Retrained Deep Neural Network Model for Breast Cancer Imaging Classification

Bruno Smarsaro Bazelato
Universidade Federal do Espírito Santo – Campus São Mateus, São Mateus, ES, Brazil
brunosmarsaro@gmail.com

Helder Roberto de Oliveira Rocha
Universidade Federal do Espírito Santo – Campus São Mateus, São Mateus, ES, Brazil
helder.rocha@ufes.br

Wilian Hiroshi Hisatugu
Universidade Federal do Espírito Santo – Campus São Mateus, São Mateus, ES, Brazil
wilian.hisatugu@ufes.br

Leonardo José Silvestre
Universidade Federal do Espírito Santo – Campus São Mateus, São Mateus, ES, Brazil
leonardo.silvestre@ufes.br

Luís Otávio Rigo Júnior
Universidade Federal do Espírito Santo – Campus São Mateus, São Mateus, ES, Brazil
luis.rigo@ufes.br

RESUMO

Aprendizagem profunda tem sido amplamente aplicado a imagens médicas. Considerando que o câncer de mama é uma das principais causas de morte entre mulheres ao redor do mundo, o uso de técnicas de aprendizagem profunda para realizar um diagnóstico mais preciso ajudaria a diminuir esses números. Assim, é proposto neste artigo o processo de retreinamento do modelo Inception V3 para classificar câncer de mama a partir de imagens médicas entre benigno e maligno. Comparado a trabalhos recentes, o desempenho foi satisfatório, considerando que foi utilizado mamografias completas e sem processamento.

PALAVRAS CHAVE. Câncer de Mama, Aprendizagem Profunda, Imagens Médicas.

Tópicos (SA - PO na Área de Saúde)

ABSTRACT

Deep learning has been largely applied for medical images. Since breast cancer is one of the main causes of death among women around the globe, the use of deep learning techniques to make the diagnosis more accurate could help to decrease these numbers. Thus, it proposed on this paper the retraining of the Inception V3 model to classify breast cancer from mammographies between benign and malignant. Comparing to recent works, it performed satisfactorily, considering it was used raw and full mammogram images.

KEYWORDS. Breast Cancer, Deep Learning, Medical Images.

Paper topics (SA - OR in Health)

https://proceedings.science/p/85436?lang=pt-br

1. Introduction

Breast cancer is the type of cancer that has the highest incidence in women and is the second in death rate [Jemal et al., 2017]. Many factors can be the cause of this increasing numbers, *e.g.* biological characteristics, socio economic factors and cultural factors [Ho et al., 2010]. However, early stage diagnosis contributes to lower mortality rates, since the stage at diagnosis is a powerful indicator of cancer survival. Thus, a way of satisfactorily performing low cost and fast diagnosis is through computational methods, more specifically Machine Learning (ML) techniques [Bolón-Canedo et al., 2016].

A particular feature of ML that makes it one of the most effective techniques to be applied to problems with high degree of complexity is its capability of learning from data [Agrawal et al., 2017]. For instance, a set of labeled images, such as mammographies, can be used to train an algorithm, that after many iterations, begins to associate certain findings and patterns, for example masses, microcalcifications and location of lesions [Balleyguier et al., 2007], to make a prediction. Hence, ML is a useful tool to help radiologists and clinicians for the categorization of mammographic images.

As Bolón-Canedo et al. [2016] exemplifies, several methods, including artificial neural networks, Bayesian networks, support vector machines and decision trees, have been broadly applied to medical problems – *e.g.* classification of Alzheimer's disease and glaucoma screening. However, conventional ML techniques were restricted to data in a simpler representation that allowed them to identify or classify patterns in the input [LeCun et al., 2015]. Nevertheless, due to Deep Learning (DL) methods, this limitation has been overcome.

Essentially, DL is a technique of artificial neural networks (ANN), which consists of a set of connected neurons organized in layers and each of them produces a sequence of values that is propagated all the way through the network [Schmidhuber, 2015]. Neurons in the first layer, called input neurons, for example, get activated directly by what is being fed to the network. Neurons in the next layers get activated by weighted connections from previously active neurons. Therefore, the learning process is basically the search for a set of weights that makes the ANN behave as desired. When plenty of these layers are used, the level of abstraction increases in each layer and the network is able to deal with higher complex problems, and the result is a deep learning model [LeCun et al., 2015].

For the reason that computational resources, such as CPU speed, data storage and sensors, have advanced so much, many ML methods have expressively progressed [Agrawal et al., 2017] and with DL has not been any different. From the 80's to the 90's, the research on ANN grew exponentially, but it got stuck until 2006 when new techniques and technologies emerged and DL came as state-of-the-art in ML [Deng et al., 2014].

Consequently, DL has been applied to solve a large variety of problems from distinct domains, such as science, business and government [LeCun et al., 2015]. Moreover, it has demonstrated exceptional performance in image recognition [Simonyan e Zisserman, 2014; Szegedy et al., 2015a; He et al., 2016], speech recognition [Graves et al., 2013; Amodei et al., 2016] and natural language processing [Sarikaya et al., 2014; Dahl, 2015; Hirschberg e Manning, 2015; Hassan e Mahmood, 2017]. Also, it has been an excellent support tool to big data analysis [Najafabadi et al., 2015].

Furthermore several researches on deep learning applied to medicine have been done. For instance, an approach for classification of electrocardiogram signals based on DL showed itself very robust and outperformed previously approaches [Al Rahhal et al., 2016]. Likewise, DL has been used for phenotyping from patient electronic health records [Cheng et al., 2016], machine health monitoring [Zhao et al., 2016], protein fold recognition [Ibrahim e Abadeh, 2017] and diagnosis system of lung cancer [Shimizu et al., 2016].

As mentioned earlier, DL has exhibited great performance in image recognition, and so for medical imaging [Shen et al., 2017]. For example, it has been applied for brain tumor segmen-

https://proceedings.science/p/85436?lang=pt-br

tation [Havaei et al., 2017], classification of video images of echocardiography [Gao et al., 2017], prostate cancer detection on multiparametric magnetic resonance images [Tsehay et al., 2017], skin lesion classification [Lopez et al., 2017; Esteva et al., 2017] and tuberculosis detection on chest radiographs [Lakhani e Sundaram, 2017]. For breast cancer detection it has also been largely applied and showed itself very successful [Arevalo et al., 2015; Carneiro et al., 2015; Huynh et al., 2016; Kallenberg et al., 2016; Wang et al., 2016; Geras et al., 2017; Kooi et al., 2017].

The main DL technique applied for medical images and images in general is deep convolutional neural networks [Tajbakhsh et al., 2016]. It is named convolutional because of its convolutional layers that are responsible for local feature detection. This method will be addressed in the second section of this paper and the third section describes recent works on DL for breast cancer classification.

In this work, it is used the CNN model Inception V3 [Szegedy et al., 2015b]. This model has achieved great performance in many tasks [Esteva et al., 2017; Kurakin et al., 2016; Hershey et al., 2017]. It is composed by several layers and it is available on the TensorFlow repository on Github[1]. The experiment using it is described in the fourth section. In the fifth section is exposed and discussed the results and lastly, in the sixth section, the paper is concluded.

2. Techniques

The main advantage in the use of deep networks is because of its capability of handling complex problems. Meanwhile, comparing it to a shallow network, a deep neural network needs less nodes (neurons), and so less parameters, to perform the same task Lai [2015], which can be a benefit in medical application where data is often a huge problem. Amongst the variety of DL techniques, deep convolutional neural networks (CNNs) has been the most applied for medical imaging [Shen et al., 2017]. The central concepts behind CNNs that makes it so successful are: local connections, shared weights, pooling and the use of many layers [LeCun et al., 2015].

In a CNN it is assumed that the data and the functions have a structure. For images, this structure assumes that there is a relevant information around each element from the data (pixel) to be analyzed. Formally, each data z will be an array of $M \times N \times K$ values, where there are $M \times N$ pixels (spatial position) and K channels per pixel (for RGB images, K is equal to 3, and for gray scale images, K is equal to 1). The functions work on the data with local operators and translate invariants in order to capture information from the pixels and their neighborhood to generate feature maps. Thus, only the input of the ANN will be an image, while the intermediate data will be feature maps derived from the original image.

The CNN is composed by a set of components and the functions that operate on the data change according to the used component. Some of the main components used are the convolution, rectified linear unit (ReLU), pooling and softmax.

2.1. Convolution

The convolution is a linear operator and the main operator of the CNN. In an image (or feature map) it is performed sliding a mask over each pixel from the image. The result for each pixel generates a feature map. The convolution is very well known in the signal and image processing field [Gonzalez e Woods, 2010; Andreas, 1986]:

$$g(x,y) = z(x,y) \star f(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} z(s,t) f(x-s, y-t) \tag{1}$$

where $z(x,y)$ and $f(x,y)$ are the image and the mask used in the convolution, respectively, $g(x,y)$ is the result of the convolution, $\star$ is the symbol to indicate the convolution operation and a and b are the mask dimensions of the convolution.

[1] https://github.com/tensorflow

https://proceedings.science/p/85436?lang=pt-br

Figure 1 shows an example of the use of a 3 × 3 mask in an image. Each element of the mask has a value that is multiplied by the value of the pixel of the image under the mask. The sum of these multiplications generates a value to a pixel in a new feature map.

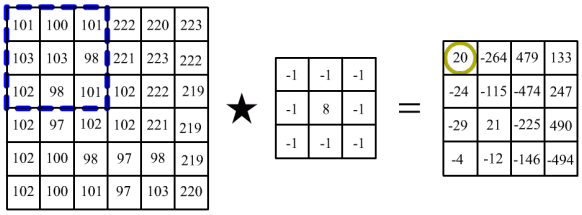


Figure 1: Example of a 2D convolution procedure. To the left of the equation is shown a 6 × 6 image and a 3 × 3 mask. The result of the convolution is the feature map to the right of the equation. The highlighted element with a circle in the feature map is the result of the convolution between the mask and the blue area of the image centered in the pixel 103. Note that the feature map is slightly smaller than the image. This reduction can be avoided by using a zero padding procedure.

In a typical CNN, each neuron in the convolution layer is a mask, whose value at each position of the mask is a weight, in addition to having an associated bias (this element is a value that is added to the convolution result). So, besides having to define the number of neurons in the convolution layer, it is also necessary to define the dimensions of the masks. For instance, if a 3 × 3 mask were defined, each neuron will have 9 weights a one bias.

Initially, the weights and biases of the neurons are generated randomly and, through a learning algorithm, these parameters are calibrated and may also be pretrained in an unsupervised manner before final calibration [Wu e Razavi, 2015].

There are a number of reasons for using convolution instead of the classical multiplication of neuron weights by the inputs. Imagine a 2D gray scale image (K = 1) of size 200 × 200 pixels, totalizing an input vector of 40,000 attributes. If the hidden layer of the ANN has 200 neurons, it would be necessary to calibrate a matrix of weights of 40,000 × 200 = 8 million weights (not counting the bias of the neurons). This only in the first layer. Moreover, vectoring an image ignores the entire spatial structure of the image [Srinivas et al., 2016]. In turn, using a convolution procedure drastically reduces the number of weights to be learned. If in the first layer it is used 96 convolution masks of dimension 11 × 11, the number of parameters to be calibrated (including bias) is (121+1) × 96 = 11,712. In addition, convolution naturally considers the spatial structure of images.

Convolution can be thought of as a regular neural network with two constraints [Srinivas et al., 2016]:

- Local connections: the convolution mask has smaller dimensions than the image to be processed. So, it is only considered a small neighborhood around the pixel to be analyzed. In contrast, in the global connectivity, typical of the vectorized representation, the image as a whole is considered.

- Weight sharing: the same convolution mask is applied on each pixel of the image and its neighborhood, different from the conventional approach in which different weights may arise for pixels in different positions. Thus, the feature detector is independent of location.

It is worth emphasizing that this paradigm of convolution is not entirely treatable for certain tasks. As an example, face images require different feature detectors at different spatial locations [Taigman et al., 2014].

2.2. Rectified Linear Unit (ReLU)

ReLU is the simplest nonlinear operator applied after a linear filtering, being applied to each pixel of the feature map in order to rectify negative values:

https://proceedings.science/p/85436?lang=pt-br

$$ReLU(x, y) = max[0, g(x, y)] \tag{2}$$

where $g(x, y)$ is the feature map and (x, y) is the coordinate of the processed pixel.

This function has a good sparsity property because it generates several values equal to zero, and since it does not have an upper limit, it does not suffer as much from the weakening effect of the flow.

Other variants of ReLU have been proposed. [Maas et al., 2013] proposed the leaky-ReLU, where α is a pre-determined parameter (Equation 3). In turn, He et al. [2015] proposed that α is a parameter to be learned by the network.

$$L - ReLU(x, y) = max[0, g(x, y)] + \alpha min[0, g(x, y)] \tag{3}$$

2.3. Pooling

Another important operator is pooling. The pooling operator is, in general, a way of reducing dimensionality, discarding unnecessary information and preserving the most critical ones [Boureau et al., 2010]. The pooling operation reduces the output of multiple neurons from the convolution layer and helps to make the detected features invariant from the location in the input image. Thus, the neural network becomes less sensitive to the exact location of the pixels [van Doorn, 2014; Boureau et al., 2010].

A pooling operator operates on channels of individual features, combining the near feature values to a value by applying a suitable operator. Common choices include max-pooling, which returns the maximum input value, and average pooling, which returns the average value. For example, max-pooling is defined as:

$$Pooling(x, y) = \max_{s=[-n,n]} g(x + s, y + s) \tag{4}$$

where $g(x, y)$ is the feature map and $n \times n$ is the dimension of the pooling mask. The max-pooling operation substitutes each region $n \times n$ with the maximum obtained value, then the activation size for the next layer is reduced by a factor of n^2. A common procedure for decreasing computational effort is to reduce the size of feature maps by a factor of n in each dimension after passing through a pooling operator [Srinivas et al., 2016].

2.4. Softmax

The output layer requires a special property not present in the other layers. A softmax layer is commonly used in the final layer of the classifier. The objective is to minimize the error of the softmax function through the appropriate adjustment of the weights of the neurons of the neural network.

Barbu et al. [2016] compared softmax function with other approaches and achieved better results with the Lorenz loss function in their study.

3. Recent Works

It has not been long since deep learning began to be applied for medicine and clinical imaging, and so for breast cancer diagnosis. In 2016 the International Symposium on Biomedical Imaging (ISBI) started holding a competition to evaluate automatic detection systems for metastatic breast cancer, in its first edition the winner team got 0.925 of area under the ROC curve (AUC) and it got to 0.9948 when human pathologists made the prediction alongside the system [Wang et al., 2016]. They proposed a deep convolutional neural network to make patch-level decisions to discriminate tumor-patches and normal-patches. To improve the search for cancer metastasis in the slides and to decrease computation time, they implemented a procedure to pre-process the images. First, it is detected tissue within the whole slide images and then it is deleted the background white space. To train and test the model they used a dataset provided by the competition called Camelyon16, which contains 400 images.

https://proceedings.science/p/85436?lang=pt-br

Geras et al. [2017] suggested a system that even thought is also based on deep convolutional neural networks, uses a multi-view approach. This means that it is used more than one image of the same breast to come to a conclusion. The authors explain that alike an exam in medical imaging which comes with a set of views and all of them are taken into account to make the diagnosis, using these multiple views in the neural network more precise results could be achieved. They used four views, two from the left breast and other two from the right breast. Each image was inputted to separated convolution and pooling layers and afterwards, they were all given as input to a softmax layer to produce the output. Nearly 103 thousand images on high-resolution, being 80% for training, 10% for testing and the last 10% for validation, were used in order to evaluate how much the size of the dataset and the resolution of the images affect the results.

Using a similar but simpler approach, Arevalo et al. [2015] proposed a convolutional neural network to classify mammography mass lesions. The dataset used is called BCDR-F03 and is composed by 736 images – 60% training and 40% testing. The pre-processing of the images was done in three stages, firstly the images were cropped to make them all the same dimensions, secondly for each image in the training set they generated 7 other images using combinations of flipping and rotations, and then the images were normalized because of the digitalization process that affects all pixel values of the image. For the classification layer, the authors implemented a linear SVM and the AUC was 0.86.

Kooi et al. [2017] also applied deep convolutional neural networks for classification of cyst and mass patches through mammographies. However, their approach uses two different datasets, one is used to pretrain the neural network and the other to retrain. According to the authors, the pretraining is responsible for extracting features, so they used a large dataset of soft tissue lesion candidates, and the retraining takes care of the diagnosis task using data from a set of screening mammograms. To evaluate their model, they performed several experiments, for instance combining two different views of a breast to make the classification and changing the image resolution, similarly to what Geras et al. [2017] did. Their system achieved an AUC of 0.80, nonetheless it got up to 0.866 when the lesions were larger than 20 mm.

Huynh et al. [2016] combined transfer learning with deep convolutional neural networks to build a breast imaging computer-aided diagnosis. Their database contained 219 mammographic images from which in each image they extracted a region of interest producing 607 images – 261 benign and 346 malignant images. They experimented three different models, a pretrained CNN, an analytical classifier and an ensemble classifier which combines the output from the other two classifiers. The highest AUC was achieved from the ensemble model, being 0.86.

Alike Geras et al. [2017], Carneiro et al. [2015] came up with a pretrained CNN model for multiview classification of mammographies. They used the InBreast and DDSM mammogram datasets, totalizing 1090 images. The chosen model was CNN-F, which was pretrained with Imagenet dataset. Also, as a base for classification, they used BI-RADS Balleyguier et al. [2007] classes to define the output categories, negative (BI-RADS 1), benign (BI-RADS 2 and 3) and malignant (BI-RADS 4,5 and 6). Their results showed that for a 2-class classification (benign and malignant) using InBreast and DDSM they achieved an AUC of 0.91 and 0.97, respectively.

Unlike the previous authors, Kallenberg et al. [2016] proposed a different architecture of deep convolutional neural networks. Instead of using an entirely supervised learning model, part of the network used unsupervised feature learning. They call this model convolutional sparse autoencoder and it consists of convolutional, pooling and a softmax layer(s). Multiple scales of an image are given as input, each one is processed separately and then joint in the second last layer. The unsupervised parts of the network are convolutional layers that are trained as autoencoders and the supervised part was fine-tuned using softmax regression. To evaluate the model's performance, they tested it through two breast cancer tasks (1) automated segmentation of breast density and (2) characterization of textural patterns in mammographies aiming to predict the chance of a woman developing cancer in the future.

https://proceedings.science/p/85436?lang=pt-br

Table 1: Summary of the recent works cited in this paper

Reference	Summary	Dataset	AUC
Wang et al. [2016]	Discrimination of tumor-patches and normal-patches using CNN with pre-processing procedures.	400 images	0.925 - 0.9948
Geras et al. [2017]	Multi-view classification.	103,000 images	Several experiments
Arevalo et al. [2015]	It uses pre-processed images and SVM in the classification layer.	736 images	0.86
Kooi et al. [2017]	Used two datasets – one to train and another to retrain the network. Also, used multiview approach.	1600 images	0.80 - 0.866
Huynh et al. [2016]	Used transferlearning combining two different models to come to a classification.	607 images produced from 219 images	0.86
Carneiro et al. [2015]	Pre-trained model for multi-view classification.	1090 images	0.91 - 0.97
Kallenberg et al. [2016]	Unsupervised learning model from breast cancer density segmentation and mammographic risk scoring using three different datasets	493, 668 and 1576 images	0.57 - 0.61

4. Method

The dataset used was CBIS-DDSM (Curated Breast Imaging Subset of DDSM) [Lee et al., 2016], it was launched by the The Cancer Imaging Archive (TCIA) project [Clark et al., 2013]. To run this experiment, 1146 full mammogram images were used (601 malignant and 545 benign) and split into three sets training, testing and validation, respectively 80%, 10% and 10% of the dataset. Amongst the images, there were mediolateral oblique (MLO) and craniocaudal (CC) views from left and right breasts.

As mentioned previously, the convolutional neural network model applied to this work was Inception V3 [Szegedy et al., 2015b]. Nonetheless, our purpose was not to train it from scratch, but to retrain the last layer of the network which was previously fully trained using ImageNet database. The training parameters were defined as 6,000 training steps, 100 batch size and 0.01 learning rate. No pre-processing procedure was applied to the images, simply, raw images were input to the network.

5. Results and Discussion

After several experiments, the average AUC was 0.72 (0.71 – 0.73). This result shows a performance slightly below the average as shown in Table 1. However, different from previous works, it was used pure raw data without any pre-processing or extraction of regions of interest. Moreover, it is not made distinction between MLO and CC mammograms in the experiment, then the model has to learn how to classify any image. Hence, its performance is great considering the difficulty of the task, even though has room for improvements.

6. Conclusion

Convolutional neural networks has been widely applied in breast image classification. However, because of the peculiarities of mammograms, small details that means a lot, it is a difficult task to build a classifier able to distinguish between a negative, benign and malignant exams. Recent works showed that the performance of these classifiers has not yet reached the same results as for other applications, such as general image recognition. One explanation can be the lack of medical

https://proceedings.science/p/85436?lang=pt-br

imaging data, which makes very tough the work of researchers. Also the necessity of using very large databases with high resolution images that make the training process more time consuming.

More recent researches suggested to apply certain transformations to the mammographies to increase the size of the datasets, for instance rotating and flipping the images or extracting areas of interest from them. In addition, pretraining CNN models using databases, such as ImageNet, has also shown to be a way to improve the adjustment of parameters of the network implying very good results.

The experiment showed in this paper used a pretrained model, but instead of adjusting all of its parameters, it was only retrained its last layer. Although it was a 2-class classification, it had some difficulties, for instance the use of raw and full mammograms and no distinction between MLO and CC views. Therefore, even though CNN is a promising technique and it could be seen by the results, a lot has to be done to make better the performance in breast cancer diagnosis.

References

Agrawal, A. K., Gans, J. S., e Goldfarb, A. (2017). What to expect from artificial intelligence. *MIT Sloan Management Review*, 58(3):23.

Al Rahhal, M. M., Bazi, Y., AlHichri, H., Alajlan, N., Melgani, F., e Yager, R. R. (2016). Deep learning approach for active classification of electrocardiogram signals. *Information Sciences*, 345:340–354.

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G., et al. (2016). Deep speech 2: End-to-end speech recognition in english and mandarin. In *International Conference on Machine Learning*, p. 173–182.

Andreas, A. (1986). Digital filters: Analysis and design.

Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., e Lopez, M. A. G. (2015). Convolutional neural networks for mammography mass lesion classification. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, p. 797–800. IEEE.

Balleyguier, C., Ayadi, S., Van Nguyen, K., Vanel, D., Dromain, C., e Sigal, R. (2007). Birads^{TM} classification in mammography. *European journal of radiology*, 61(2):192–194.

Barbu, A., Lu, L., Roth, H., Seff, A., e Summers, R. M. (2016). An analysis of robust cost functions for cnn in computer-aided diagnosis. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, p. 1–6.

Bolón-Canedo, V., Remeseiro, B., Alonso-Betanzos, A., e Campilho, A. (2016). Machine learning for medical applications. In *Proceedings of the European Symposium on Artificial Neural Networks*, p. 225–234.

Boureau, Y.-L., Ponce, J., e LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, p. 111–118.

Carneiro, G., Nascimento, J. C., e Bradley, A. P. (2015). Unregistered multiview mammogram analysis with pre-trained deep learning models. In *MICCAI (3)*, p. 652–660.

Cheng, Y., Wang, F., Zhang, P., e Hu, J. (2016). Risk prediction with electronic health records: A deep learning approach. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, p. 432–440. SIAM.

https://proceedings.science/p/85436?lang=pt-br

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., e Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*, 26(6):1045–1057.

Dahl, G. E. (2015). *Deep learning approaches to problems in speech recognition, computational chemistry, and natural language text processing*. PhD thesis, University of Toronto.

Deng, L., Yu, D., et al. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., e Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639): 115–118.

Gao, X., Li, W., Loomes, M., e Wang, L. (2017). A fused deep learning architecture for viewpoint classification of echocardiography. *Information Fusion*, 36:103–113.

Geras, K. J., Wolfson, S., Kim, S., Moy, L., e Cho, K. (2017). High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint arXiv:1703.07047*.

Gonzalez, R. C. e Woods, R. (2010). Processamento digital de imagens. tradução de cristina yamagami e leonardo piamonte.

Graves, A., Mohamed, A.-r., e Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, p. 6645–6649. IEEE.

Hassan, A. e Mahmood, A. (2017). Deep learning approach for sentiment analysis of short texts. In *Control, Automation and Robotics (ICCAR), 2017 3rd International Conference on*, p. 705–710. IEEE.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., e Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31.

He, K., Zhang, X., Ren, S., e Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, p. 1026–1034.

He, K., Zhang, X., Ren, S., e Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 770–778.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., et al. (2017). Cnn architectures for large-scale audio classification. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, p. 131–135. IEEE.

Hirschberg, J. e Manning, C. D. (2015). Advances in natural language processing. *Science*, 349 (6245):261–266.

Ho, R., Muraoka, M., Cuaresma, C., Guerrero, R., e Agbayani, A. (2010). Addressing the excess breast cancer mortality in filipino women in hawai ‘i through aancart, an nci community network program. *Hawaii medical journal*, 69(7):164.

https://proceedings.science/p/85436?lang=pt-br

Huynh, B. Q., Li, H., e Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3): 034501–034501.

Ibrahim, W. e Abadeh, M. S. (2017). Extracting features from protein sequences to improve deep extreme learning machine for protein fold recognition. *Journal of Theoretical Biology*, 421:1–15.

Jemal, A., Ward, E. M., Johnson, C. J., Cronin, K. A., Ma, J., Ryerson, A. B., Mariotto, A., Lake, A. J., Wilson, R., Sherman, R. L., et al. (2017). Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *JNCI: Journal of the National Cancer Institute*, 109(9).

Kallenberg, M., Petersen, K., Nielsen, M., Ng, A. Y., Diao, P., Igel, C., Vachon, C. M., Holland, K., Winkel, R. R., Karssemeijer, N., et al. (2016). Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE transactions on medical imaging*, 35(5):1322–1331.

Kooi, T., Ginneken, B., Karssemeijer, N., e Heeten, A. (2017). Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Medical physics*, 44(3):1017–1027.

Kurakin, A., Goodfellow, I., e Bengio, S. (2016). Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*.

Lai, M. (2015). Deep learning for medical image segmentation. *arXiv preprint arXiv:1505.02000*.

Lakhani, P. e Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, p. 162326.

LeCun, Y., Bengio, Y., e Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, R. S., Gimenez, F., Hoogi, A., e Rubin, D. (2016). Curated Breast Imaging Subset of DDSM. `http://dx.doi.org/10.7937/K9/TCIA.2016.7O02S9CY`.

Lopez, A. R., Giro-i Nieto, X., Burdick, J., e Marques, O. (2017). Skin lesion classification from dermoscopic images using deep learning techniques. In *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*, p. 49–54. IEEE.

Maas, A. L., Hannun, A. Y., e Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., e Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2 (1):1.

Sarikaya, R., Hinton, G. E., e Deoras, A. (2014). Application of deep belief networks for natural language understanding. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22(4):778–784.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61: 85–117.

Shen, D., Wu, G., e Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, (0).

https://proceedings.science/p/85436?lang=pt-br

Shimizu, R., Yanagawa, S., Monde, Y., Yamagishi, H., Hamada, M., Shimizu, T., e Kuroda, T. (2016). Deep learning application trial to lung cancer diagnosis for medical sensor systems. In *SoC Design Conference (ISOCC), 2016 International*, p. 191–192. IEEE.

Simonyan, K. e Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S., e Babu, R. V. (2016). A taxonomy of deep convolutional neural nets for computer vision. *arXiv preprint arXiv:1601.06615*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., e Rabinovich, A. (2015a). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., e Wojna, Z. (2015b). Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567.

Taigman, Y., Yang, M., Ranzato, M., e Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 1701–1708.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., e Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312.

Tsehay, Y. K., Lay, N. S., Roth, H. R., Wang, X., Kwaka, J. T., Turkbey, B. I., Pintob, P. A., Woodc, B. J., e Summersa, R. M. (2017). Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images. In *SPIE Medical Imaging*, p. 1013405–1013405. International Society for Optics and Photonics.

van Doorn, J. (2014). Analysis of deep convolutional neural network architectures. In *21st Twente Student Conference on IT*, volume 23, p. 3–5.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., e Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*.

Wu, Y. e Razavi, R. (2015). An introduction to deep learning: Examining the advantages of hierarchical learning. In *Technical report*. SAP affiliate company.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., e Gao, R. X. (2016). Deep learning and its applications to machine health monitoring: A survey. *arXiv preprint arXiv:1612.07640*.

https://proceedings.science/p/85436?lang=pt-br