Salud, Ciencia y Tecnología - Serie de Conferencias. 2025; 4:669
doi: 10.56294/sctconf2025669

ORIGINAL

Deep learning techniques for breast mass malignancy classification on digital mammography

Técnicas de aprendizaje profundo para la clasificación de la malignidad de lesiones tipo masa en mamografía digital

Ariel Coto Santiesteban[1], Lisbel Garzón Cutiño[2], Damian Valdés Santiago[1]

[1]Universidad de La Habana, Departamento Matemática Aplicada, Facultad de Matemática y Computación. La Habana, Cuba.
[2]Hospital Clínico Quirúrgico Docente "Dr. Miguel Enríquez". Departamento de Neurología. La Habana, Cuba.

Cite as: Coto Santiesteban A, Garzón Cutiño L, Valdés Santiago D. Deep learning techniques for breast mass malignancy classification on digital mammography. Salud, Ciencia y Tecnología - Serie de Conferencias. 2025; 4:669. https://doi.org/10.56294/sctconf2025669

Submitted: 08-02-2024 **Revised:** 07-08-2024 **Accepted:** 28-12-2024 **Published:** 01-01-2025

Editor: Dr. William Castillo-González

Corresponding author: Damian Valdés Santiago

ABSTRACT

Introduction: breast cancer is one of the most common type of cancer with a high mortality rate. Mammography is widely used to identify breast cancer. Computer Aided Diagnosis systems are used for automatic detection of breast lesions.
Method: we propose and evaluate a deep learning model, called VGG16-C300, for breast mass malignancy classification. CBIS-DDSM dataset was used for training and evaluation. Image contrast enhancement methods like CLAHE and Mean Blur where previously applied to regions of interests.
Results: the trained model achieved and area under the curve of 0,80, after 10 iterations of a 5-fold Cross-Validation.
Conclusions: VGG16-C300 could be used as a component in a computer-aided diagnosis system for breast cancer detection.

Keywords: Mammography; CLAHE; Mean Blur; Convolutional Neural Networks; Malignancy Tumor Classification.

RESUMEN

Introducción: el cáncer de mama es uno de los más comunes con un alto índice de mortalidad en el mundo. La prueba que más utilizan los especialistas es la mamografía de rayos X. En los centros de diagnóstico se emplean sistemas computarizados para la detección automática de lesiones.
Método: se propone y evalúa un modelo automatizado de aprendizaje profundo, llamado VGG16-C300, para la clasificación de anomalías de tipo masa en mamas. El entrenamiento y la evaluación se realizaron sobre la base de datos CBIS-DDSM. Fueron aplicados algoritmos de mejoramiento de contraste como CLAHE y Mean Blur.
Resultados: se obtuvo como resultado un área bajo la curva de 0,80, tras ejecutar 10 veces 5-fold Cross Validation, tomando como estimador el promedio de sus resultados.
Conclusiones: VGG16-C300 puede utilizarse como componente de un sistema de diagnóstico asistido para detectar cáncer de mama.

Palabras clave: Mamografía; CLAHE; Mean Blur; Redes Neuronales Convolucionales; Clasificación de Malignidad de Tumores.

INTRODUCTION

Breast cancer is one of the most common type of cancer worldwide with a high mortality rate, even men could suffer this disease. According to the World Health Organization, 2,1 million new cases are detected every year, causing 627 000 deaths in 2018. More than 3 500 women are diagnosed every year in Cuba, according to the Cuban Health Statistics Yearbook.[1]

There are some techniques for breast cancer detection. The most commonly used technique in the field is the X-ray mammography, due to its high effectiveness.[2] Low contrast and some other features like breast density, make difficult the diagnostic process. For diagnosis, two mammograms are needed per breast: one with a craneocaudal (CC) view, and the other with mediolateral-oblique (MLO) view, summing four images per patient.[3] Specialists analyze this views and determined if there is any benign or malignant signal.

Tumor malignancy is confirmed by a biopsy.[2] This is used when there is a high probability of malignancy, because it is an invasive technique for the patient. Computerized systems are used worldwide for tumor detection and classification, named on the literature as Computer-Aided Diagnosis systems (CADs). Breast cancer early detection is necessary due to the amount of patients affected by it and the severity of this disease.

CADs have been developed to assist radiologists in the reading process by indicating possible anomalies or regions of interest (ROIs) on mammograms, and showing the probability of malignancy for each one. CADs are considered a tool for improving early breast anomaly detection, and it could help young radiologist's formation.

The architecture of CADs typically have three stages. The first one is the input pre-processing stage, where some computer vision algorithms are applied over the input for feature improvements. The next stage is ROIs segmentation, where candidate anomalies are extracted from the mammograms. This process could be made manually or automatically. At the end, there are two final stages related with tumor classification: feature extraction, and tumor binary classification in benign or malignant. There are two main types of features: local (texture, density, shape), and global (features extracted from a Convolutional Neural Network). This features are input to machine learning (ML) models for binary classification, like Support Vector Machines (SVM), Dense Neural Networks (DeNN), or Random Forrest (RF). It is important to point out that the last two phases are dependent of the output of the previous ones, so it is necessary to take account of the full pipeline, and the nature of the mammograms for a CADs analysis.

Automated binary classification of mammographic masses malignancy have been developed by many authors in the literature. Artificial Intelligence algorithms used on CADs are based on mammogram features extracted to make a prediction.[4-7] To make that happen, they must be trained first over an image dataset as large as possible. Over the last five years, CNNs with transfer-learning and data augmentation techniques have been applied on breast cancer detection.[4-6,8] A common image dataset used for generic feature extraction comes from the ImageNet Challenge.[9]

In 2019, McKinney et al.[5] worked with three mammograms dataset. One of them was the *UK National Health Service Breast Screening Programme*.[10] This database contains 8 277 ROIs from 7 672 images of 3 871 women, annotated by OPTIMAM with centered squares that contain a ROI.[11] The *US Northwestern Memorial Hospital, Chicago database* was also used with 3 549 ROIs from 1 917 images of 694 women. Finally, they added to its training dataset the public database CBIS-DDSM.[12] They developed a CADs with three deep learning models and different architectures. The average of the three model output was taken as final prediction. Data augmentation techniques like randomized *flips*, *shifts*, elastic deformations and *shearing*, were applied. Transfer learning was used on all feature extractor models over the ImageNet Challenge database.[9] They resized all images to 2048 x 2048 pixels. It was proposed as input the four views images taken for patient (two MLO and two CC). RetinaNet model[13] was used for ROI segmentation and ResNet-v2-50[14] for global feature extraction. The authors added as a final feature to the classifier patient's age, and achieved an AUC of 0,88. They report that, after a comparative study between the AI system and six radiologists, in average, the algorithm over-performed the specialists.

Hai et al.[4] developed a model that used multiple levels features. They trained a CNN named S-DenseNet, that followed an architecture similar to DenseNet,[15] with four dense blocks. Each block weights were initialized with the ones from the ImageNet Challenge.[9] The neural network was used for high level semantic feature extraction. Also, local features based on gray levels and texture were extracted. All features passed through a selector algorithm named Lasso regression. Dense layers classifier took as input all the concatenated features, and predicts the results. The training images were collected from the Department of Radiology of the Henan Province Hospital, China, with 204 patients. Each ROI was extracted manually. Images were resized to 224 x 224 pixels. They achieved an AUC of 0,71. The authors claimed that multiple levels features extracted by CNNs and other methods could capture more significant information in mammograms and improve classification metrics.

In this paper, we present a breast mass malignancy classifier from digital mammographic images, based on deep neural networks with transfer learning and data augmentation techniques. Also, we applied and image pre-processing algorithm to mammograms. Further, we evaluated four different classifiers that used features extracted from a CNN.

METHOD

Some images on the CBIS-DDSM database have low resolution, bad contrast, and significant random noise. For this reason a pre-processing step was done by applying two consecutive filters like CLAHE[16] and Mean Blur.[17] EMs were resized to 256 x 256 pixels with nearest neighbor interpolation for case analysis, due to lack of computing power.[18] Each image was also normalized.

Based on the success of deep CNN architectures, this paper proposes an artificial neural network that used as feature extractor the convolutional blocks of the pre-trained network VGG16.[19] Figure 1 describes our model proposed architecture, named as VGG16-C300.

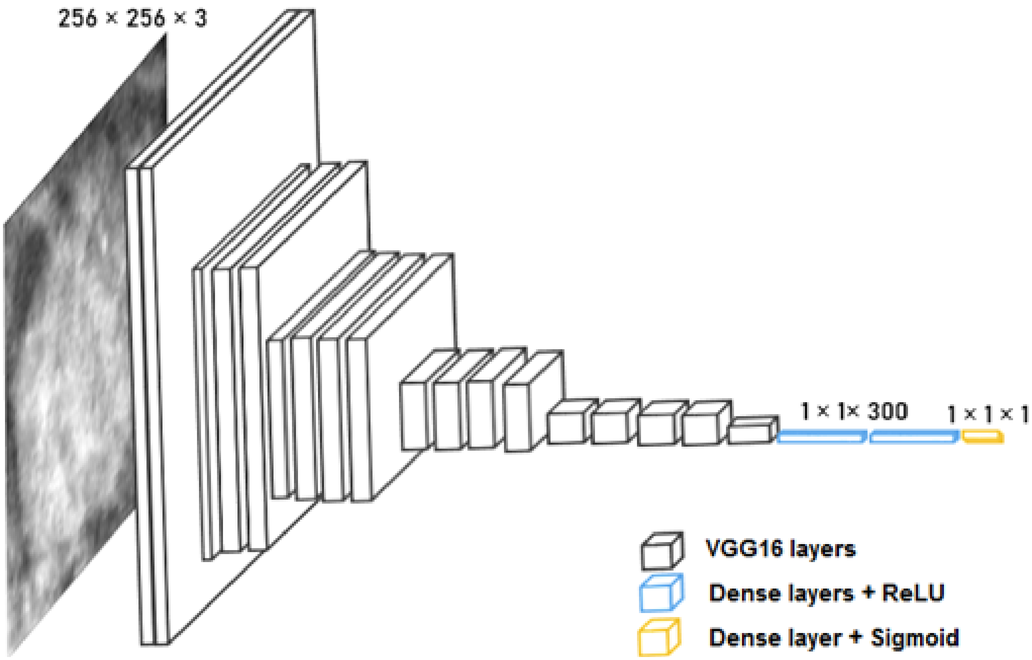


Figure 1. Proposed model proposed architecture named VGG16-C300. It contains VGG16's convolutional blocks, two dense layers randomly initialized with 300 units, and an output dense layer with one unit

To avoid overfitting, Dropout layers were added between dense layers, with a random parameter between 0,3 and 0,5 (following Lu, Loh, & Huang[20]). Dense layers have 300 units. During experiments, a higher number of units was tested, but the network tended to quickly overfit. For all layers, the activation function ReLU was used, but the last one, which used the sigmoidal activation function. VGG16-C300 outputs a number in the interval [0,1]. While prediction result closer to 1, there is a higher malignancy probability.

During VGG16-C300 training, data augmentation random techniques were applied to reduce overfitting. This helped to improve model generalization capacity. Some of the techniques applied were random rotations, horizontal and vertical *flips*, *shearing*, *zoom* and *shifting*.[21]

The transfer-learning technique was also used due to the relatively low number of training cases. The VGG16 convolutional blocks weights were initialized with the ImageNet trained weights.[9] Hence, the general problem of object detection was reduced to the breast cancer detection with mass anomalies.

Other classifiers were tested, like SVM (C = 2^5), RF (1 000 estimators), and *k*-NN (*k*=30). Models parameters were obtained after execute the algorithm Grid-Search 5-Folds. This classifiers took as input the VGG16-C300 second dense layer output. Given the low number of training cases, the algorithm 5-Fold Cross Validation was executed 10 times, so that, in each run, five different subsets were generated. For the training phase, 388 malignant and 498 benign cases were selected, summing 886 ROIs. The validation phase was run with a total of 223 ROIs, 98 malignant and 125 benign cases. The average of each fold prediction result was selected as prediction estimator.

Imaging Data

The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is a public updated standardized version of DDSM.[12] CBIS-DDSM includes selected and pre-processed images with a more precise segmentation of the ROIs. It contains 753 calcifications and 891 masses cases. Almost all cases have the ROI mask and a square that contains the ROI.

This research is centered on ROI masses, named as Extracted Mass (EM). After a manual process of image selection from EMs, 623 benign and 486 malignant were obtained, from a total of 1109 images. CBIS-DDSM images format is DICOM. All images were converted to PNG format to make easier the compatibility with programming frameworks.

Software and Hardware

The programming language used was Python 3.5.[22] For artificial neural network, we used the Keras framework[21] over Tensorflow with GPU.[23] Other models like SVM, RF and *k*-NN were trained and tested in the sklearn framework.[24] We used OpenCV in the Python framework for other image manipulation methods.[25] All experiments were executed over a PC i5 of 4th generation with 8GB of RAM, and with a NVIDIA GFORCE 1060 3GB as GPU. The code for training the VGG16-C300 network is available in GitHub (https://github.com/Ariel96cs/VGG-C300).

RESULTS

Figure 2 shows VGG16-C300's ROC curves before and after applying data augmentation techniques mentioned.

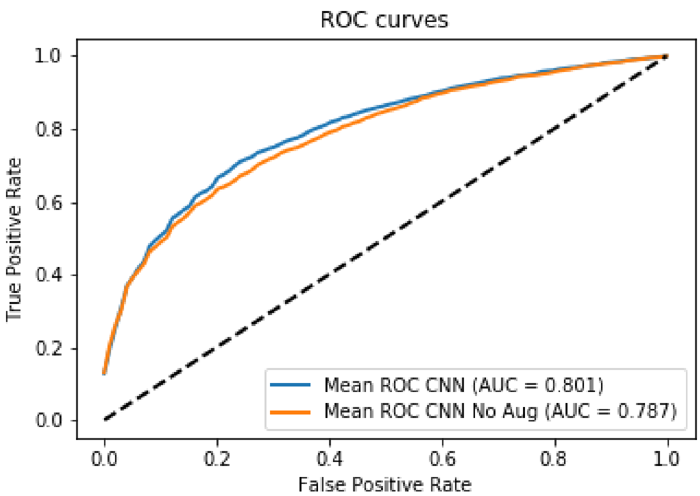
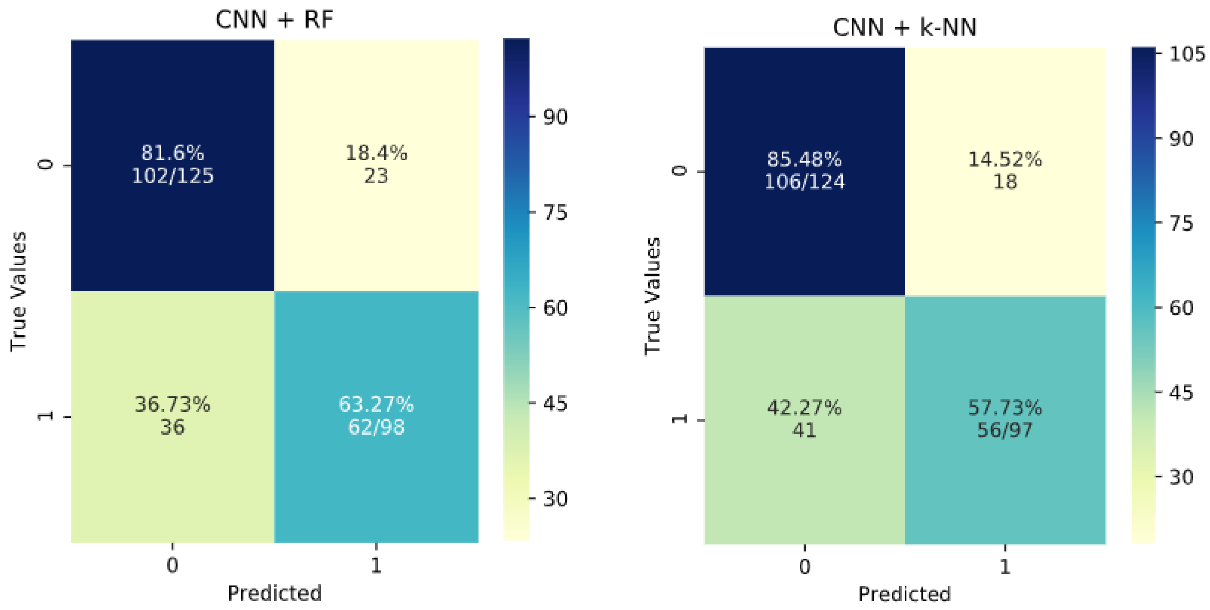
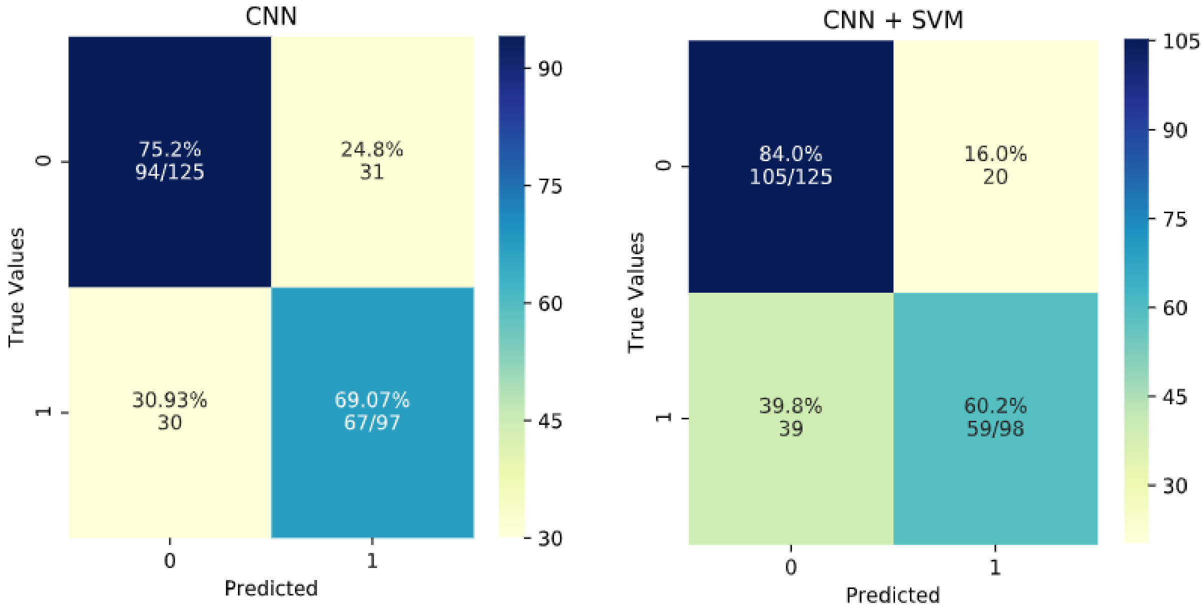


Figure 2. VGG16-C300's ROC curves with and without data augmentation

Given the models described, their confusion matrices were computed, as showed on figure 3. According to this results, benign cases were better detected by CNN + k-NN model, and malignant cases were better detected by CNN model (VGG16-C300). The model with the highest false malignant cases number was the CNN and the worst model detecting malignant cases was the CNN + k-NN. The CNN was selected as the best of the 4 models, because it can detected more malignant tumors than the others.



https://doi.org/10.56294/sctconf2025669



(c) (d)

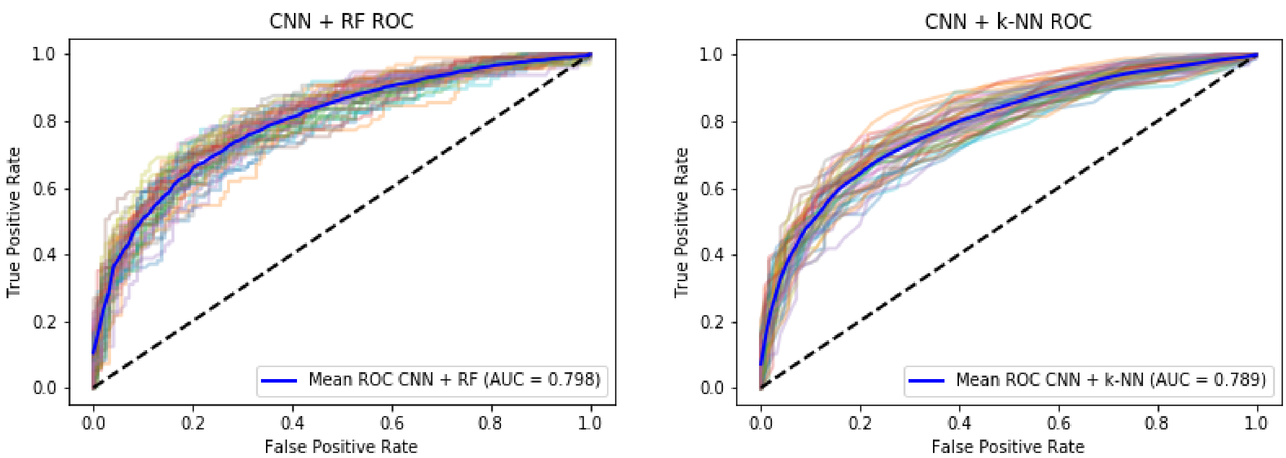
Figure 3. Five experimental models and its confusion matrices: (a) CNN + RF, (b) CNN + k-NN, (c) CNN and (d) CNN + SVM

Table 1 shows the obtained metrics by the trained models. The best results for each metric are written in bold. The CNN + k-NN model had the highest accuracy with 0,756. The CNN achieved the best recall and F1 with 0,691 and 0,688, respectively.

Table 1. Obtained metrics by the experimental models. The best results for each metric are written in bold

Models	Accuracy	Recall	F1
CNN + SVM	0,747	0,602	0,665
CNN + RF	0,729	0,634	0,677
CNN + k-NN	0,756	0,574	0,651
CNN	0,700	0,691	0,688

ROC curves were calculated for each experimental model after applying cross-validation. Figure 4 shows this curves and each model AUC average. It shows the differences between each model average ROC curve. After this results, we select CNN as the best model with an AUC of 0,80.



(a) (b)

https://doi.org/10.56294/sctconf2025669

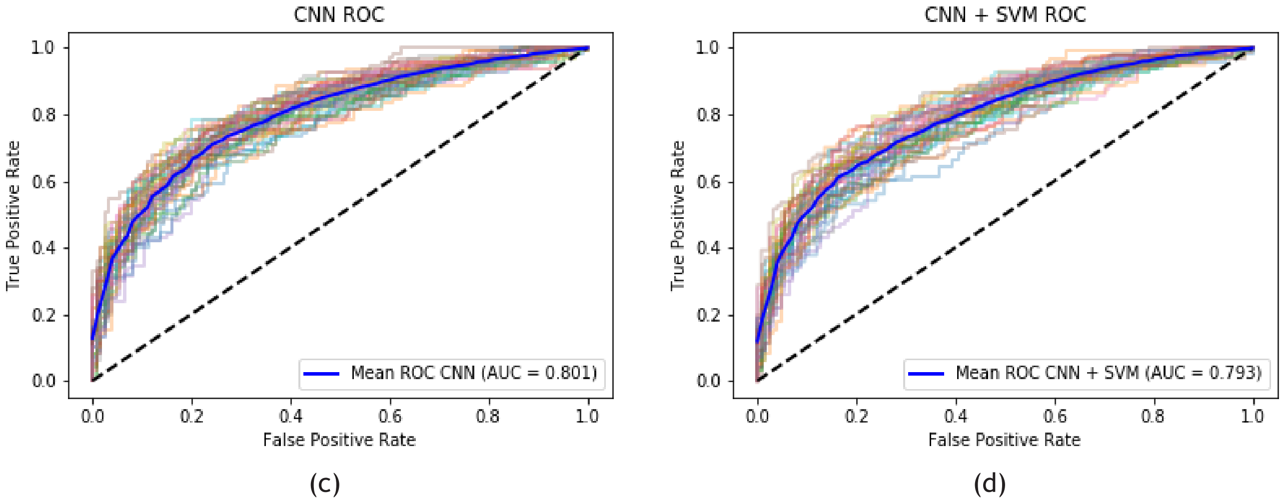


Figure 4. ROC curves of the trained models after 10 times 5-folds cross-validation: (a) CNN + RF, (b) CNN + k-NN, (c) CNN and (d) CNN +SVM

DISCUSSION

A comparative study between our model with others is difficult due to the differences between the variety of training datasets, the cases distribution, the mammograms characteristics (resolution, noise, contrast), and some patients characteristics (age, country, etc.). Also, the computer power used on each research varies. For example, McKinney et al.(5) processed more than 12 000 cases and reported an AUC of 0,88, S-DenseNet runs over 204 cases showed an AUC of 0,71,(4) and our model VGG16-C300 reported an AUC of 0,80.

One disadvantage of our model is that VGG16-C300 pre-trained weights are prepared to receive RGB images, but mammograms only have one channel (gray level images). This implies that the network extracts three times more features and leads to overfit more quickly. To deal with this problem it's necessary a larger training set.

CONCLUSIONS

This paper presents a deep neural network with the capacity of, given a ROI with a mass, and, after applying an image enhancement algorithm with two consecutive filters, predicts the anomaly malignancy probability, with an AUC of 0,80. Data augmentation and transfer learning techniques were applied, and improved the prediction performance, compared to shallow neural networks and traditional approaches. The potential usefulness of transfer learning for the task of breast cancer detection was shown, but on the other side, it forces us to be restrictive with the original architecture. Similarly, the use of data augmentation techniques lead to better model prediction results (Figure 4). Both techniques are widely use on the state of the art of the problem. Our model was limited to 256 x 256 pixels ROIs, and it is pretended to be included on a real time CADs for helping radiologist on the task of breast cancer detection, therefore, it needs an algorithm that can provide a ROI with a mass, and then it can calculate the anomaly malignancy probability.

REFERENCES

1. Dirección Nacional de Registros Médicos y Estadísticas de Salud. Anuario Estadístico de Salud 2019 [Internet]. La Habana: Ministerio de Salud Pública de Cuba; 2020. Available from: https://files.sld.cu/bvscuba/files/2021/08/Anuario-Estadistico-Español-2020-Definitivo.pdf

2. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. Radiology [Internet] 2019;292(2):331–42. Available from: http://pubs.rsna.org/doi/10.1148/radiol.2019182622

3. Rangayyan RM. Biomedical image analysis [Internet]. Calgary, Alberta, Canada: CRC Press; 2004. Available from: https://www.routledge.com/Biomedical-Image-Analysis/Rangayyan/p/book/9780849396953?srsltid=AfmBOor_VKv0-EBU0xaPSbkHWeBAhwfgGQanXO0wkrusw5QsRh5QKMrF

4. Hai J, Tan H, Chen J, Wu M, Qiao K, Xu J, et al. Multi-level features combined end-to-end learning for automated pathological grading of breast cancer on digital mammograms. Comput Med Imaging Graph [Internet] 2019;71:58–66. Available from: https://doi.org/10.1016/j.compmedimag.2018.10.008

5. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation

of an AI system for breast cancer screening. Nature [Internet] 2020;577(7788):89–94. Available from: https://www.nature.com/articles/s41586-019-1799-6

6. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. CoRR [Internet] 2017;abs/1707.0. Available from: http://arxiv.org/abs/1707.08401

7. Touahri R, AzizI N, Hammami NE, Aldwairi M, Benaida F. Automated Breast Tumor Diagnosis Using Local Binary Patterns (LBP) Based on Deep Learning Classification [Internet]. In: 2019 International Conference on Computer and Information Sciences (ICCIS). IEEE; 2019. 1–5.Available from: https://ieeexplore.ieee.org/document/8716428/

8. Jiang F, Liu H, Yu S, Xie Y. Breast mass lesion classification in mammograms by transfer learning [Internet]. In: Proceedings of the 5th International Conference on Bioinformatics and Computational Biology. New York, NY, USA: ACM; 2017. 59–62.Available from: https://dl.acm.org/doi/10.1145/3035012.3035022

9. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis [Internet] 2015;115(3):211–52. Available from: http://link.springer.com/10.1007/s11263-015-0816-y

10. Halling-Brown MD, Warren LM, Ward D, Lewis E, Mackenzie A, Wallis MG, et al. OPTIMAM Mammography Image Database: A Large-Scale Resource of Mammography Images and Clinical Data. Radiol Artif Intell [Internet] 2021;3(1):e200103. Available from: http://pubs.rsna.org/doi/10.1148/ryai.2020200103

11. Elangovan P, Hadjipanteli A, Mackenzie A, Dance DR, Young KC, Wells K. OPTIMAM Image Simulation Toolbox - Recent Developments and Ongoing Studies [Internet]. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2016. 668–75.Available from: http://link.springer.com/10.1007/978-3-319-41546-8_83

12. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data [Internet] 2017;4(1):170177. Available from: https://www.nature.com/articles/sdata2017177

13. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. IEEE Trans Pattern Anal Mach Intell [Internet] 2020;42(2):318–27. Available from: https://ieeexplore.ieee.org/document/8417976/

14. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition [Internet]. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016. 770–8.Available from: http://ieeexplore.ieee.org/document/7780459/

15. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks [Internet]. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017. 2261–9.Available from: https://ieeexplore.ieee.org/document/8099726/

16. Georgieva V, Kountchev R, Draganov I. An Adaptive Enhancement of X-Ray Images [Internet]. In: Kountchev R, Iantovics B, editors. Advances in Intelligent Analysis of Medical Data and Decision Support Systems. Heidelberg: Springer International Publishing; 2013. 79–88.Available from: https://link.springer.com/10.1007/978-3-319-00029-9_7

17. Gonzalez RC, Woods RE. Digital Image Processing [Internet]. 4th ed. New York: Pearson; 2018. Available from: https://elibrary.pearson.de/book/99.150005/9781292223070

18. Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size [Internet]. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). IEEE; 2015. 730–4. Available from: http://ieeexplore.ieee.org/document/7486599/

19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv Prepr arXiv14091556 [Internet] 2014;Available from: https://arxiv.org/abs/1409.1556

20. Lu HC, Loh EW, Huang SC. The Classification of Mammogram Using Convolutional Neural Network with

https://doi.org/10.56294/sctconf2025669

Specific Image Preprocessing for Breast Cancer Detection [Internet]. In: 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE; 2019. 9–12.Available from: https://ieeexplore.ieee.org/document/8837000/

21. Chollet F. Deep Learning with Python [Internet]. USA: Manning Publications Co.; 2017. Available from: https://www.manning.com/books/deep-learning-with-python

22. Van Rossum G, Drake FL. Python 3 Reference Manual [Internet]. Scotts Valley, CA: CreateSpace; 2009. Available from: https://dl.acm.org/doi/book/10.5555/1593511

23. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning [Internet]. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016. 2016. 265–83.Available from: https://dl.acm.org/doi/10.5555/3026877.3026899

24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res [Internet] 2011;12:2825–30. Available from: https://www.jmlr.org/papers/v12/pedregosa11a.html

25. Kaehler A, Bradski G. Learning OpenCV 3 [Internet]. O'Reilly Media, Inc.; 2016. Available from: https://www.oreilly.com/library/view/learning-opencv-3/9781491937983/

FINANCING

This research is part of the research project "Wavelets, frames, spectral techniques, partial differential equations and scientific machine learning in image analysis", National Basic Science Program P223LH010-036, 2024-2026.

CONFLICT OF INTEREST

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Damian Valdés Santiago.
Data curation: Ariel Coto Santiesteban.
Formal analysis: Ariel Coto Santiesteban, Damian Valdés Santiago.
Research: Ariel Coto Santiesteban, Damian Valdés Santiago.
Methodology: Ariel Coto Santiesteban, Lisbel Garzón Cutiño, Damian Valdés Santiago.
Project management: Damian Valdés Santiago.
Resources: Damian Valdés Santiago.
Software: Ariel Coto Santiesteban.
Supervision: Damian Valdés Santiago.
Validation: Ariel Coto Santiesteban, Damian Valdés Santiago.
Display: Ariel Coto Santiesteban.
Drafting - original draft: Ariel Coto Santiesteban, Damian Valdés Santiago.
Writing - proofreading and editing: Ariel Coto Santiesteban, Lisbel Garzón Cutiño, Damian Valdés Santiago.