Transfer Learning in Breast Mammogram Abnormalities Classification With Mobilenet and Nasnet

Lenin G. Falconí[1], María Pérez[2], Wilbert G. Aguilar[3]
[1,2,3]Facultad de Ingeniería de Sistemas
Escuela Politécnica Nacional
Quito, Ecuador 170525
{lenin.falconi, maria.perez, wilbert.aguilar}@epn.edu.ec
[2]*Member, IEEE*

***Abstract*—Breast cancer has an important incidence in women mortality worldwide. Currently, mammography is considered the gold standard for breast abnormalities screening examinations since it aids in the early detection and diagnosis of the illness. However, both identification of mass lesions and its malignancy classification is a challenging problem for artificial intelligence. Research has turned to the use of deep learning models in mammography which can enhance the performance of Computer Aided Diagnosis Systems (CADx). In this paper, we present our preliminary results on the use of transfer learning for malignancy classification of breast abnormality. We experiment with models that, according to our literature review, have not yet been explored thoroughly such as NasNet and MobileNet. Their performance is compared with InceptionV3 and Resnet50. The best results were obtained with Resnet50 and MobileNet with 78.4% and 74.3%, respectively. Also, some image pre-processing steps are studied in order to increase classification accuracy.**

***Keywords*—deep learning, digital mammogram, image pre-processing, machine learning, transfer learning**

I. Introduction

Breast cancer is one of the main causes of death among women and its incidence is expected to increase during the next decade. According to the International Agency for Research on Cancer (IARC), this illness is the second most frequent form of cancer among women worldwide with 2,088,849 (11.6%) new cases and 626,679 (6.6%) of deaths [1]. Early detection of breast cancer has proven to be successful in lowering the death risk rate [2].

As of today, mammography is considered the gold standard for screening examination [3]–[5]. Although direct digital mammography (Full Field Digital Mammography) has improved the sensitivity of the method, especially in dense breasts, the number of false negatives (FN) is still high, largely due to the presence of dense tissue that may affect lesions conspicuity. Other modalities have been developed such as ultrasound, magnetic resonance, tomosynthesis, infrared thermography [6] [7]. Process for detection is done manually and it is related, to radiologist's experience which shows that 10% of all woman screened for cancer are called back for additional testing and just as little as 0.5% of them are diagnosed with breast cancer [4]. This background shows that it is important to design CAD systems to aid specialists and train new ones in breast lesions detection. Recent developments in image analysis using Deep Learning based neural networks have motivated researches to enhance the performance of Computer Aided Diagnosis systems (CADx).

However, deep learning performance relies on training with large labelled datasets [8]. Unfortunately, mammogram public datasets are not "deep enough". In this context, transfer learning is a deep learning technique that can aid the development of accurate enough classifiers by transferring knowledge from another domain where large datasets are available. Several convolutional neural networks models such as: Alexnet [8], VGG [9], GoogLeNet [10], Resnet [11], Inception [12], [13], and others have been developed in the last years to solve visual classification tasks in natural images. Also researchers have began to explore the use of transfer learning of these models to mass pathology classification.

In this work, we perform a literature review process to find currently used pre-trained models and discover models that have not yet been tested in the field. Then we experiment with transfer learning on these new models found to measure their accuracy in classifying breast mammogram abnormalities in benign and malignant. The dataset that our proposed method uses for training and testing correspond to the Region of Interest (ROI) mass images from the public dataset of CBIS-DDSM [14]. Our research results indicate that MobileNet and NasNet have not yet been used. For comparison we also train InceptionV3 and Resnet50 which are found to have been used in literature.

The remainder of this paper is organized as follows: In Section II, we perform a literature review of the use of transfer learning in mammogram abnormalities classification. Our proposed experimental method, dataset and model is presented in Section III. Section IV presents experimental data results. Finally, discussion and future works are presented in Section V and VI, respectively.

978-1-7281-3227-3/19/$31.00 ©2019 IEEE

II. Related works

As of today, deep learning models have achieved top accuracy scores in computer visual tasks [15]. This has motivated research in the use, adaptation and application of such models in medical image analysis [16]. One of the common problems in this domain is the lack of large labeled training data sets. This problem is addressed by transfer learning (TL), where a high performance learner in a target domain is created from a related source domain [17]. This is possible since TL allows to use different domains, tasks, and distributions in training and testing [18], making it possible to improve the target learner by transferring information from a related domain [17]. In fact, according to Yosinski et al. [19], the first layer features of CNN trained on natural images resembles a Gabor filter [19]. Because of this, it is reasonable to expect to use a pre-trained CNN in a natural domain to be transferred to a medical image. In this work we examine the use of transfer learning in the classification of mammogram abnormalities.

A. Study Selection Process

In order to find relevant scientific literature related to our study, we performed a systematic literature research based on Kitchenham's methodology [20] that we aim to publish as a future work. The search string used to obtain all possible existing literature is formed of the following terms:"breast cancer" **AND** ("classification" **OR** "detection" **OR** "prediction") **AND** ("ensemble learning" **OR** "transfer learning") **AND** mammo*. A total of 173 studies were gathered for analysis from electronic scientific repositories such as: Springer Link, Science Direct (Elsevier), IEEE Xplore, Scopus, Web of Science, ACM digital library, and PubMed. From these studies, a total of 40 primary documents were retrieved according to a selection study process, where documents should have experimental methodology with results regarding the use of transfer learning in mammogram breast cancer classification.

B. Literature Review Discussion

A traditional machine learning approach for mammogram classification is shown in Fig.1. It has 5 stages [5]: image pre-processing, image segmentation, feature extraction, classification and performance evaluation. The core of this approach relies in the feature extraction step. For instance, Pérez et al [21] uses Haralick texture descriptor to get the feature vector of the image. On the other hand, deep learning through CNN helps to solve the problem of image feature representation, because a CNN synthesizes its own feature extractor [22]. However, successful deep learning models such as AlexNet [8] need to be trained on large datasets. This is a problem for mammogram images where there is not enough labelled data to train a CNN from scratch. Because of that, research is focused in using transfer learning in order to improve a mammogram abnormality classifier by transferring knowledge from a pre-trained CNN. In literature, we have found that TL could be used as a feature extractor or as fine tuned CNN.

1) Transfer Learning as a Feature Extractor: In this case, the mammogram image is feed to the pre-trained CNN and a feature vector is extracted from one of the layers of the model. Huynh et al [23] trains a support vector machine (SVM) based on the feature vector extracted from the full convolutional layer of an AlexNet pre-trained model. In the case of Huynh et al. the SVM is used because the training set is very limited.

2) Fine Tuning: This is the most common technique found in literature. In this case, only some of the last layers of the model are re-trained with the new data. For example, Chougrad et al. [24] uses VGG16 [9], InceptionV3 [13], and ResNet50 [11]. Also, the author indicates that when the number of convolutional blocks exceeds 2, the accuracy of the fine tuned model drops. In a similar fashion, Hamidinekoo et al. [25] uses VGG16 [9] and GoogLeNet [10] while Jiang et al [26] uses AlexNet [8]. One different approach is found in Morrell et al. [27] where the InceptionV3 [13] is totally re-trained in the new data. Despite of the result, this last approach is considered as computationally complex, since TL is performed not only because of lack of examples in the training set but also because fine tuning a model should be computationally less expensive since the first layers of a CNN are common to different and only deeper layers start to become specific in the task at hand.

Most researchers use region of interest (ROI) images focused on mass lesions. However, Carneiro et al. [28] and Morrell et al. [27] have worked on whole mammogram image models looking for an end to end design.

3) Data Augmentation and Pre-processing: As stated by Litjens et al. [29], the achievements in medical images visual tasks with deep learning do not only rely in the CNN model but also in the pre-processing of images. For instance, some of the pre-processing methods found in literature are: global contrast normalization (GCN), local contrast normalization, and Otsu's threshold segmentation. Also, since datasets are not so large, data augmentation is used by almost all researchers. Some of the most common techniques used are: rotations and cropping.

III. Proposed Approach

In this paper we aim to determine the accuracy in image classification of MobileNet [30] and NasNet [31] pre-trained CNN by using transfer learning. In Fig. 2 our methodology is shown. First, images will be pre-processed to enhance contrast. After that, two datasets will be created. One with images segmented by using Otsu's algorithm, named Otsu Dataset, and another one keeping the tissue information, named ROI dataset. As a third step, transfer learning training is conducted. Finally, classification results are displayed comparing accuracy values of classification with InceptionV3 and Resnet50.

A. Data set used for experiments

One of the most used mammogram datasets among researchers is the DDSM database [32]. In this study, we use the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [14]. CBIS-DDSM is an updated and standardized version of the Digital Database for Screening Mammography (DDSM) which includes a subset of the DDSM data selected and

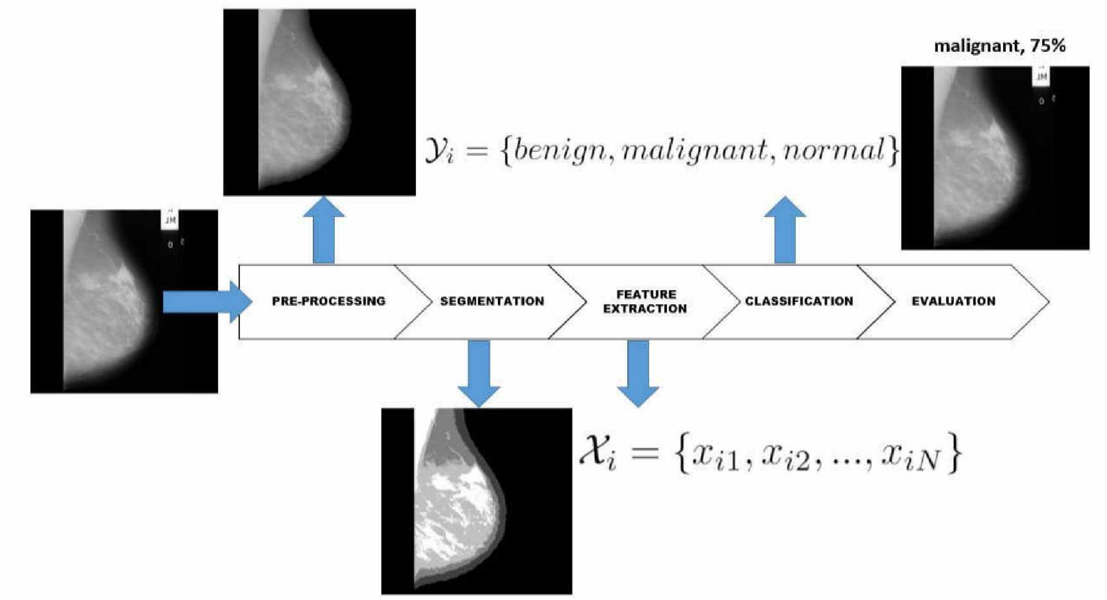

Fig. 1. Stages of a Mammogram CAD System in traditional machine learning

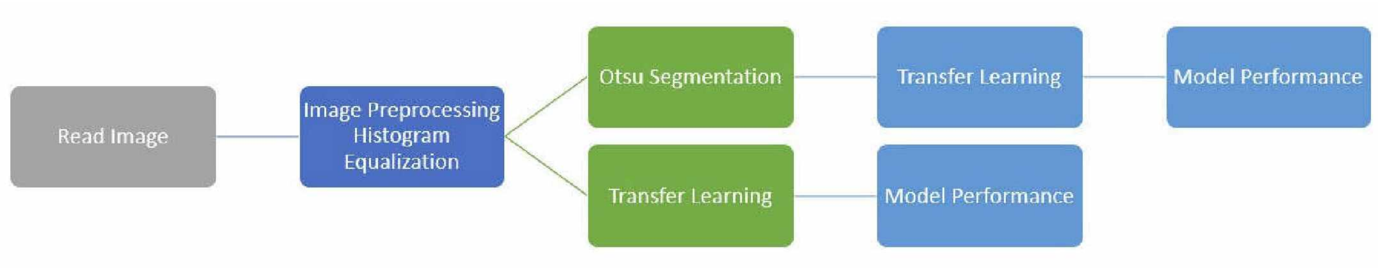

Fig. 2. Experimental Methodology

curated by a trained mammographer. In this paper we use the ROI images for breast abnormalities organized as follows:

- Training:
  - Benign: 681
  - Malignant: 637
- Testing:
  - Benign: 231
  - Malignant: 147

B. Image pre-processing

In this section we present the steps performed in the pre-processing stage of our proposed method. First of all, we have to convert DICOM files (.dcm) to image .jpg format. Then, we consider it necessary to normalize the image to the range of values of 256 levels of gray. After that, in order to increase the contrast of image we perform an histogram equalization. This method allows to enhance the global contrast of each image. In Fig.3 an example of histogram equalization of a ROI mass image is presented.

Since ROI images provided by CBIS-DDSM are of different sizes, it was necessary to apply a change of size in the image considering the aspect ratio as defined in (1), where r is the aspect ratio, w and h are the width and height of the image respectively.
Aspect ratio was considered to resize the image as a parameter that helps to preserve the best quality possible from the original image in both upsampling and downsampling procedures. Most pre-trained models use an image size of 224x224 (width, height) which is used as target size. For upsampling, cubic interpolation was used, whereas for downsampling, area interpolation gives best results. Also, images with an aspect ratio inferior to 0.4 and superior to 1.5 where removed from dataset.

$$r = w/h \tag{1}$$

C. Image Segmentation Based on Global Threshold

In this step, we applied an automatic threshold selection region based segmentation method. Thus, we use the Otsu algorithm [33] which is a global thresholding technique depending only on the gray value of the image. Next, morphological opening operation was used to remove breast tissue and extract mass from background to create a second training and testing set. Morphology is a large set of image processing operations that process images based on shapes. Morphological operations apply an structural element to an input image, creating an output image of the same size. These operations where carried out with OpenCV [34] and Skimage [35] in python. The resulting image is shown in Fig.4

D. Transfer Learning model

In our approach we want to transfer knowldege from the source domain D_S of natural images to our target domain of ROI mammogram mass images D_T. This means that a target classifier $h_T(\mathcal{X}_T)$ is to be trained such that given a ROI mammogram image $\mathcal{X}_{T_i}$ from target domain D_T, we get a prediction about the malignancy of the image in two categories $\mathcal{Y}_{T_i} = \{benign, malignant\}$. This is stated in (2) for a sample i. In this paper, transfer learning is used as a feature extractor i.e. that all layers are frozen and only top layer from the source classifier $h_S(\mathcal{X}_S)$ is retrained for the new target classes.

$$\mathcal{Y}_{T_i} = h_T(\mathcal{X}_{T_i}) \tag{2}$$

In our approach, transfer learning will be used as a feature extractor of the ROI image.

IV. Experimentation and Results

A. Evaluation metrics

In this study, we used traditional measures to evaluate the performance of our transfer learning such as classification accuracy which is indicated in (3), where P is the total of positive cases or benign, N is the total of negative cases or malignant. TP is the true positive quantity i.e. all benign cases that the classifier predicts correctly as benign. TN stands for true negatives.

$$Accuracy = \frac{TP + TN}{P + N} \tag{3}$$

B. Transfer learning model comparison

According to our literature review, MobileNet and NasNet have yet not been tested in mammogram images. In this paper, we test the accuracy in classification of abnormal mammograms by using transfer learning. This means that only the top layer of each model will be substituted from original ImageNet

Table I
TRANSFER LEARNING RESULTS

Model	Roi Dataset Accuracy	Otsu Dataset Accuracy
MobileNet	74.3%	66.0%
Resnet50	78.4%	62.4%
InceptionV3	68.9%	67.5%
NasNet	73.1%	68.0%

architecture, by replacing the 1000 categories with benign and malignant. In our experiments we trained for 6000 epochs, with a learning rate of 0.01. A batch size of 100 was used. In all cases, original training and testing datasets were merged to form a train set of 903 benign images and 772 malignant images. Validation dataset is of 10%. Testing set is of 10% and Training dataset is 80%. No data augmentation was used in these tests.

Table I summarizes the classification accuracy of each model tested in this paper. As indicated in Section III, we have created two datasets to compare the accuracy performance of using Otsu segmentation algorithm to the ROI images. Results show that transfer learning in each of this models are better if we do not use Otsu's segmentation. The best model so far for our study is Resnet50.

V. Discussion

In this study, we have compared InceptionV3, Resnet50, MobileNet, and NasNet pre-trained models in the task of mammogram abnormalities classification. From our results we conclude that transfer learning from natural images domain to mammogram images domain is possible. However, classifier performance also depends on the image pre-processing applied to the images. In fact, even though image augmentation was not used in the present work, this will be reviewed in a future study, since there is a consensus in the use of data augmentation in general for transfer learning problems, but it is not so certain to say that geometric transformations improve mammogram classifiers. Resnet50 is the model that achieved the best accuracy for CBIS-DDSM dataset in our study. Another interesting finding is that Otsu segmentation did not improve classification results.

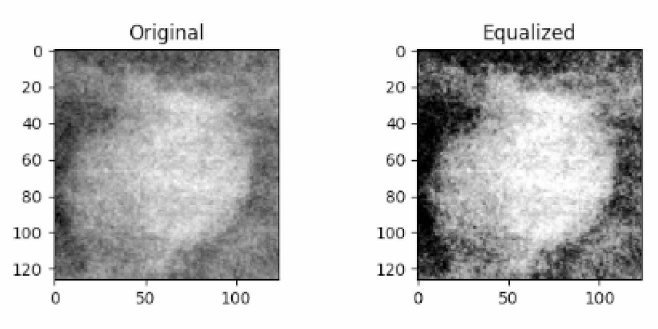

Fig. 3. Histogram equalization of ROI image.

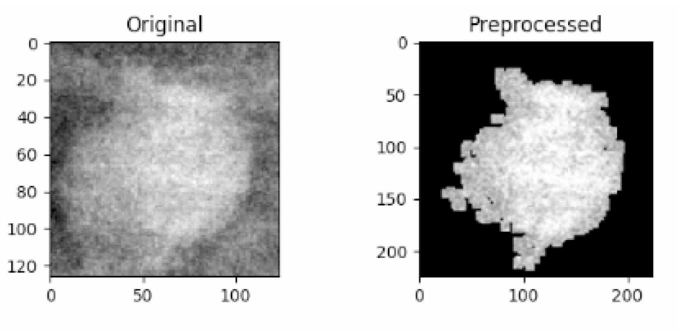

Fig. 4. Removal of tissue background.

VI. FUTUREWORKS

Since, this is our first work in transfer learning, there are some future works that we aim to continue investigating. First, the systematic literature review about transfer and ensemble learning. Second, in order to improve current results we will be using ROC curves and Confusion Matrix to measure model classification. Also, pre-processing will be reviewed, since literature has not used histogram equalization. Another, future work is the use of these pre-trained models as feature extractors to train an ensemble learning model. It is also necessary to research fine-tuning in a next study in order to find if classification accuracy can be improved by training some of the last layers and not only the top layer as in present study.

ACKNOWLEDGMENT

The authors would like to thank EPN by project REDU: PREDU 2016-013 in the development of this work. We also want to thank technical team of CEDIA for all the help provided during our experimentation stage.

REFERENCES

[1] World health organization, global health observatory (2018). Geneva: health organization. [Online]. Available: who.int/gho/database/en/

[2] S. Charan, M. J. Khan, and K. Khurshid, "Breast cancer detection in mammograms using convolutional neural network," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1–5, March 2018.

[3] I. Thomassin-Naggara, A. Tardivon, and J. Chopier, "Standardized diagnosis and reporting of breast cancer," Diagnostic and Interventional Imaging, vol. 95, no. 7, pp. 759 – 766, 2014.

[4] X. Zhang, Y. Zhang, E. Y. Han, N. Jacobs, Q. Han, X. Wang, and J. Liu, "Whole mammogram image classification with convolutional neural networks," pp. 700–704, 2017.

[5] A. Jalalian, S. Mashohor, R. Mahmud, B. Karasfi, A. R. B. Ramli, C. S. Engineering, U. Putra, H. Science, U. Putra, and Q. Branch, "Review article : Foundation and methodologies in computer-aided," pp. 113–137, 2017.

[6] D. Selvathi and A. Aarthypoornila, "Performance analysis of various classifiers on deep learning network for breast cancer detection," pp. 359–363, January 2018.

[7] N. I. Yassin, S. Omran, E. M. El Houby, and H. Allam, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review," Computer Methods and Programs in Biomedicine, vol. 156, pp. 25 – 45, 2018.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Tech. Rep.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Tech. Rep., 2015.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Tech. Rep. , 2015.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Inception v3," pp. 4278–4284, 2015.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," Tech. Rep. 2015.

[14] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," Scientific Data, vol. 4, pp. 170-177, Dec 2017.

[15] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," Neurocomputing, vol. 187, pp. 27 – 48, 2016.

[16] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, pp. 60–88, 2017.

[17] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big Data, vol. 3, no. 1, p. 9, 2016.

[18] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[19] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" Advances in neural information processing systems, pp. 3320–3328, 2014.

[20] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Engineering, vol. 2, pp. 1051, 2007.

[21] M. Perez, M. E. Benalcazar, E. Tusa, W. Rivas, and A. Conci, "Mammogram classification using back-propagation neural networks and texture feature descriptors," 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), pp, 1–6, October 2017.

[22] Y. Bengio, "Convolutional networks for images, speech, and time-series unsupervised learning of speech representations view project parsing view project," Tech. Rep., 1997.

[23] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," J. Med. Imag, vol. 3, no. 3, pp. 34501, 2016.

[24] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," Computer Methods and Programs in Biomedicine, vol. 157, pp. 19–30, April 2018.

[25] A. Hamidinekoo, Z. Suhail, E. Denton, and R. Zwiggelaar, "Comparing the performance of various deep networks for binary classification of breast tumours," 14th International Workshop on Breast Imaging (IWBI 2018), pp. 39, 2018.

[26] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," ACM International Conference Proceeding Series, pp. 59–62, 2017.

[27] S. Morrell, Z. Wojna, C. S. Khoo, S. Ourselin, and J. E. Iglesias, "Large-scale mammography CAD with deformable conv-nets," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11040 LNCS, pp. 64 – 72, 2018.

[28] G. Carneiro, J. Nascimento, and A. P. Bradley, " Deep learning models for classifying mammogram exams containing unregistered multi-view images and segmentation maps of lesions," 1st ed. Elsevier Inc., 2017.

[29] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," pp. 60–88, December 2017.

[30] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017.

[31] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8697–8710, 2018.

[32] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography," Tech.Rep, 2000.

[33] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, January 1979.

[34] O. D. Team, "Feature detection," OpenCV2, vol. 4, pp. 2011–2013, 2013.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," Journal of machine learning research, vol. 12, pp. 2825–2830, October 2011.