ISSN 1870-4069

Shallow Convolutional Neural Network to Classify Microcalcifications Clusters in Digital Mammograms

Ricardo Salvador Luna-Lozoya[1], Humberto de Jesús Ochoa-Domínguez[1], Juan Humberto Sossa-Azuela[2], Vianey Guadalupe Cruz-Sánchez[1], Osslan Osiris Vergara-Villegas[1]

[1] Universidad Autónoma de Ciudad Juárez, Instituto de Ingeniería y Tecnología, Mexico

[2] Instituto Politécnico Nacional, Centro de Investigación en Computación, Laboratorio de Robótica y Mecatrónica, Mexico

al216618@alumnos.uacj.mx, hochoa@uacj.mx, hsossa@cic.ipn.mx
{vianey.cruz,overgara}@uacj.mx

Abstract. Convolutional Neural Networks (CNNs) have proven to be an efficient tool to classify medical image data. In this paper, we propose a new shallow CNN to classify into presence or absence of microcalcifications clusters in digital mammograms. The network consists of two convolutional layers of 6 and 16 filters of size 9×9, respectively without pooling layers. After, a GlobalPooling layer is used to reduce the dimensionality from 3D to 1D and to avoid the flattening operation and the dense layers. The output layer is a sigmoid function for binary classification purposes. The loss function used was the Binary Cross Entropy. The network was trained using the INbreast database. The overall accuracy of the network is 99.3% with 8,301 parameters with as compare to the MobileNetV2 network that achieves 99.8% with 67,797,505 parameters.

Keywords: Microcalcifications detection, shallow CNN, deep learning.

1 Introduction

Breast cancer is the most prevalent cancer among women being a significant problem of public health [10]. Microcalcifications (MCs) are the most significant indirect signs of early breast cancer and its detection represents a 99% survival at 5 years or more [3]. The MCs are small deposits of calcium typically in the range of 0.1 mm to 1 mm [4]. The microcalcifications clusters (MCCs) are found in up to 50% of mammograms with confirmed cancer and correspond to at least

three MCs per cm^2 [17, 19, 20]. Mammography is described as the most widely used technique for the detection of breast cancer in early stages [4, 7].

Currently, Deep Learning (DL) models [13], trained with large amounts of data, have achieved high degrees of accuracy. In this sense, the CNNs techniques are studied in the field of MCCs detection [4]. The problem is that the mammographs or X-ray scanners have specific configurations, making them different from each other. Consequently, each hospital or clinic would have to train their own model, which would be a complicated situation because of the CNN architectures have become more complex and deeper, requiring a lot of computational power, specialized equipment and time invested to train a model.

Automatic systems to discriminate MCCs from normal tissue are always demanded. Hence, we propose a shallow CNN to classify patches of digital mammograms into presence or absence of MCCs. The network consists of two convolutional layers of 6 and 16 filters of size 9×9, respectively without pooling layers. A GlobalPooling is used to reduce the dimensionality and to avoid the flattening operation and the dense layers. For binary classification, the output layer is a sigmoid function. The loss function used was the Binary Cross Entropy (BCE). This work is a continuation of the research presented in [12].

The main contribution of the paper is:

– A light and shallow CNN to classify MCCs in digital mammograms. The network is light because of its reduced number of parameters and shallow because it has two convolutional layers only.

The rest of the article is organized as follows: in Section 2, the state of the art. In Section 3, the material and methods are shown. In Section 4, the experiments and results are presented. In Section 5, the discussion of the results is presented. Finally, the conclusions are presented in Section 6.

2 State of the Art

In our perusal, around 90 journal papers were analyzed. This section describes the current evidence found related to the detection of MCCs using DL.

Hsieh et al. [9], implemented a VGG-16 network to find MCCs in mammograms. A Mask R-CNN was used to segment the MCCs of the previously found MCCs and remove background noise. Then, the InceptionV3 was used to classify MCCs into benign or malignant. Accuracy for classification and detection was 93%, for MCs labeling 95%, and classification 91%. Precision, specificity, and sensitivity of the entire method was 87%, 89%, and 90%, respectively.

Rehman et al. [15], proposed a Fully Connected Deep CNN (FC-DSCNN) diagnosis system to detect MCCs and classify them into the classes benign and malignant. The proposed system has four steps: 1) image processing and data augmentation, 2) RGB to a grayscale transformation, 3) suspicious regions segmentation and 4) MCCs classification. First, the mammogram is divided

into subregions and sent to the FC-DSCNN to classify them into malignant and benign class. A total of 6,453 mammograms, from the DDSM and PINUM databases, were used. The results showed a sensitivity, a specificity, a precision and a recall of 99%, 82%, 89% and 82%, respectively.

Valvano et al. [18], developed two CNNs, one to detect possible Regions of Interest (ROIS) containing MCs and the other to segment the ROIS. They used 283 mammograms with a resolution of 0.05 mm from a private database. Square patches of $n \times n$, overlapped by n/2 pixels were extracted from the mammogram. For each patch, a positive label was associated when the patch contains MCs and a negative label when it does not. Each patch is processed by a CNN that detects the presence or absence of MCs. Afterwards, the ROIS found are entered into a segmentation CNN, which returns a mask. The mask is analyzed by a labeling algorithm to locate the position of each MC within the mask. Both CNNs are built of six convolutional layers with kernel of 3×3 and stride of one. Patches of 29, 39 and 49 squared pixels were used. The last size yielded the best results with an accuracy of 98.22% for the detector and 97.47% for the segmenter.

Gómez et al. [8], proposed a methodology to preprocess digital mammograms from the mini-MIAS and the UTP databases to detect presence or absence of MCCs. First, they divided mammograms in 4,292 patches of size 101×101 pixels, 3,500 used for training and 792 for testing purposes. In total, 2,360 patches out of the 4,292 contained MCCs and 1,932 did not contained these lesions. The CNN proposed used seven hidden layers with 8, 16, 32, 64, 128, 256, and 512 filters, respectively and kernel size of 3×3. After each convolutional layer, a 2×2 MaxPooling layer and a layer of Rectified Linear Unit (ReLU) activation functions were added. A softmax layer was used for multiclass classification purposes. The model yielded an accuracy of 96.26% during training while during testing 95.83%.

Luna et al. [12], presented a comparison of the CNN architectures InceptionV3, DenseNet121, ResNet50, VGG-16, MobileNetV2, LeNet-5 and AlexNet for classifying MCCs. In the best testing accuracies, InceptionV3 achieved 99.71%, DenseNet121, ResNet50 and VGG-16 yielded 99.74%, MobileNetV2 obtained the best overall accuracy with 99.84%, LeNet-5 99.30% and AlexNet 99.40%.

3 Materials and Methods

The CNN model was created using the Google Colaboratory Integrated Development Environment (IDE) [6], Python 3.0 language, and TensorFlow framework 2.0 [1]. The IDE automatically allocates the necessary computational resources.

The INbreast public database [14] was used to train the model. It contains 410 digital mammograms of size $2,560 \times 3,328$ pixels and $3,328 \times 4,084$ pixels with 8-bit depth (grayscale). The mammograms are labeled with various types of lesions such as asymmetries, calcifications, microcalcification clusters, distortions, regions spiculate, masses or nodules, and pectoral muscle. These

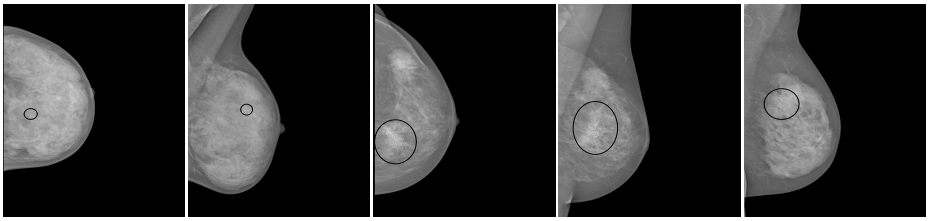
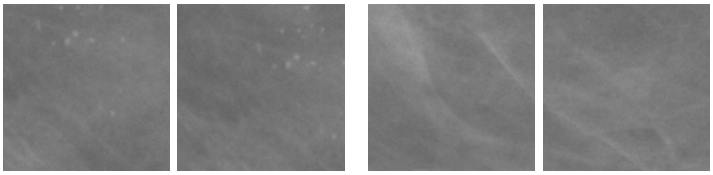


Fig. 1. Examples of mammograms showing MCCs highlighted within a circle.



(a) (b)

Fig. 2. Patches of digital mammograms (a) with MCCs and (b) normal tissue.

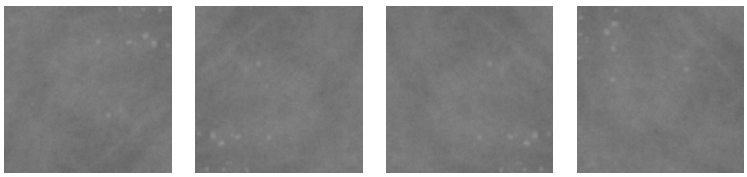
mammograms were acquired using the MammoNovation Siemens FFDM scanner. Each pixel in the image represents 70 microns. In this work, only cases labeled as MCCs were utilized. In Fig. 1, five example images labeled as MCCs from the database are shown.

3.1 Data Preparation

Digital mammograms in the INbreast database are stored in DICOM format. Therefore, we converted them into the PNG format. The labeling and coordinates of the breast lesions were searched in their corresponding XML file, independent of the images. To mark the lesions on the digital mammograms, it was necessary to develop a computer program to read the coordinates of the MCCs from the file.

Patch Extraction. The proposed CNN model processes the mammograms in patches of 1 cm^2. This corresponds to squares of area 144×144 pixels. We developed a computer program to extract the patches from the mammogram. Figs. 2(a) and 2(b) show two examples of patches with MCCs and normal tissue respectively. An expert radiologist doctor analyzed the patches manually discarding the wrong ones. She selected a total of 1,576 patches with MCCs and 1,692 without these injuries.

Data Augmentation. Due to the limited availability of mammograms labeled with MCCs in the INbreast database, there were not enough patches to adequately train, validate, and test the models. To address this issue, we augmented the database to enhance precision. In Fig. 3, we illustrate four



(a) (b) (c) (d)

Fig. 3. Examples of geometric transformations on one patch. (a) Reflection, (b) 180^o turn, (c) reflection and 180^o turn, and (d) 90^o turn.

geometric transformations applied to the patches. This augmentation process resulted in a total of 6,304 transformed patches with MCCs and 6,768 without MCCs. Only geometric transformations were employed to augment the data in order to ensure that the features of each patch remained unaltered. As a result, the augmented database comprises 7,880 patches with MCCs and 8,460 patches without MCCs, totaling 16,340 patches.

The Datasets. A total of 15,760 patches were collected, consisting of 7,880 patches containing MCCs and 7,880 patches representing normal tissue samples. It was determined by the Pareto's Principle [2] that 80% of the data would be used for both training and validation, while the remaining 20% would be reserved for testing purposes. To be specific, for training purposes, we utilized 64% (10,088 patches).

For validation, we allocated 16% (2,520 patches), and for testing, we reserved 20% (3,152 patches). This partitioning strategy proved effective in achieving optimal results. To ensure consistency, all patches were normalized by dividing their pixel values by 255, given that the pixel depth was eight bits.

The Proposed Architecture. In [12], we compared different CNN architectures and recommend the best performing layer types for classifying MCCs. From here, the CNN MobileNetV2 [16] yielded the highest overall accuracy of 99.8% with 67,797,505 parameters.

In the same work, the CNN LeNet-5 [11] yielded an accuracy of 99.3% with 2,233,365 parameters. The difference in accuracies is only 0.539%. However, the LeNet-5 is 30 times smaller. Therefore, to classify MCCs it is not necessary to implement Deep CNNs. Hence, after extensive testings and combinations of the layers suggested in [12], we obtained the shallow CNN model shown in Fig. 4, that comprises only two convolutional layers without any intervening pooling layers. To optimize the number of trainable parameters, the conventional combination of a flat layer with a fully connected layer was replaced with a GlobalPooling layer, resulting in a significant reduction of parameters.

The input layer extracts the characteristics of the patches x from mini batches of size 64. The layer comprises 6 filters W_0 of size 9×9 with biases B_0 and a layer of ReLU activation functions. The output can be defined as in Eq. (1):

$$F_0 = \max(0, W_0 x + B_0). \tag{1}$$

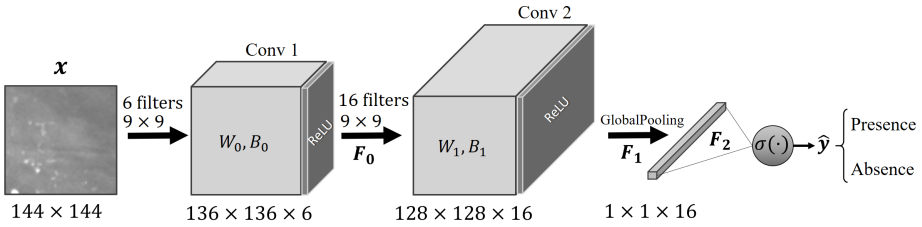


Fig. 4. Proposed CNN architecture to classify MCCs.

The second layer has 16 filters W_1 of 9×9 with another ReLU layer. Similarly, at the output of this layer we have Eq. (2):

$$F_1 = \max(0, W_1 F_0 + B_1). \tag{2}$$

The output volume F_1 represents the feature maps of the second layer which are further processed by a GlobalPooling layer, which extracts the highest value of each map as shown in Eq. (3):

$$F_2 = \max(F_1). \tag{3}$$

The features from the GlobalPooling are sent to the sigmoid to predict the probability of the binary variable as represented in Eq. (4):

$$\hat{y} = \sigma(F_2). \tag{4}$$

The BCE cost function is shown in Eq. (5):

$$\mathcal{L}(y, \hat{y}) = -\frac{1}{m} \sum_{i=1}^{m} \left[y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \right]. \tag{5}$$

Where m is the size of the training set used, y_i is the target value that can take two possible values, 0 or 1 (presence or absence of MCCs), and $\hat{y}_i$ is the predicted value. We used dropout to regularize the network.

Hyperparameter Tuning. The tuning of hyperparameters consists of choosing the values that achieve the maximum performance of the model. We used the random search method [5] for hyperparameter tuning, because the proposed network is very short and allows us to specify the number of models to train. Besides, we can base our search interactions on our computational resources (which is limited) or the time taken by iteration. The validation loss was monitored for up to 100 epochs. The resulting hyperparameters are shown in Table 1.

Table 1. Resulting hyperparameters for the proposed CNN architecture.

Hyperparameter	Value
First Convolutional Layer	
Number of features map	6
Kernel size	9 x 9
Activation funcion	ReLU
Second Convolutional Layer	
Number of features maps	16
Kernel size	9 x 9
Activation Function	ReLU
GlobalPooling Layer	
Number of features	16
Dense Layer	
Units	1
Activation Function	Sigmoid
Network Training	
Loss function	Binary Cross Entropy (BCE)
Optimization algorithm	Adaptive Moment Estimation (ADAM)
Learning rate	0.001
Batch size	64
Epochs	100
Dropout	Keep 80%

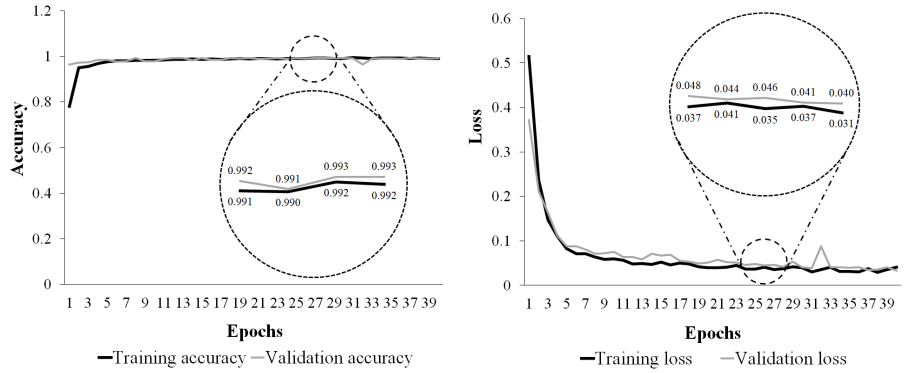
4 Experiments and Results

The output layer is a sigmoid with real output between 0 and 1. In our work, the range between 0 and 0.49 indicates '0' or absence and from 0.5 to 1 indicates '1' or presence of MCCs in a patch. All the models tested were trained with one hundred epochs. At the end of each epoch, the models were validated with the validation set to obtain the accuracy and the error. We selected the model with the highest accuracy. The dense layer is the output sigmod whose inputs are the 16 output features delivered by the GlobalPooling layer plus a bias term. Table 2 presents the number of parameters per layer, as well as the sum of these parameters after hyperparameter tuning.

The accuracy and loss plots during the training and validation are depicted in Fig. 5. Upon closer inspection, it becomes evident that both the training and validation results exhibit remarkable similarity. Table 3 shows the results of the proposed network compared with the state-of-the-art MobileNetV2 and LeNet-5.

Table 2. Parameters of the proposed CNN architecture.

Type of Layer	Parameters
Convolution 1	492
Convolution 2	7,792
Dense	17
Total	**8,301**



(a) (b)

Fig. 5. Proposed CNN performance during training and validation epochs. (a) shows the accuracy performance and (b) shows the loss performance.

5 Discussion

From Figure 5, it shows that the training and validation accuracies follow each other very closely. Also, the losses of training and validation are similar and the validation loss does not increase after a number of epochs, showing that no overfitting is present.

Table 3 shows that the accuracy of the MobileNetV2 is greater by only 0.5%. However, the number of parameters of the proposed network is much less even than the LeNet-5, which makes the proposed network light and shallow suitable for MCCs classification.

Clean data is important for optimal model performance. In this work, the data set was cleaned by an expert. She validated the model by observing the patch and checking the decision made by network (presence/absence).

The continuation of the present investigation is a faster residual network, with better performance, under the assumption that it is not necessary a deep neural network to classify MCCs. Also, the networks reported in [12] are being investigated to include other type of lesions.

Table 3. Performance comparison of the proposed CNN versus the MobileNetV2 and the LeNet-5.

Architecture	Accuracy	Parameters
MobileNetV2	**99.8**%	67,797,505
LeNet-5	99.3%	2,233,365
Proposed	99.3%	**8,301**

6 Conclusions

We presented a new shallow CNN architecture to classify MCCs using only two convolutional layers without pooling layers between the convolutions and a GlobalMaxPooling layer. The results of the proposed model yielded similar results to the deeper CNN MobileNetV2. This demostrate that for MCCs classification, shallow networks produce similar results to their deeper and more complex counterparts. The proposed model is already implemented in a web application that inspects digital mammograms. The application is been tested in collaboration with the Centro de Imagen e Investigación (Medimagen) of Chihuahua, México. Currently, we are compiling a database of Mexican mammograms to train shallow models that can work in hospitals and clinics of the country.

Acknowledgments. Ricardo Luna thanks the UACJ for the support provided and the CONACYT for the scholarship granted to pursue his doctoral studies. We would like to express our gratitude to the radiologist Dra. Karina Núñez, from the Salud Digna Clinical and Imaging Laboratory, for her support in carrying out this work.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C. a.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), https://www.tensorflow.org/
2. Ali, S., Mohammed, A., Hefny, H.: An enhanced deep learning approach for brain cancer MRI images classification using residual networks. Artificial Intelligence in Medicine, vol. 102, pp. 101779 (2020)
3. American Cancer Society: Tasas de supervivencia del cáncer de seno. cancer.org. https://www.cancer.org/es/cancer/cancer-de-seno/comprension-de-un-diagnostico-de-cancer-de-seno/tasas-de-supervivencia-del-cancer-de-seno.html, accessed May. 13, 2023
4. Basile, T., Fanizzi, A., Losurdo, L., Bellotti, R., Bottigli, U., Dentamaro, R.: Microcalcification detection in full-field digital mammograms: A fully automated computer-aided system. Physica Medica, vol. 64, pp. 1–9 (2019)
5. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization, vol. 13, pp. 281–305 (2012)

6. Bisong, E.: Building machine learning and deep learning models on Google Cloud Platform: A comprehensive guide for beginners. Apress Berkeley, CA, 1st edn. (2019)
7. Cronin, K., Lake, A., Scott, S., Firth, A., Sung, H., Henley, S.: Annual report to the nation on the status of cancer, part i: National cancer statistics: Annual report national cancer statistics. Cancer, vol. 124, pp. 2785–2800 (2018)
8. Gómez, A., Echeverry-Correa, D., Gutiérrez, A.: Automatic pectoral muscle removal and microcalcification localization in digital mammograms. hir, vol. 27, no. 3, pp. 222–230 (2021)
9. Hsieh, Y., Chin, C., Wei, C., Chen, I., Yeh, P., Tseng, R.: Combining VGG16, Mask R-CNN and Inception V3 to identify the benign and malignant of breast microcalcification clusters. In: 2020 IEEE International Conference on Fuzzy Theory and Its Applications (iFUZZY). pp. 1–4. IEEE, Hsinchu, Taiwan (2020)
10. International Agency for Research on Cancer: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf, last accessed May. 08, 2023
11. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE, vol. 86, pp. 2278 – 2324 (1998)
12. Luna, R., Ochoa, H., Sossa, J., Cruz, V., Vergara, O.: Comparison of deep learning architectures in classification of microcalcifications clusters in digital mammograms. In: Pattern Recognition. pp. 231–241. Springer Nature Switzerland, Cham (2023)
13. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.: Deep learning for healthcare: review, opportunities and challenges. Briefings in Bioinformatics, vol. 19, pp. 1236–1246 (2018)
14. Moreira, I., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M., Cardoso, J.: INbreast: Toward a full-field digital mammographic database. Academic radiology, vol. 19, pp. 236–48 (2011)
15. Rehman, K., Li, J., Pei, Y., Yasin, A., Ali, S., Mahmood, T.: Computer vision-based microcalcification detection in digital mammograms using fully connected depthwise separable convolutional neural network. Sensors, vol. 21, pp. 4854 (2021)
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4510–4520. IEEE, Salt Lake City, UT, USA (2018)
17. Sickles, E., D'Orsi, C., Bassett, L.: ACR BI-RADS® mammography. In: ACR BI-RADS® atlas, breast imaging reporting and data system. American College of Radiology, 5th edn. (2013)
18. Valvano, G., Santini, G., Martini, N., Ripoli, A., C., I., Chiappino, D.: Convolutional neural networks for the segmentation of microcalcification in mammography imaging. Journal of Healthcare Engineering, vol. 2019, pp. 1–9 (2019)
19. Wang, J., Nishikawa, R., Yang, Y.: Global detection approach for clustered microcalcifications in mammograms using a deep learning network. Journal of Medical Imaging, vol. 4, pp. 024501 (2017)
20. Wang, J., Yang, Y.: A context-sensitive deep learning approach for microcalcification detection in mammograms. Pattern Recognition, vol. 78, pp. 12–22 (2018)