a system architecture with segmentation, feature extraction and classification. Various CNNs and segmentation methods are experimented to achieve maximum performance of the model. VGG with and without segmentation achieves a precision score of 100%, Densenet with and without segmentation achieves a recall score of 100%. Maximum F1-score is obtained when the system utilizes VGG model with U-NET segmentation.

## References

1. Singh, Rishav, et al. "Imbalanced breast cancer classification using transfer learning." IEEE/ACM transactions on computational biology and bioinformatics 18.1 (2021)
2. L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography", Sci. Rep., vol. 9, pp. 12495, Aug. (2019).
3. J. Xie, R. Liu, I. Joseph Luttrell, et al. Deep learning based analysis of histopathological images of breast cancer. Frontiers in Genetics, 10, (2019).
4. Adam Yala, Constance Lehman, Tal Schuster, Tally Portnoi, and Regina Barzilay. 2019. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology, (2019 )
5. S. V. Francis, M. Sasikala and S. Saranya, "Detection of breast abnormality from thermograms using curvelet transform based feature extraction", Journal of medical systems, vol. 38, no. 4, pp. 1-9, (2014).
6. M. Milosevic, D. Jankovic and A. Peulic, "Thermography based breast cancer detection using texture features and minimum variance quantization", EXCLI Journal, (2014).
7. Y. Qiu, S. Yan, R. R. Gundreddy et al., "A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology," Journal of X-Ray Science and Technology, vol. 25, no. 5, pp. 751–763, (2017).
8. S.Prabu, v. Balamurugan, k. Vengatesan: Design of cognitive image filters for suppression of noise level in medical images, Measurement, 141, 296-301 (2019).
9. Abdulla, Srwa Hasan, Ali Makki Sagheer, and Hadi Veisi. "Breast Cancer Classification Using Machine Learning Techniques: A Review." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.14 1970-1979, (2021)
10. Pragya Chauhan and Amit Swami "Breast Cancer Prediction Using Genetic Algorithm Based Ensemble Approach" 18 October (2018)
11. Abien Fred M. Agarap, "On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset" 7 February (2019).
12. Priyanka Gupta and Prof. Shalini L of VIT university, vellore "Analysis of Machine Learning Techniques for Breast Cancer Prediction", 5 May (2018).
13. Muhammet Fatih Aslan, Yunus Celik , Kadir Sabanci and Akif Durdu, "Breast Cancer Diagnosis by Dierent Machine Learning Methods Using Blood Analysis Data" 31 December, (2018).
14. Yixuan Li, Zixuan Chen "Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction", October 18, (2018).
15. Mumine Kaya Keles "Breast Cancer Prediction and Detection Using Data Mining Classification Algorithms: A Comparative Study", Feb (2019)
16. Haifeng Wang and Sang Won Yoon, Department of Systems Science and Industrial Engineering State University of New York at Binghamton Binghamton, "Breast Cancer Prediction Using Data Mining Method " May (2015).
17. Wenbin Yue , Zidong Wang "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis", 9 May (2018)