UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

Detecting breast cancer lesion on mammography images

Juan Sebastián Chimbo Martínez

Ingeniería en electrónica y automatización

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en electrónica y automatización

Quito, 28 de diciembre de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

**HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA**

Detecting breast cancer lesion on mammography images

Juan Sebastián Chimbo Martínez

Nombre del profesor, Título académico	**Diego Benítez, PhD.**
Nombre del profesor, Título académico	**Noel Pérez, PhD.**

Quito, 28 de diciembre de 2021

© DERECHOS DE AUTOR

Nombres y apellidos: Juan Sebastián Chimbo Martínez

Código: 00201807

Cédula de identidad: 1723584270

Lugar y fecha: Quito, 28 de diciembre de 2021

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

El cáncer de mama es el tumor maligno más frecuente en las mujeres y la principal causa de muerte en los países desarrollados. En el año 2020 en Ecuador hubieron al menos 1033 muertes y su presencia anda en aumento, pero los diagnósticos tempranos han evitado que exista mayor mortalidad. Este tumor se presenta tanto como microcalcificaciones y masas, siendo estas últimas las más difíciles de detectar por un radiólogo. Por tal motivo, en este proyecto se presenta un método para detectar las masas en mamografías para su temprano diagnóstico aplicando aprendizaje profundo. La estructura propuesta para la detección y segmentación de las masas fue U-Net con ciertas modificaciones como la adición de más niveles y dropout en cada capa. El modelo fue entrenado con las bases de datos públicas de INBreast y DDSM, las cuales tienen las imágenes y anotaciones en donde se encuentra el tumor. Durante la evaluación del modelo se obtuvo un Dice Coefficient de 0.667 para las dos bases de datos. Los resultados obtenidos sobresalen en el aspecto de que se utilizan imágenes de alta resolución a diferencia de otras metodologías que llegan a comprimir la información de las imágenes al rescalar su tamaño hasta cuatro veces menor al tamaño original. A su vez, el modelo entrenado puede llegar a ser de utilidad para radiólogos para tener una segunda opción al momento de realizar diagnósticos.

Palabras clave: Redes neuronales convolucionales, Modelo U-Net, Cáncer de mama, segmentación de masas en mamografías.

ABSTRACT

Breast cancer is the most common malignant tumor in women and the leading cause of death in developed countries. In the year 2020 in Ecuador there were at least 1033 deaths and its presence is increasing, but early diagnosis has prevented further mortality. This tumor presents itself both as microcalcifications and masses, the latter being the most difficult to detect by a radiologist. For this reason, this project presents a method to detect masses in mammograms for early diagnosis by applying deep learning. The proposed structure for the detection and segmentation of masses was U-Net with certain modifications such as the addition of more levels and dropout in each layer. The model was trained with the public databases of INBreast and DDSM, which have the images and annotations where the tumor is located. During the evaluation of the model a Dice Coefficient of 0.667 was obtained for the two databases. The results obtained stand out in the aspect that high resolution images are used, unlike other methodologies that compress the information of the images by rescaling their size up to four times smaller than the original size. At the same time, the trained model can be useful for radiologists to have a second option when making diagnoses.

Key words: Convolutional neural networks, U-Net model, Breast cancer, mammogram mass segmentation.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

INTRODUCTION

Cancer happens when some cells in the body grow uncontrollably, affecting other cells and spreading through the body until creating tumors (National Cancer Institute, 2021). Each year the presence of cancer increases worldwide. In 2020, at least 19.2 million new cases were estimated, which has caused several health agencies to seek ways to reduce them (Sung et al., 2021). For women, the most diagnosed cancer was breast cancer. In 2020 there were at least 2.3 million confirmed cases and about 685,000 deaths globally (WHO,2021). In Ecuador, 3563 cases were diagnosed and about 1033 deaths during 2020 (Ministerio de salud pública, 2021). Breast cancer can be found in two primary forms, as masses according to their shape, orientation, margin (indistinct, angular, microlobulated, or spiculated), or as microcalcifications. The latter can be inside or outside a mass and present as an intraductal calcification (National Cancer Institute, 2021). The most commonly used technique to detect it is mammography, an X-ray image that doctors use to detect tumors that could not be felt, and microcalcifications that indicate the presence of breast cancer. On the other side, some potential harms of screening mammograms are: false-positive results, when there seems to be an abnormality but no cancer is present, over-diagnosis and over-treatment, when small cancers or Ductal Carcinoma in Situ are diagnosed as dangerous but ultimately never cause symptoms or do not affect the woman's life, and false-negative results when mammograms appear normal, but there is cancer present (National Cancer Institute, 2020).

American College of Radiology (ACR) created the late 1980s Breast Imaging Reporting and Data System® (BI-RADS®) in order to classify the types of tumors between benign and malignant and the effects they can have on the patient (Burnside et al., 2009). BI-RADS® consists of 7 categories, which are: 0 when the information is incomplete, 1 when there is a negative diagnosis of cancer, 2 when the tumor found is benign and does not affect

the patient's health, 3 when there is a high probability of benign tumor, 4 when it is suspicious of malignant tumor, 5 when it is very likely to be malignant, requires a biopsy, and finally, 6, when the tumor is malignant and proven by biopsy (Sickles et al.., 2013). When diagnosed in category 4 or 5, a problem arises, since radiologists cannot differentiate on mammography whether the tumor is benign or malignant because the form in which the tumor presents itself is not very easy to differentiate from the rest of the masses. This implies performing a biopsy on the patient, which is very invasive and can cause discomfort and pain. Some techniques such as image machine learning and deep learning (DL) have emerged to avoid this.

 Tsochatzidis et al. compared some DL models for breast cancer diagnosis where ResNet-50 and ResNet-101 performed well with an area over the ROI curve (AUC) ≈ 0.85 when initialized with pre-trained weights (fine-tuning). On the other hand, a relatively new DL structure called U-Net has been used by Tardy, M & Mateus, D to segment masses and micro-calcifications at a resolution of 1536x1536 pixels, obtaining a Sorensen-Dice Coefficient (DICE) of 0.58 in the validation dataset. Soulami et al. segmented and classified benign, malignant, and normal dense masses at low resolution, obtaining a DICE of 0.905. While Abdelhafiz et al. performed a new deep learning model called "Vanilla U-Net" that achieved a DICE of 0.909 for mass detection.

This project explores the use of CNN to detect masses in breast imagines trying to achieve higher or comparable DICE values than other works in terms of segmentation breast cancer tumors for masses in high-resolution images. When working with medical images, it should be taken into account that the least desirable is compression or reduction of the resolution of the image because of the loss of details that become very important in the detection of masses that are not too big. Being able to segment correctly with high-resolution

images allows future work to focus on classification according to tumor type with BI-RADS®

applying artificial intelligence.

MATERIALS AND METHODS

Database

Two datasets were used for this research: first INBreast, a database for mammographic research created at the Centro Hospitalar de S. Joao (CHSJ) in Porto, Portugal; its images were adquired between April 2008 and July 2010 from anonymous patients in the region. Second, the Digital Database for Screening Mammography (DDSM) was used; this was designed with The Massachusetts General Hospital, The University of South Florida, and Sandia National Laboratories, among other collaborators (Rose et al., 2008). The purpose of both datasets is to provide the information required to develop programs that serve as CAD (computer-aided design) and the creation of algorithms to assist as second criteria for radiologists.

According to Moreira's, et al. technical report, INBreast consists of 410 mammograms, including masses, calcifications, asymmetries, and distortions in DICOM format, exactly 116 masses in 107 images and 6880 calcifications in 299 images. XML files contain the annotations made by a radiologist, which were validated by a second specialist, detailing the region where they found: asymmetries, masses, cluster, calcifications, distortions, spiculated region, and pectoral muscle. On the other hand, in Rose's et al. 2008 report, DDSM is classified by cases, which is the information of a single patient with mammographic images, and by volumes, which is the collection of cases with common characteristics. There are 12 volumes of normal cases, 15 of malignant cancer, and 14 of malignant cases, each with annotations and segmentation masks. 1658 images and mass masks of this dataset were used for this research.

Deep learning model

The implementation of deep learning began as a need to generate models that can learn how humans interpret the environment. Its focus is primarily on computer vision, which aims at image and video recognition, image analysis and classification, and language processing. The algorithm that has emphasized the development of deep learning has been the Convolutional Neural Network (CNN). This is in charge of receiving an input image, assigning weights and biases based on the convolution with filters that obtain the main characteristics of the image that differentiate it from others. The architecture of this model is analogous to how neurons work in human's brain, each neuron responds to a specific stimulus, and by joining them together, they recreate the entire visual área (Tatan, 2019). The neural network comprises the following processes:

- Convolution layer, which convolves the image in strides with a specific filter.

- Image padding to manage the borders.

- Pooling layer, which reduces the spatial size of each feature taken from convolutions, and the fully connected layer, which classifies each class.

It has been possible to build different types of deep learning models from this network to achieve different results.

The framework used for training the breast cancer segmentation model was U-Net. This topology was presented by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015 at the International Conference of Medical Image Computing & Computer Assisted Intervention (MICCAI), and its purpose was for biomedical image segmentation, achieving good results in terms of intersection over unión (IoU), processing time, and a smaller amount of training data compared to other structures such as ImageNet. Its architecture is shown in figure 1,   which consists of two main stages. The contraction path (encoder), which is similar to CNNs, where two 3x3 convolutions with the activation by Rectified Linear Unit (ReLU) are performed for

each level and 2x2 max pooling is used between levels, doubling the number of channels. Furthermore, the expansion path (decoder) where 2x2 up-convolution and two convolutions with ReLU activation are used where the number of levels is halved. In addition, for each upsampling level, the up-convolution feature maps are concatenated with their respective downsampling feature from the encoder, which allows preserving the location of the pixels by the edges obtained in each convolution in the downsampling. Once the encoding and decoding process is finished, a 1x1 convolution is performed with the required channels according to the classes (1 channel for binary images), giving a binary mask with the segmented desired by the trained model.
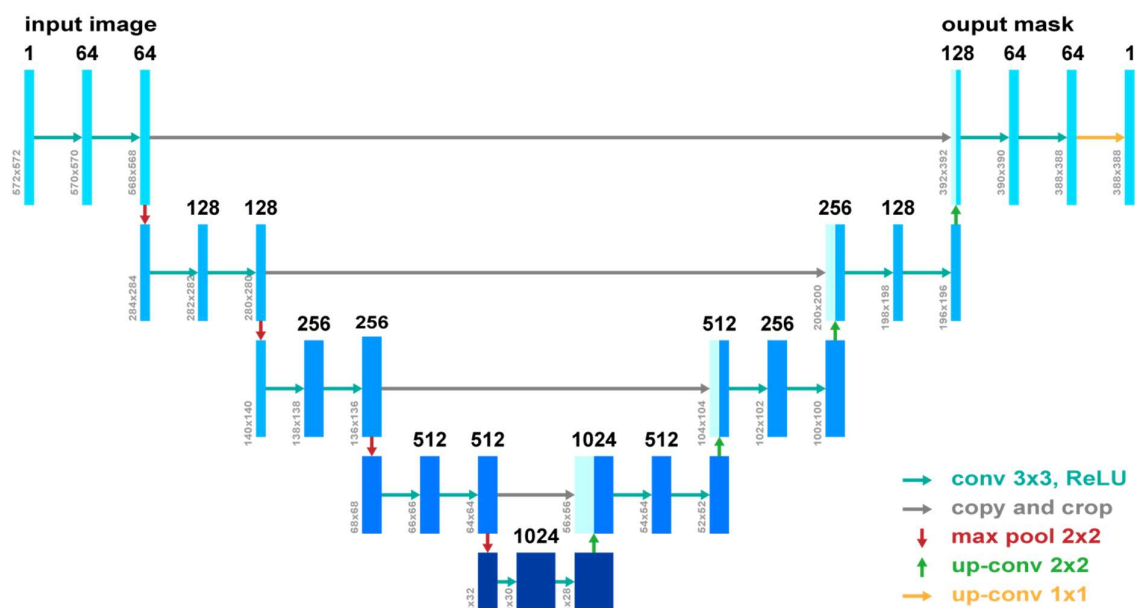

Figure 1 U-Net original architecture level described

Proposed method

Medical imaging uses high resolution to preserve the smallest detail of the mammogram. In turn, the model to be implemented requires the same amount of pixels for both height and width. Because the loss of information of small masses is not wanted, a

rescaling of all images and maks to a resolution of 1024x1024 pixels was performed. In the original U-Net model, lower resolution images are used with good results with the proposed model structure. On the contrary, two additional levels were added to the encoder and decoder stage for this research to obtain more features due to the high-resolution images.

In this model, ReLU was maintained, the size of the convolutional layers of 3x3, strides of 2x2, and there is only one class to be segmented. The last convolution is 1x1 with activation by sigmoid. This proposed method adds dropouts in each convolutional layer of downsampling and upsampling with values between 0.1 and 0.5 to avoid overfitting.

Adding 7 layers in the structure led to a number of 4096 feature maps channel in the last layer, being 12x12 the lowest resolution, giving 497668417 trainable parameters. Figure 2 shows the implemented U-Net structure.
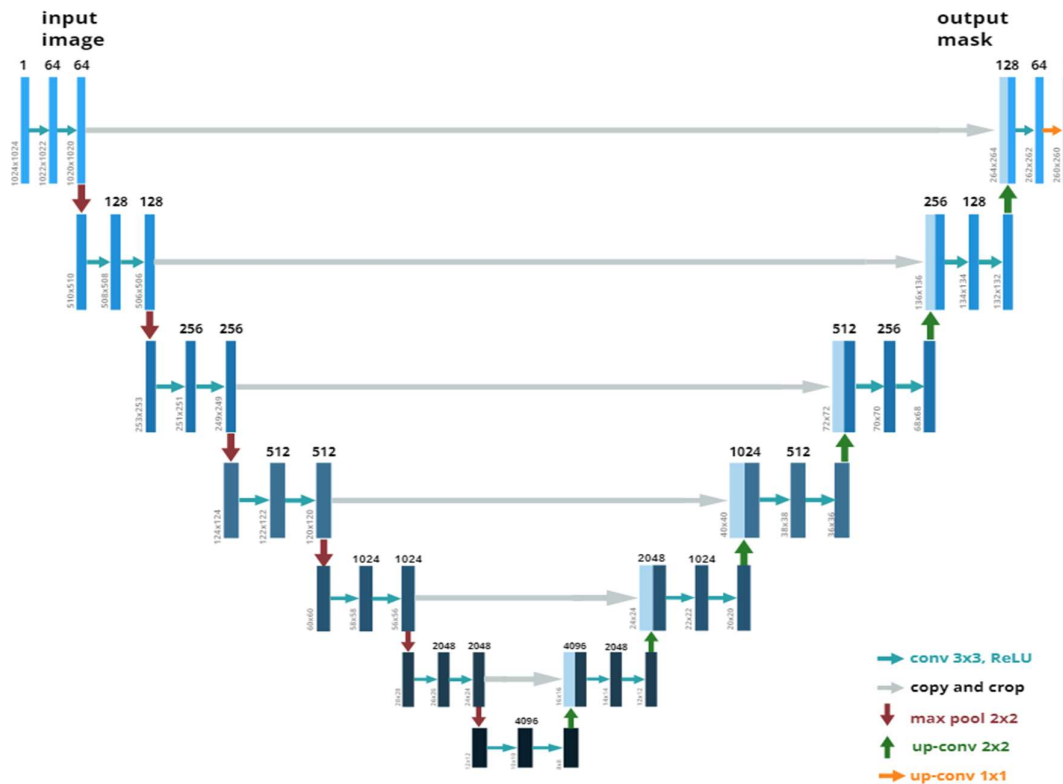

Figure 2 U-Net proposed architecture level described

Experimental setup

Data processing

The whole project was developed with version 3.6.9 of the Python programming language (Van Rossum & Drake, 2009) and tensorflow version 2.3.0. (Abadi et al., 2015). Certain modifications were made to the images for the databases used to facilitate the algorithm's processing and training. First, it was necessary to convert the images in DICOM (format used in medicine that has both the image and the patient information) format to .tiff, which is a non-compression format, using a library called pydicom (Mason et al. 2021). Once the images were available, noise reduction was used with Gaussian Blur, Otsu thresholding and finding contours of cv2 were then applied to cut out the black areas of both the images and the masks to have the most relevant information about the breast, all of these functions were taken from open cv version 4.5.4 library (Alekhin et al., 2021). After these, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used. This type of equalization sweeps the image along small sections, it calculates the histogram for each of them to obtain an average for each tile to redistribute the values with higher luminance to those with lower luminance (Pizer et al., 1987). Finally, to prepare the images and masks for training the U-Net model, they were converted into a NumPy array with Numpy library version 1.19.5 (Harris, 2021), rescaled to a resolution of 1024x1024 with Pillow version 8.4.0 (Clark, 2015), and normalized from uint 16 to float 64 because the model to be trained requires that all images have the same size and that the arrays have values between 0 and 1. In addition, data augmentation (DA) with elasticity, brightness change, and rotation transformations was performed using Albumentations library from Buslaev et al. on the IN Breast dataset due to the small number of mass images available, from this, a total of 2060 images and masks were obtained.

Training and test sets

After The sklearn train test split library was used to distribute the images of the datasets in training and validation (Pedregosa, 2011). Of 2060 images with their respective masks, 10% correspond to validation sets and the remaining to training data. On the other hand, 166 random images and masks were separated from the INbreast and DDSM dataset to be used as test data to perform the predictions with the trained model.

Assessment metrics

In order to validate the segmentation prediction during training, Tiu E., 2019 recommends three main metrics as follows. The first is pixel accuracy, which detects the pixels that match the original mask or Ground Truth (GT) and the prediction mask (PM).

The second metric was Intersection Over Union (IoU) or Jaccard's index. This index is calculated by dividing the area of overlap of PM with GT by the area of the union of PM with GT. The IoU value is 0 when PM is not equal to GT and 1 when PM is equal to GT. With the definition of IoU, a Loss function can be made by subtracting one minus IoU, thus, the error has to go down as a function of how much IoU increases.

The last metric implemented was the Dice score or F1 score, which consists of multiplying by two the area of overlap of GT and PM and dividing them by the total number of pixels of the two masks. In the same way that in IoU, its values range from 0 to 1. In this case, it was applied only for the test set because callbacks were used to save the model with the highest metric, where only one was allowed to be used.

The difference between IoU and DICE is that IoU tends to quantitatively penalize individual cases of misclassification, even when both can agree that the prediction is wrong. In a few words, Dice is more likely to measure the average performance, while the IoU measures the worst-case performance during prediction (Willem, 2017).

Model configuration

500 epochs were used for training, where the model with the highest IoU value during validation was saved because, at each epoch, it monitors how correct the prediction is compared to the original mask for the training set and the validation set. A batch size of 4 was also used because higher values were not enough in the memory of the video card. As for optimizer, the adaptive moment estimation (Adam) was used, which was presented by Kingma and Ba in 2014, as an algorithm for first-order gradient-based optimization of stochastic objective functions. Adam is a good optimizer when having large datasets and converges in less time, and uses less memory than other algorithms. For the model, a learning rate of 0.0001 was used, with dropouts with different probability percentages in each U-Net layer.

RESULTS AND DISCUSSIONS

In order to evaluate the segmentation performed by the U-Net model modified for experimentation, the metrics of the training and validation set were obtained as well as for the 166 images outside the training set for testing. The values of Accuracy, Loss, IoU, and Dice Metric are necessary to understand the performance of the model for the detection of benign and malignant breast cancer masses.

Perfomance evaluation

During the training phase, about 2060 images and masks were used, resulting in 0.998 accuracy for the training and validation data. Despite showing an ideal result, this metric was not entirely reliable for segmentation, because it considers all the pixels of the masks and indicates when they match across the image. The problem is that black pixels are present in most of the GT and PM, and white pixels are in the minority, making the accuracy to reach high values but resulting in an incorrect prediction. Figure 5 shows the development of ACC for training and validation. For the reasons mentioned above, the IoU metric and Dice Loss were used to show how well the model was trained and whether it could generalize the results to images other than the training images. For the loss function in the train set, it reached 0.051. In the validation set, it was 0.263, as shown in Figure 6, which indicates that the error for the latter does not reach values of zero. Therefore the weights of the model to be trained do not fit to generalize the segmentation in a significant way, which affects the IoU value calculated as explained below. In terms of IoU, the following results were obtained for the training, validation, and test set: 0.920, 0.656, 0.566, respectively. The evolution in each epoch is shown in figure 7. In the case of the last one, it was calculated with the IoU equation from the predicted masks. In addition to obtaining these values, the Dice Coefficient was calculated to obtain a more realistic value of GT and PM. The average Dice obtained was 0.661. The reason for these

results is linked to the data imbalance since two datasets with different formats and image quality are mixed, the rescaling to a resolution almost half the average of the original images due to memory limits, and finally, the complexity of the U-Net architecture since it needs a more considerable amount of training data to avoid overfitting. Figure 3 shows correct predictions, and figure 4 shows incorrect predictions. This was done by running the test images through the algorithm, and the output of both the ground truth and the predicted mask was shown.
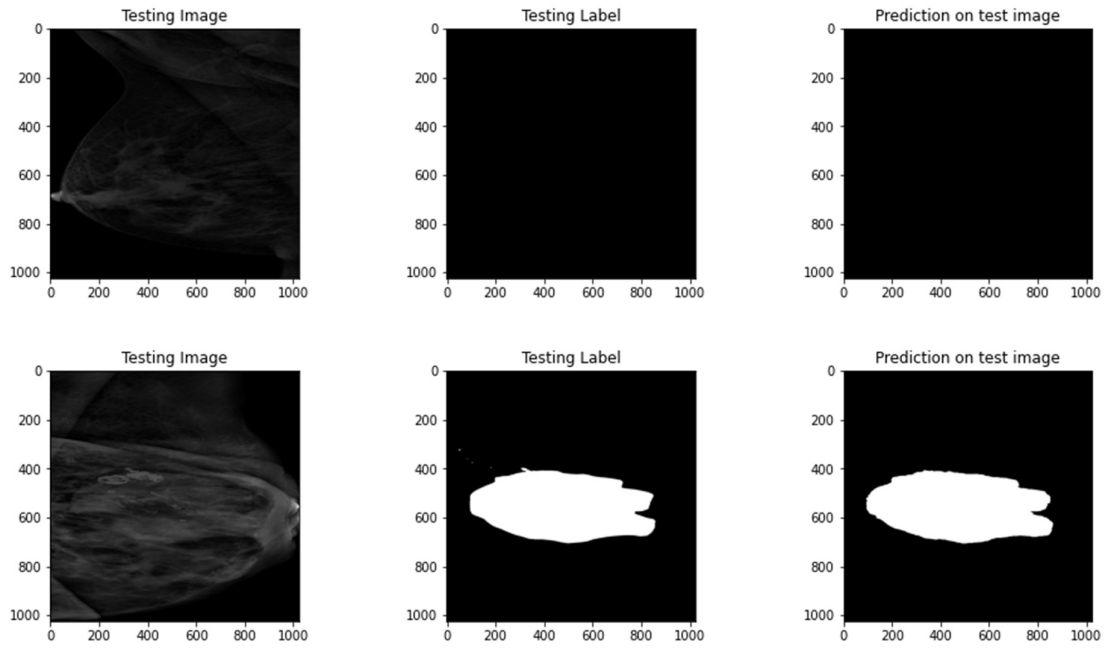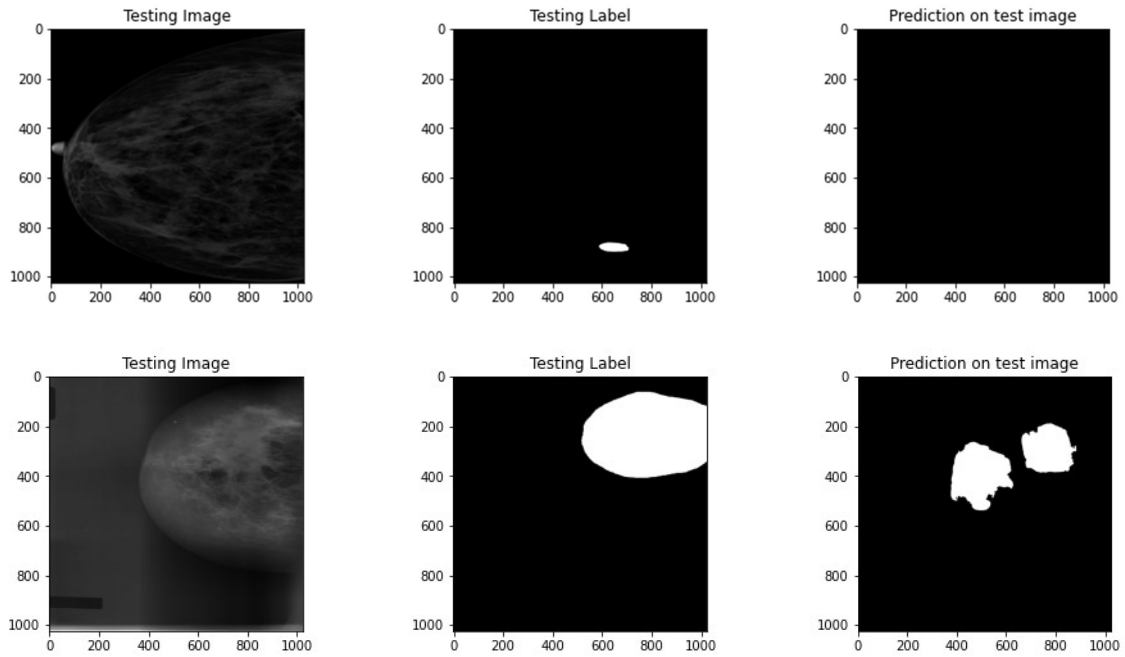
Figure 3 Correct prediction in test images

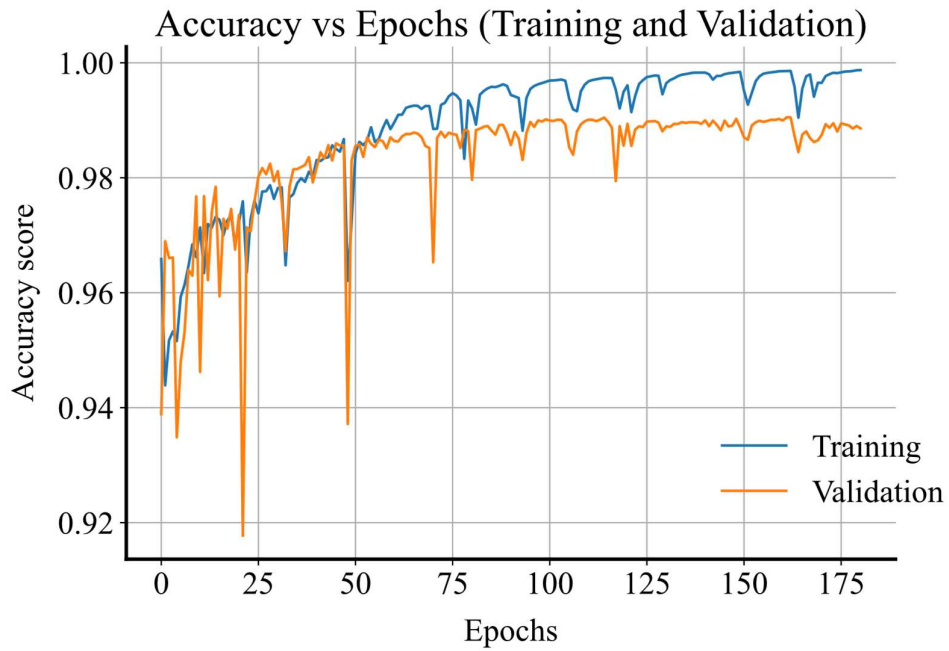*Figure 4 Incorrect prediction in test images*

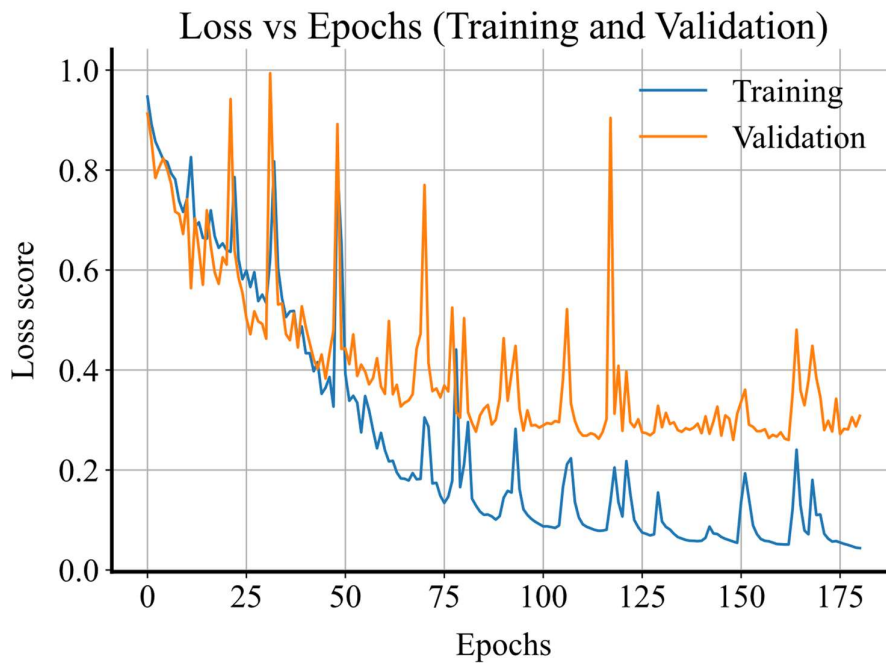*Figure 5 Performance of Accuracy during training and validation phase for U-Net proposed architecture*

Figure 6 Performance of Loss during training and validation phase for U-Net proposed architecture
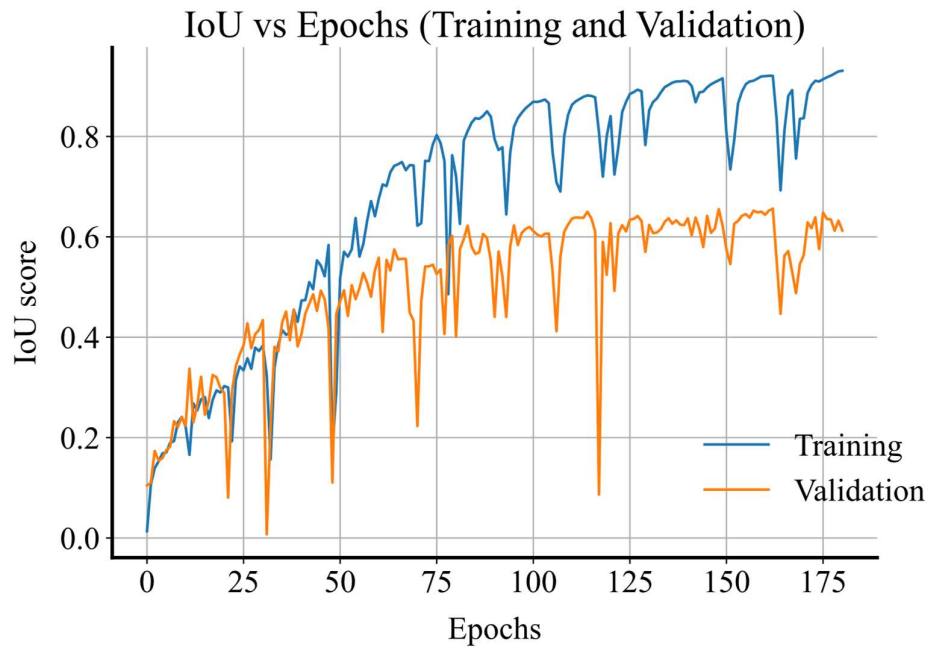
IoU vs Epochs (Training and Validation)

Figure 7 Performance of IoU during training and validation phase for U-Net proposed architecture..

State of art based comparison

The results make it possible to compare this research with other works. Table 1 summarizes the results of Dice Metric obtained from other studies, the database used, the type of lesion, the number of images, the resolution of the images retrieved, and the configuration of the model. Firstly, in comparison to the research performed by Tardy & Mateus, our research is superior in terms of DICE. However, it should be taken into account that Tardy & Mateus segmented masses and microcalcifications, while our model only segments masses. Secondly, the results of Soulami et al. are superior, this may be because rescaling the image to a very small size such as 128*128 and 256*256 causes the loss of smaller masses and therefore facilitates the algorithm to detect big masses, they use less images than our project but there is no information if used data augmentation. Third, the results of Zheng et al. are also high. In this case, they rescaled to 512*512 images. They obtained a high Dice metric in the same way

as the previous author, and this is because they used a more complex U-Net structure by mixing it with the convolutional neural network "VGG-16" in the encoding stage. Also, they trained with four different datasets, so they had ten times more images and masks which helped the algorithm perform better. Their proposed method can serve as an input to further improve the results obtained during this research and at the same time to experiment with higher image resolution.

	Database	Lession	# Images + DA	Image resolution	U-Net structure	Dice Metric
Tardy, M & Mateus, D	INbreast	Masses and Microcalci-fications	410	1536*1536	7 layers	0.58
Soulami et al.	DDSM, CBIS-DDSM & INbreast	Masses	1079	128*128 & 256*256	5 layers with multiclass rgb	0.905
Zheng et al.	CBIS-DDSM, INbreast, UCHCDM & BCDR-01	Masses	20660	512*512	Vanilla U-Net 6 Layers VGG-16	0.909
Proposed Method	DDSM & INbreast	Masses	2060	1024*1024	7 layers	0.661

Table 1 Comparison based on Dice Metric between different U-Net trained research.

CONCLUSIONS AND FUTURE WORK

In this project, a deep learning model capable of performing mass segmentation for breast cancer detection was developed based on the U-Net architecture. The modification made to the architecture was to add two more depth levels and dropouts in each convolution. This allowed us to obtain a better Dice metric of 0.667. Compared to other researchs, the trained model works on high-resolution images, which is important since it is beneficial for a radiologist to have as much information as possible about the lesion to make a correct diagnosis. Due to computational limitations at the moment, it was not possible to train the model with resolutions closer to the original images because the higher resolution requires more video memory (VRAM) to store and process the images for training. On the other hand, for future work adding databases would allow the algorithm to generalize mammograms even further and avoid overfitting. Also the increase in resolution and the modification of this model with the strategy used in "Vanilla U-Net" is considered, these may improve the algorithm in general because it would have more context of the lesion and more parameters to be trained. At the same time, it could be trained by classes, so in addition to segmentation, it could classify the type of lesion between benign and malignant and thus be able to locate masses and microcalcifications. Finally, the current model can be helpful for radiologists to localize masses because they will have a second opinion on the position of a tumor, which can avoid biopsy and thus improve the quality of life of women both in Ecuador and the world.

ACKNOWLEDGMENT

REFERENCES

Alekhin, A., Egor, S., Talamanov, A., Ivanov, S., Panov,A., Turkmen, S., Phurrough, D., Payne, G., Bareeva, J., Levin,V., Budnikov, D., Liutong, H., Smorkalov, A., Smirnov A., Petrogalli, F., Ahmed, I., Vautherin, J., Pashchenkov, M., Halik, S., …, Yin, Z. (2021). Open Source Computer Vision Library. Retrieved 5 October 2021, from https://opencv.org/opencv-4-5-4/

Abadi, D., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G. Isard, M. Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., … Zheng, X. (2015). Tensorflow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abdelhafiz, D., Bi, J., Ammar, R., Yang, C., & Nabavi, S. (2020). Convolutional neural network for automated mass segmentation in mammography. BMC Bioinformatics, 21(S1). doi: 10.1186/s12859-020-3521-y

Burnside, E., Sickles, E., Bassett, L., Rubin, D., Lee, C., & Ikeda, D. et al. (2009). The ACR BI-RADS® Experience: Learning From History. Journal Of The American College Of Radiology, 6(12), 851-860. doi: 10.1016/j.jacr.2009.07.023

Buslaev, A., Parinov, A., Khvedchenya, E., Iglovikov & I., Kalinin A. (2018). Albumentations: fast and flexible image augmentations. ArXi e-prints doi: 1809.06839

Clark, A. (2015). Pillow (PIL Fork) Documentation. readthedocs. Retrieved from https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf

D´Orsi, C., Sickles, E., Mendelson, E., Morris, E., & et al. (2013). Mammography. In: ACR BI-RADS® Atlas. Reston, VA: American College of Radiology. Retrieved from https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads

Harris, A., Millman, J., Van der Walt, J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M., Brett, M., Haldane, A., Ferrnández del Río, J., Wiebe, M., Peterson, P., Gérard-Merchant, P., …, Oliphant, T. (2021). Array programming with NumPy. Nature 585, 357-362. doi: 10.1038/s41586-020-2649-2.

Kingma, D., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. arXiv.org. Retrieved 24 October 2021, from http://arxiv.org/abs/1412.6980.

Mason, D., Suever, J., Lemaitre, G., Panchal, A., Rothberg, A., Massich, J., Kerns, J., Van Golen, K., Herrmann, M., Robitaille, T. Biggs, S., Bridge, C., Shun-Shin, M., Conrad, B., Mattes, M., Lyu, Y., Morency, F., Meine, H., Wortmann, J., Hahn, K., …, Rachum, R. (2021). Pydicom: an open source DICOM library. Retrieved 5 October 2021, from https://zenodo.org/record/5294594#.YcPtWepBxD9

Ministerio de salud pública. (2021). Cifras de Ecuador – Cáncer de Mama. https://www.salud.gob.ec/cifras-de-ecuador-cancer-de-mama/.

Moreira, I., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M., & Cardoso, J. (2012). INbreast: Toward a Full-field Digital Mammographic Database. Academic Radiology, 19(2), 236-248. doi: 10.1016/j.acra.2011.09.014

National Cancer Institute. (2020). Mammograms. Retrieved 18 October 2021, from https://www.cancer.gov/types/breast/mammograms-fact-sheet

National Cancer Institute, N. (2021). What Is Cancer?. Retrieved 9 December 2021, from https://www.cancer.gov/about-cancer/understanding/what-is-cancer

Pizer, S., Amburn, E., Austin, J., Cromartie, R., Geselowitz, A., & Greer, T. et al. (1987). Adaptive histogram equalization and its variations. Computer Vision, Graphics, And Image Processing, 39(3), 355-368. https://doi.org/10.1016/s0734-189x(87)80186-x

Pedregosa, F., Varoquaux, Gael, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 2825–2830.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes In Computer Science, 234-241. doi: 10.1007/978-3-319-24574-4_28}

Rose, C., Turi, D., Williams, A., Wolstencroft, K., & Taylor, C. (2008). DDSM: Digital Database for Screening Mammography. Retrieved 13 November 2021, from http://www.eng.usf.edu/cvprg/mammography/database.html

Soulami, K., Kaabouch, N., Saidi, M., & Tamtaoui, A. (2021). Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. Biomedical Signal Processing And Control, 66, 102481. doi: 10.1016/j.bspc.2021.102481

Sung, H., Ferlay, J., Siegel, R., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal For Clinicians, 71(3), 209-249. doi: 10.3322/caac.21660

Tardy, M., & Mateus, D. (2020). Lightweight U-Net for high-resolution breast imaging.

Tatan, V. (2019). Understanding CNN (Convolutional Neural Network). Retrieved 14 December 2021, from https://towardsdatascience.com/understanding-cnn-convolutional-neural-network-69fd626ee7d4

Tiu, E. (2019). Metrics to Evaluate your Semantic Segmentation Model. Medium. Retrieved 3 November 2021, from https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2#:~:text=1.,image%20that%20are%20classified%20correctly.

Tsochatzidis, L., Costaridou, L., & Pratikakis, I. (2019). Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study. Journal Of Imaging, 5(3), 37. doi: 10.3390/jimaging5030037

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

WHO. (2021). Breast cancer. Retrieved 9 November 2021, from https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Willem. (2017). F1/Dice-Score vs IoU. Retrieved 12 December 2021, https://stats.stackexchange.com/q/276144