Active Learning Strategies and Convolutional Neural Networks for Mammogram Classification

João Marcelo Tozato[(✉)] ⬤, Pedro Henrique Bugatti ⬤,
and Priscila Tiemi Maeda Saito ⬤

Department of Computing, Federal University of Technology - Parana,
1640 Alberto Carazzai Ave, Cornelio Procopio, Parana, Brazil
`tozato@alunos.utfpr.edu.br`, {`pbugatti,psaito`}`@utfpr.edu.br`

Abstract. Deep learning has been used successfully in a variety of applications due to the large data availability and the growth in computing power. However, some domains present a shortage of both samples and labels, for instance, the medical area. In this work, we propose machine learning approaches that include traditional supervised classifiers and active learning methods for the breast lesion domain, in order to aid breast cancer diagnosis. We propose the introduction of active learning strategies in this process, to sort out the most informative samples in the dataset. The active learning process reduces the burden of the dataset annotation, while also improving the robustness of our models. Hence, we achieved considerable gains with fewer labeled training images, minimizing the specialist's annotation effort. The validation of our proposed methodology is done on a public breast lesion-related dataset and our results show considerable accuracy gains over the traditional supervised learning approach and reductions of up to 68% in the labeled training sets.

Keywords: Machine learning · Active learning · Pattern recognition.

1 Introduction

The usage of deep learning for classification or segmentation tasks has become extremely efficient in a variety of application domains, such as medical problems. For instance, in hemorrhage detection in CT scans [6], segmentation of MRI scan images of the brain [1], breast cancer detection [9], amongst other applications.

According to the American Institute for Cancer Research [4], breast cancer is the most common type and the fifth major cause of death in women worldwide. Hence, applying intelligent systems in this scenario can assist in the decision process of the professional in charge of the diagnosis. Furthermore, the early diagnosis of this disease can reduce its mortality rate.

Despite some efforts found in the literature [10,14,18], most of these consider learning approaches that require a substantial amount of labeled data and do not take into account some mandatory restrictions of medical applications (e.g. mostly related to computational time and resources).

In addition, there are other challenges related to data acquisition and labeling processes. The gathering of mass-related breast lesion images is not trivial due to the maintenance of patients' privacy, availability of data by hospitals, among others. Moreover, generally, there is the requirement of sample labeling by one or more specialists (e.g. considering radiologists with different levels of experience) to ensure that the correct labels are assigned. It also impacts and contributes to the lower availability of labeled samples. The data labeling process requires time and effort from the specialist and is highly susceptible to errors.

Therefore, this paper addresses the study, development and validation of active learning strategies, in order to compare them to the traditional supervised learning approach. Active learning strategies have been widely used and successful in several other application domains. Such strategies allow for obtaining a reduced set of the most informative samples to the learning process of pattern classifiers. More effective and efficient classifiers can be obtained, achieving higher accuracies faster and minimizing the effort of the expert in the labeling process.

2   Background

The active learning approach considers the usage of the classifier in the selection of the most informative samples from a designated dataset. This method is advantageous on tasks that require hundreds or even thousands of labeled data (such as images), mainly by reducing the burden of annotating the whole dataset, which demands plenty of time and effort from a specialist in a given domain to execute the labeling of these samples [3].

It is an iterative process that makes use of a selection strategy to gradually obtain a fixed number of samples and incorporate them in the training dataset of the learning algorithm. In such way, by using the active learning approach, it is possible to create robust classification models with far less labeled instances and hence reducing the cost of the data labeling process.

At each iteration of the active learning process, a fixed number of samples (for our methodology twice the number of existing classes) is selected by the selection strategy and then gradually incorporated into the training set. New instances of the classifier are obtained and evaluated in the test set. Each active learning method used in this work to select the most informative samples is related to the uncertainty criterion [16].

There are different active learning strategies in the literature, one of them is based on Entropy (EN) [17], which can be understood as the degree of uncertainty of a variable, prioritizing samples that have a greater value for this measure. It calculates this according to the Eq. 1, where y is the probability of a given label for a sample x.

$$EN(x) = -\sum p_i(y|x) \log p_i(y|x) \tag{1}$$

In the technique called Least Confidence (LC) [11], the model selects the sample that presents a lower confidence for the most probable class. Equation 2 shows

the inner working of this technique, where y' is the highest probability given by the model for a sample x. Thus, a lower value for the probability of the most probable class leads to a higher chance of this sample being selected, due to the low confidence assigned to it.

$$LC(x) = 1 - p(y'|x) \tag{2}$$

Equation 3 shows the Margin Sampling (MS) [15] technique, which takes into account not only the most likely label, as in the previous strategy, but it is based on the smallest difference between the first and second most likely labels for sample selection, where y' and y'' represent the highest probabilities for a sample x.

$$MS(x) = p(y''|x) - p(y'|x) \tag{3}$$
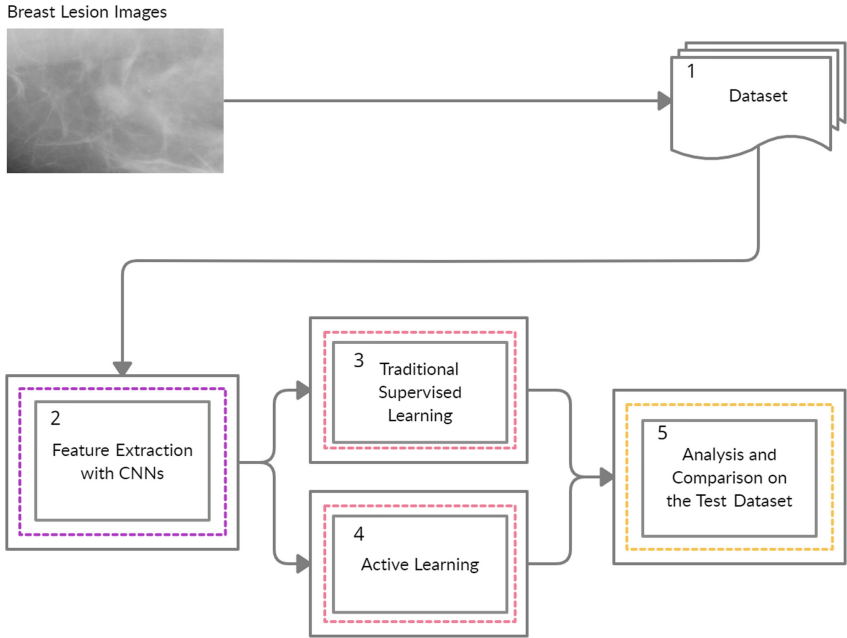
3   Proposed Methodology

Initially, according to the first step of our pipeline (Fig. 1) we obtained and organized our dataset as described in Subsect. 4.1. Our methodology consists in two key approaches (traditional supervised learning and active learning, respectively) as shown in Steps 3 and 4 of the pipeline. Both approaches depend on the extraction of deep features through the use of CNNs (according to Step 2), which are acquired by removing the classification layers of a CNN model and getting the output of a given layer. In the present work we consider the last layer before the fully connected layers for this process.

For the feature extraction process we apply the Transfer Learning strategy, which allowed the initialization of our network's weights based on the weights of another neural network that was already trained on the ImageNet dataset [5]. The parameters of the old network are reused for the inference process of this new network, therefore reducing the computational cost to train neural networks from scratch.

We have also applied normalization to our input data. It is a technique that aims to adjust the mean and standard deviation of the input data values of a given neural network on a common scale, such as close to zero and one, respectively. It becomes especially important when using pre-trained neural networks, due to the fact that the model only knows how to work with data of the type that it has seen before. If the inputs of the new network using these parameters, do not share these normalization statistics, the results will not be as expected.

The fourth step of our methodology consisted in using active learning strategies alongside with traditional classifiers. Active learning strategies allow the selection of the most informative samples for the learning process. Therefore, we can reduce the amount of annotated images required for classification tasks while achieving significant or equivalent results when compared to the supervised approach, which requires a completely annotated training set.

We performed comparisons between the active learning strategies and the random sample selection at each iteration. In addition, we also compared the

Breast Lesion Images

Fig. 1. Pipeline of the proposed methodology.

traditional supervised learning approach (which require a fully annotated training set) and the active learning strategy.

4   Experiments

4.1   Dataset

We used in this work the public dataset called MAMMOSET [12], which contains images regarding to three types of lesions: mass, calcification and normal (with no kind of lesion). For this work we have considered exclusively the subset of mass-related lesions, which are divided into malignant or benign. Figure 2 shows samples of the two distinct classes of the dataset.

The subset contains 1381 images in total and are distributed in the following manner: the training and test set are composed of 568 and 67 images for the malignant class and 671 and 75 images for the benign class, respectively.

4.2   Scenarios

Our training set is divided into 10 mutually exclusive stratified splits so that the percentage of each class samples is preserved. Each split has its own validation set, which is used to check the performance and model's biases during its training

(a)   (b)
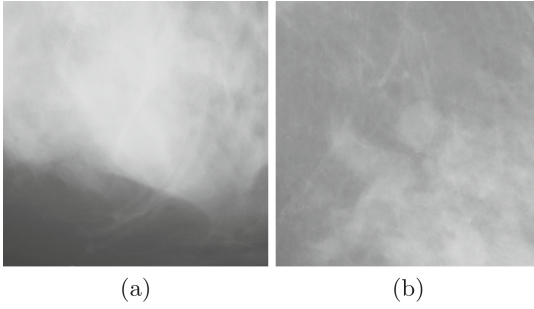
Fig. 2. Examples of images from the MAMMOSET dataset: (a) malignant sample, (b) benign sample.

in that split. At the end of the training of a given split, we start the inference process in a fixed test set that is the same throughout the process for each split. The traditional supervised learning and active learning experiments were conducted using the deep features extracted from the CNN architectures (DenseNet-121, DenseNet161, EfficientNetB3, EfficientNetB4 ResNet34 and ResNet50) in conjunction with traditional classifiers such as k-Nearest Neighbors (k-NN) [13], Naive Bayes (NB) [7], Random Forest (RF) [2] and Support Vector Machines (SVM) [8]. The images of the dataset were resized to 224 x 224 pixels before being used as input to the CNNs. For every network considered in the feature extraction process, we have not updated its weights, these architectures were just used as fixed feature extractors. Table 1 shows the dimensionality of each feature map of the CNNs used in the feature extraction process. The hyper-parameters chosen for the classifiers are the standard as provided in their literature.

Table 1. Description of the deep feature extractors with respect to their feature map dimensionality.

| Extractor	Feature dimensionality
DenseNet121	1024
DenseNet161	1664
EfficientNetB3	1536
EfficientNetB4	1792
ResNet34	512
ResNet50	2048

5   Results and Discussion

Initially, we show the results obtained by the traditional learning approach (Fig. 3), considering the deep features extracted from each of the CNN architectures and the supervised classifiers (k-NN, NB, RF and SVM, respectively). We
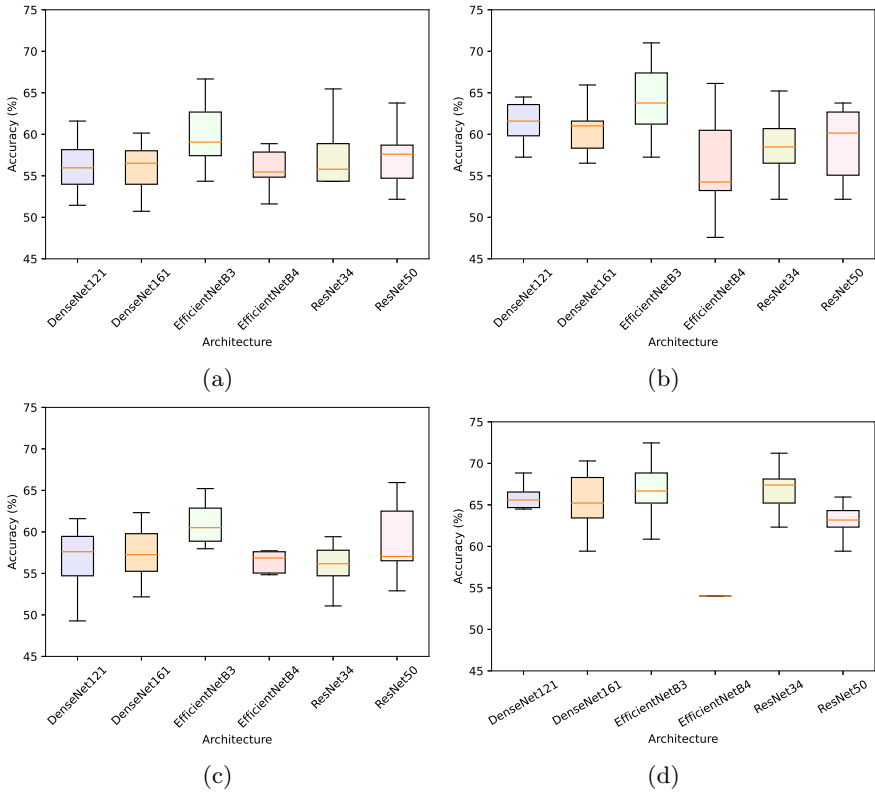
Fig. 3. Mean accuracies obtained by the traditional learning approach considering the deep features extracted from each architecture (DenseNet121, DenseNet161, Efficient-NetB3, EfficientNetB4, ResNet34 and ResNet50) and the supervised classifiers: (a) k-NN, (b) NB, (c) RF and (d) SVM.

can notice that, in general, the deep features obtained from the EfficientNetB3 architecture presented higher accuracy values in relation to the other architectures for all classifiers. The EfficientNetB3 architecture achieved the highest accuracy (up to 66.98 ± 3.43) with the SVM classifier. It is important to note that these experiments requires all samples from the dataset labeled.

Then, we performed experiments considering the active learning approach with the features extracted from the EfficientNetB3 architecture, in order to reduce the need to annotate all samples in the dataset. The reason for choosing this particular architecture is mainly due to its consistency and overall high accuracies presented in the supervised learning experiment. Figure 4 presents the mean accuracies obtained by the selection strategies (EN, LC, MS and RANDOM) along the iterations of the learning process, considering each of the supervised classifiers (k-NN, NB, RF and SVM, respectively).
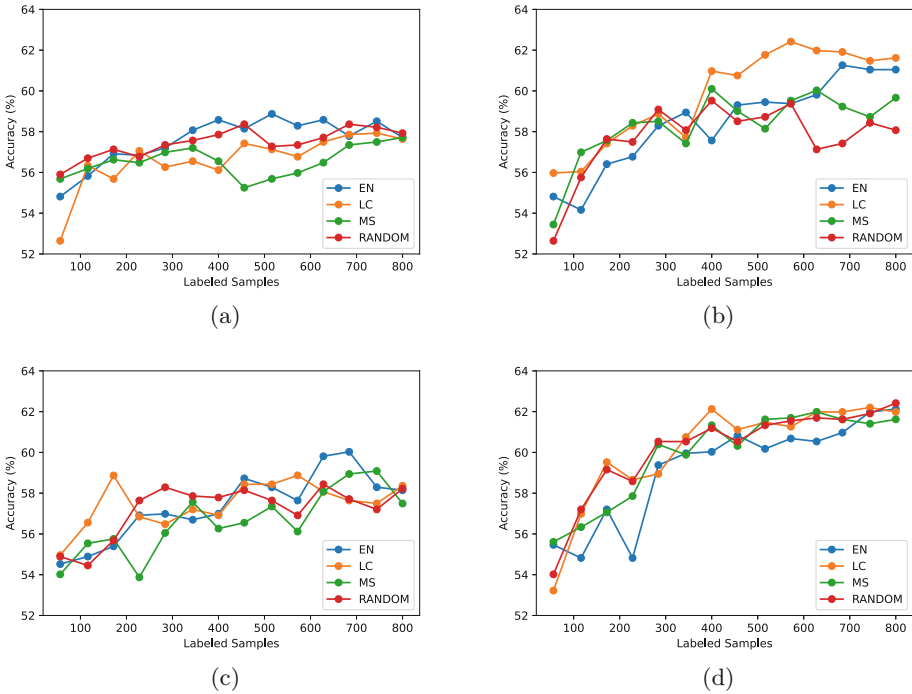
Fig. 4. Mean accuracies obtained by the active learning approach considering the selection strategies (EN, LC, MS and RANDOM), features extracted from the Efficient-NetB3 architecture and the supervised classifiers: (a) k-NN, (b) NB, (c) RF and (d) SVM.

It is possible to note that the active learning approach achieves better results with a reduced labeled training dataset, when comparing to the traditional supervised learning approach, which requires the dataset to be completely labeled. The active learning strategies allow a significant reduction (up to 44%, 68%, 61%, 60%) in the labeled training set required to reach accuracies equivalent to those obtained by the traditional supervised approach, considering the k-NN, NB, RF and SVM classifiers, respectively.

Table 2 shows the computational times for each combination of active learning strategy and classifier. We can verify that, as a more complex classifier, the SVM model presented classification times much higher than the others. Moreover, the random strategy (RANDOM) achieved smaller selection times, since it does not have a specific criterion for the selection of the samples that will integrate the training dataset.

Table 2. Total selection and classification times in seconds obtained by the active learning approach considering the selection strategies (EN, LC, MS and RANDOM), features extracted from the EfficientNetB3 architecture and the supervised classifiers (k-NN, NB, RF and SVM).

Strategy	Classifier	Selection time	Classification time
EN	k-NN	1.002e+00	1.477e-01
LC		8.151e-01	1.101e-01
MS		8.132e-01	1.103e-01
RANDOM		1.619e-02	4.839e-02
EN	NB	9.541e-02	3.975e-02
LC		5.920e-02	2.184e-02
MS		6.288e-02	2.509e-02
RANDOM		2.500e-02	1.145e-02
EN	RF	2.273e-02	1.181e-01
LC		1.303e-02	8.439e-02
MS		1.565e-02	9.341e-02
RANDOM		2.785e-02	9.621e-02
EN	SVM	6.496e-01	3.309e+00
LC		5.404e-01	2.652e+00
MS		7.282e-01	3.657e+00
RANDOM		1.288e-02	2.666e+00

6   Conclusion

In the present work, we conduct extensive experiments following the proposed methodology in order to assist in the diagnosis of breast cancer lesions. We compared the results obtained from two main different approaches to this image classification task: supervised learning and active learning strategies alongside with traditional classifiers. Regarding the active learning approach, we were able to verify that, in contrast to the supervised approach, which requires the whole labeled dataset for its learning process, the selection strategies provide a way to create representative training sets and achieve high accuracies for this particular classification task.

We also explored the significance of using active learning strategies with CNNs on image classification tasks, exhibiting that it is possible to reach expressive results with a reduced labeled dataset, specially when comparing to traditional supervised learning. The results obtained in the carried out experiments show the benefits of the usage of active learning strategies in the development process of a classifier, since in this medical context is not common that all the available images are annotated. Then, by selecting the most informative samples for the model's learning there is a reduction in both time and effort required for the labeling process of a dataset by a specialist.

References

1. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L., Erickson, B.J.: Deep learning for brain MRI segmentation: state of the art and future directions. J. Digit. Imag. **30**(4), 449–459 (2017)
2. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
3. Bressan, R.S., Bugatti, P.H., Saito, P.T.: Breast cancer diagnosis through active learning in content-based image retrieval. Neurocomput. **357**, 1–10 (2019)
4. American Institute of Cancer Research: Breast cancer: how diet, nutrition and physical activity affect breast cancer risk. https://www.wcrf.org/dietandcancer/breast-cancer
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255. IEEE (2009)
6. Grewal, M., Srivastava, M.M., Kumar, P., Varadarajan, S.: Radnet: radiologist level accuracy using deep learning for hemorrhage detection in CT scans. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 281–284. IEEE (2018)
7. Hand, D.J., Yu, K.: Idiot's bayes—not so stupid after all? Int. Stat. Rev. **69**(3), 385–398 (2001)
8. He, Z., Xia, K., Niu, W., Aslam, N., Hou, J.: Semisupervised SVM based on cuckoo search algorithm and its application. Math. Prob. Eng. **2018**, 1–13 (2018). https://doi.org/10.1155/2018/8243764
9. Huynh, B.Q., Li, H., Giger, M.L.: Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J. Med. Imag. **3**(3), 034501 (2016)
10. Kooi, T., et al.: Large scale deep learning for computer aided detection of mammographic lesions. Med. Imag. Anal. **35**, 303–312 (2017)
11. Lewis, D.D., Catlett, J.: Heterogeneous uncertainty sampling for supervised learning. In: Machine Learning Proceedings 1994, pp. 148–156. Elsevier, Amsterdam (1994)
12. Oliveira, P., de Carvalho Scabora, L., Cazzolato, M., Bedo, M., Traina, A., Jr., C.: Mammoset: An enhanced dataset of mammograms. In: Proceedings of the satellite events - Brazilian Symposium on Databases, pp. 256–266 (2017)
13. Rani, P., Vashishtha, J.: An appraise of KNN to the perfection. Int. J. Comput. Appl. **170**(2), 13–17 (2017). https://doi.org/10.5120/ijca2017914696
14. Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I.: Detecting and classifying lesions in mammograms with deep learning. Sci. Rep. **8**(1), 1–7 (2018)
15. Scheffer, T., Decomain, C., Wrobel, S.: Active hidden Markov models for information extraction. In: Hoffmann, F., Hand, D.J., Adams, N., Fisher, D., Guimaraes, G. (eds.) IDA 2001. LNCS, vol. 2189, pp. 309–318. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44816-0_31
16. Settles, B.: Active learning literature survey. University of Wisconsin-Madison Department of Computer Sciences, Technical reports (2009)
17. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
18. Valério, L.M., Alves, D.H., Cruz, L.F., Bugatti, P.H., de Oliveira, C., Saito, P.T.: Deepmammo: deep transfer learning for lesion classification of mammographic images. In: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), pp. 447–452. IEEE (2019)