SPIE.

# Classification of breast cancer abnormalities using mixed machine learning and deep learning methods

*L. G. Rodriguez-Lopez, J. A. Gonzalez-Fraga and V. Kober*

*CICESE/ Universidad Autónoma de Baja California/ Chelyabinsk State University*

CICESE

## Introduction

The main factors related to breast cancer, as found in mammograms, are masses and calcifications (abnormalities), which are the focus of this research. First, the convolutional neural networks (CNNs) were trained to obtain the final classification. After training, one of them was used to extract features, which were subsequently by machine learning algorithms to achieve the final classification.

## Dataset

CBIS-DDSM[1] was used in these experiments. A total of 2,286 images were used for training. A stratified validation set comprising 20% of the training set was created and used initially to monitor the training of the neural networks. All image patches were resized to 224x224 pixels.

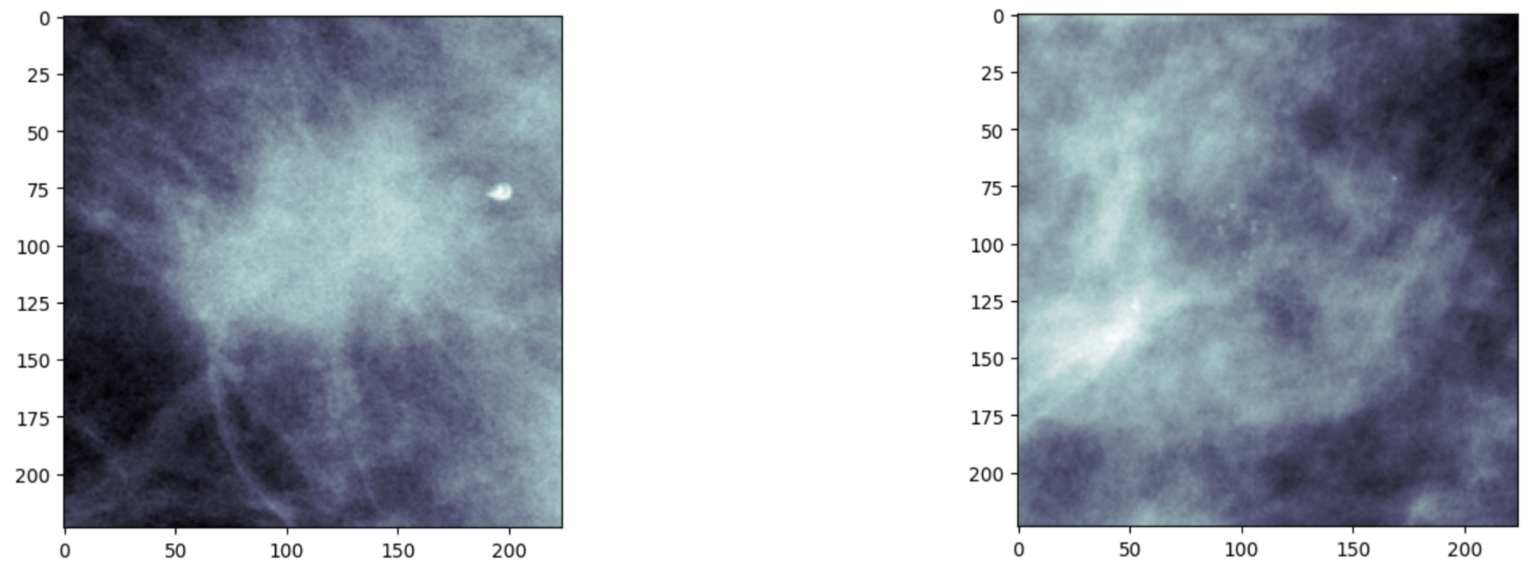

Figure 1: Examples of masses and calcifications resized to 224x224.

## Methods

VGG16, VGG19, and DenseNet121 were trained to classify abnormalities into masses and calcifications. Our approach in both Transfer Learning (TL) and Fine Tuning (FT) involves removing the dense and output layers, known as fully connected layers (FC), from each of the CNNs used for classification in the original ImageNet dataset and customizing new ones. The layers we added were consistent across all CNNs: a dropout layer with 0.5, a dense layer, and a classification layer with sigmoid activation for the binary classification task. The number of neurons in the dense layer of each CNN are: 1024 neurons for VGG16 and DenseNet121 models, while 128 for VGG19. Throughout this poster, results equal to or better than those obtained by the VGG19 network are marked in blue, and the model with the fewest misclassified images is marked in fuchsia.

Table 1: Individual results obtained from each CNN with FT and TL.

| Classifier | Accuracy | Recall | F1 Score | MCC | Precision | Misclassified Images |
|---|---|---|---|---|---|---|
| DenseNet121 | 0.8783 | 0.888 | 0.8635 | 0.7547 | 0.8394 | 73 |
| VGG16 | 0.8983 | 0.8494 | 0.8782 | 0.7924 | 0.909 | 61 |
| VGG19 | 0.9033 | 0.8455 | 0.883 | 0.8032 | 0.924 | 58 |

## Feature extraction

After training, we utilized the VGG16[2] network to extract two feature vectors. Feature Vector 1 was obtained from the final max-pooling layer. This output has a volume shape of (7,7,512), which was flattened into a feature vector of 25,088 dimensions (7x7x512=25,088). Feature vector 2 was obtained from its unique dense layer. This output is a unidimensional vector with shape of 1024. Both vectors were standarized before being input into any machine learning model. See Figure 2.

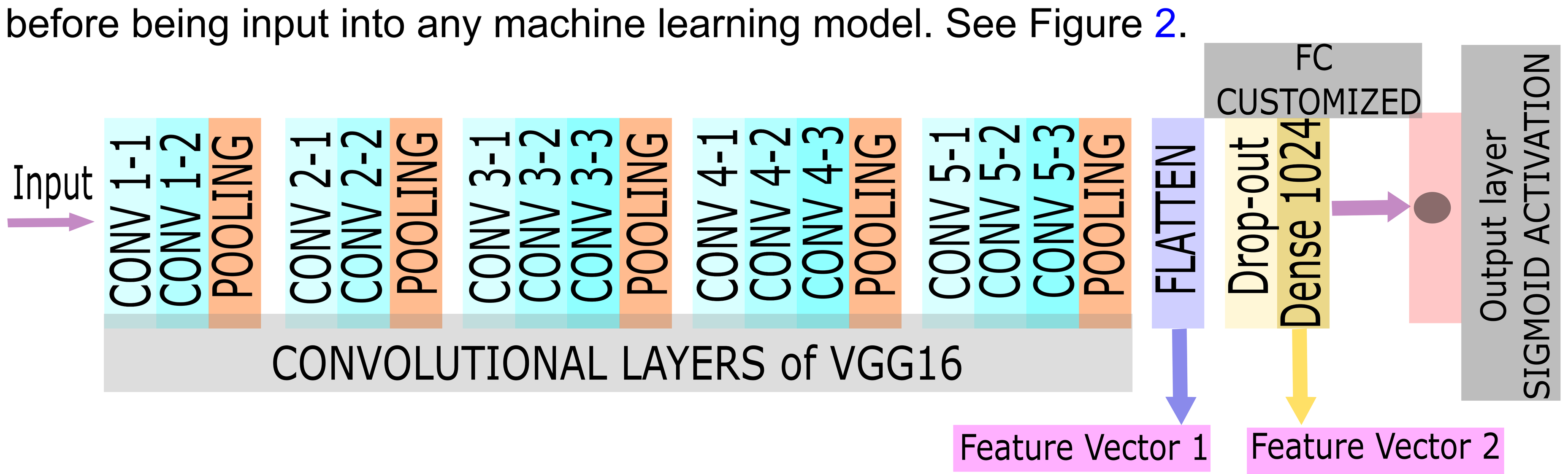


Figure 2: Feature vectors obtained from customized VGG16.

Table 2: Feature extraction results using Feature Vector 1.

| | Classifier | Accuracy | Recall | F1 Score | MCC | Precision | Misclassified Images |
|---|---|---|---|---|---|---|---|
| Feature Vector 1 Results | Naive Bayes | 0.6516 | 0.8764 | 0.6847 | 0.3774 | 0.5618 | 209 |
| | AdaBoost(Naive Bayes) | 0.8133 | 0.7876 | 0.7846 | 0.6199 | 0.7816 | 112 |

Table 3: Feature extraction results using 1024 output neurons (Feature Vector 2).

| | Classifier | Accuracy | Recall | F1 Score | MCC | Precision | Misclassified Images |
|---|---|---|---|---|---|---|---|
| Feature Vector 2 Results | AdaBoost(SVC) | 0.5683 | 0 | 0 | 0 | 0 | 259 |
| | SGD | 0.6833 | 0.9961 | 0.7308 | 0.5021 | 0.5771 | 190 |
| | Logistic regression | 0.9033 | 0.8455 | 0.883 | 0.8032 | 0.924 | 58 |
| | Voting Classifier | 0.9033 | 0.8455 | 0.883 | 0.8032 | 0.924 | 58 |
| | Random Forest (RF) | 0.88 | 0.8764 | 0.8631 | 0.7566 | 0.85 | 72 |
| | Voting Classifier with RF | 0.9083 | 0.8378 | 0.8875 | 0.8146 | 0.9434 | 55 |

## Voting and Stacking Classifiers using CNNs

Since CNNs are probabilistic classifiers, "soft voting" can be employed. Additionally, since the validation set was not used to adjust the parameters of the CNNs, it can be used to generate a "blending set" and thus utilize the Stacking technique, as used by Ref. 3 in their pathology classification problem.

Table 4: Results from soft voting of the trained CNNs.

| Voting Classifier | Accuracy | Recall | F1 Score | MCC | Precision | Misclassified Images |
|---|---|---|---|---|---|---|
| VGG16-19 | 0.9083 | 0.861 | 0.8902 | 0.813 | 0.9214 | 55 |
| VGG16-DenseNet121 | 0.9166 | 0.8996 | 0.9029 | 0.8301 | 0.9066 | 50 |
| VGG19-DenseNet121 | 0.915 | 0.8918 | 0.9005 | 0.826 | 0.9094 | 51 |
| VGG16-19-DenseNet121 | 0.9083 | 0.8725 | 0.8915 | 0.8127 | 0.9112 | 55 |

Table 5: Best results obtained with each meta-learner in each blending set formed by the indicated CNNs.

| STACKING Blending Set | Meta-Learner | Accuracy | Recall | F1 Score | MCC | Precision | Misclassified Images |
|---|---|---|---|---|---|---|---|
| VGG16-19 | Bagging(LogReg) | 0.9116 | 0.8648 | 0.8942 | 0.8199 | 0.9256 | 53 |
| VGG16-DenseNet121 | Ridge Classifier | 0.9166 | 0.9034 | 0.9034 | 0.83 | 0.9034 | 50 |
| VGG19-DenseNet121 | Logistic Regression | 0.9133 | 0.8957 | 0.8992 | 0.8232 | 0.9027 | 52 |
| VGG16-19-DenseNet121 | Bagging(SVC) | 0.9116 | 0.911 | 0.899 | 0.8207 | 0.8872 | 53 |

## Conclusions

1.Although the experiments using 25,088 features did not yield the highest success, they highlight the potential benefits of applying dimensiolnality reduction techniques to improve performance.

2.- The experiments conducted using Voting and Stacking Classifiers revealed that the bests results were achieved by combining the VGG16 and DenseNet121 networks, showcasing the effectiveness of these models working together over the VGG19 network.

3.- Overall, these findings demonstrate that the combination of CNNs with machine learning classifiers can significantly enhance performance in the classification of breast abnormalities, underscoring the value of hybrid approaches in medical image analysis.

## References

1.Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., and Rubin, D. L., "A curated mammography data set for use in computer-aided detection and diagnosis research," Scientific data 4, 170177 (2017).

2.- Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," in [3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings], Bengio, Y. and LeCun, Y., eds. (2015).

3.-Nemade, V., Pathak, S., and Dubey, A. K., "Deep learning-based ensemble model for classification of breast cancer," Microsystem Technologies 30, 513–527 (May 2023).