



Revista Brasileira de Computação Aplicada, November, 2020

DOI: 10.5335/rbca.v12i3.10343 Vol. 12, Nº 3, pp. 132-141

Homepage: seer.upf.br/index.php/rbca/index

#### ORIGINAL PAPER

# Classification of x-ray images for detection of childhood pneumonia using pre-trained neural networks

Nator Junior Carvalho Costa<sup>[0,1,2]</sup>, Jose Vigno Moura Sousa <sup>[0,1,3]</sup>, Domingos Bruno Sousa Santos <sup>[0],1</sup>, Francisco das Chagas Fontenele Marques Junior <sup>[0],2</sup>, Rodrigo Teixeira de Melo <sup>6,2</sup>

<sup>1</sup>State University of Piauí UESPI - Brazil, <sup>2</sup>State University of Ceará UECE - Brazil, <sup>3</sup>University brazil - Brazil

\*nator@prp.uespi.br; josevigno@prp.uespi.br; domingosbruno@prp.uespi.br;junior.marques@aluno.uece.br;rodrigo.t.morpheus@gmail.com

Received: 2020-05-24. Revised: 2020-08-31. Accepted: 2020-10-15.

#### Abstract

This paper describes a comparison between three pre-trained neural networks for the classification of chest X-ray images: Xception, Inception V<sub>3</sub>, and NasNetLarge. Networks were implemented using learning transfer; The database used was the chest x-ray data set, which contains a total of 5856 chest x-ray images of pediatric patients aged one to five years, with three classes: Normal Viral Pneumonia and Bacterial Pneumonia. Data were divided into three groups: validation, testing and training. A comparison was made with the work of Kermany et al. (2018) who implemented the Inception V3 network in two ways: (Pneumonia X Normal) and (Bacterial Pneumonia X Viral Pneumonia). The nets used had good accuracy, being the NasNetLarge network the best precision, which was 95.35 % (Pneumonia X Normal) and 91.79 % (Viral Pneumonia X Bacterial Pneumonia) against 92.80 % in (Pneumonia X Normal) and 90.70 % (Viral Pneumonia X Bacterial Pneumonia) from kermany's work, the Xception network also achieved an improvement in accuracy compared to kermany's work, with 93.59 % at (Normal X Pneumonia) and 91.03 % in (Viral Pneumonia X Bacterial Pneumonia).

Keywords: Pneumonia; NasNetLarge; Xception; Classification; Inception V3; Chest-X-Ray;

#### Resumo

Este artigo descreve uma comparação entre três redes neurais pré-treinadas para a classificação de imagens de radiografia de tórax: Xception, Inception V3 e NasNetLarge. As redes foram implementadas usando transferência de aprendizado; o banco de dados utilizado foi o conjunto de dados de radiografia de tórax, que contém um total de 5856 imagens de radiografias de pacientes pediátricos com idade entre um e cinco anos, com três classes: Normal Pneumonia Viral e Pneumonia Bacteriana. Os dados foram divididos em três grupos: validação, teste e treinamento. Foi feita uma comparação com o trabalho de Kermany et al. (2018) que implementou a rede Inception V3 com dois modos: (Pneumonia X Normal) e (Pneumonia bacteriana X Pneumonia viral). As redes utilizadas tiveram boa precisão, sendo a rede NasNetLarge a melhor, que foi de 95,35 % (Pneumonia X Normal) e 91,79 % (Pneumonia Viral X Pneumonia Bacteriana) contra 92,80 % na (Pneumonia X Normal ) e 90,70 % (Pneumonia viral X Pneumonia bacteriana) do trabalho de kermany, a rede Xception também obteve uma melhoria na precisão em comparação com o trabalho de kermany, com 93,59 % em (Normal X Pneumonia) e 91,03 % em (Pneumonia viral X Pneumonia bacteriana).

Palavras-Chave: Pneumonia; NasNetLarge; Xception; Classificação; Inception V3; Chest-X-Ray;

### 1 Introduction

According to the World Health Organization, pneumonia is the largest infectious cause of death in children worldwide. Pneumonia killed 920136 children under 5 by 2015, accounting for 16% of all deaths of children under five years of age. Pneumonia affects children and families everywhere, but it is more prevalent in South Asia and sub-Saharan Africa (World, 2016).

Chest X-rays are often used to assess cases of pneumonia and are the most commonly used diagnostic tests for chest-related diseases. A very small dose of ionizing radiation is used to produce images of the chest (Kermany et al., 2018).

Low precision in the diagnosis of pneumonia is harmful to the patient, and this leads to the excessive prescription of antibiotics and to the waste of stocks of the same. Antibiotics also kill beneficial bacteria, causing unintended health problems in patients (Kurt et al., 2018). It is noteworthy that overuse leads to the proliferation of drug resistant bacteria, therefore the importance of a rapid and precise diagnosis.

Computational systems able to aid in the diagnosis and to identify the potential of diseases in patients are increasingly common (Manogaran et al., 2018). Used as a specialist support tool they can minimize errors (Malmir et al., 2017), being able to carry out a screening of potential patients.

With the increase of computational power, it became possible to use techniques such as Artificial Neural Networks (RNA) (Esteva et al., 2017), a technique that requires a high processing power in the training phase. RNA has its origin in a biophysical analogy to biological neurons (Koçer and Tümer, 2017).

In this way, it inspired the development of image classifier **childhood pneumonia** images, in order to diagnose patients in an automated and fast way. The method chosen and implemented is in the classification of chest x-ray images of patients, to determine if it has pneumonia or not, and is also classified what type of pneumonia, which may be bacterin or viral, for the classification was used the Chest X-Ray Dataset in total 5856 chest X-ray images of pediatric patients aged one to five years, provided with (Kermany, Zhang and Goldbaum, 2018).

In the prediction stage the classification of the images through artificial intelligence, consisting of Convolutional Neural Networks (CNN), together with the technique known as learning transference (Douarre et al., 2018). For purposes of comparison, two different CNNs were used, these being pre-trained neural networks known as NasNetLarge (Zoph et al., 2017) and Xception (Chollet, 2016), thus, an analysis is made of which network has a better performance for the proposed system. The proposed method guarantees robust coverage in image recognition, in certain aspects that will be clarified throughout the text.

The document is divided into 7 sections, in which Section 2 is characterized by the contextualization of the work. The methodology applied and the validation metrics in Section 3. The description of the database is

in Section 4, the evaluation metrics in Section 5, the results after application of the proposal are presented in Section 6 and Section 7 consists of the conclusion.

#### 2 Related work

In Rajpurkar et al. (2017), an algorithm is proposed to detect pneumonia of chest radiographs, where the authors affirm that the proposed algorithm exceeds the performance of radiologists. The algorithm is a convolutional neural network of 121 layers trained using the set of images of ChestX-ray14, containing more than 100,000 radiographic images with 14 diseases.

In Varoquaux et al. (2017), a review is made on cross-validation procedures for neuroimaging decoding, also includes a didactic overview of the relevant theoretical considerations. Practical aspects of common decoders are highlighted in predictions across multiple data sets. The experiments describe the large margins of error of cross-validation in neuroimaging configurations.

A clinical diagnostic tool is developed in Kermany et al. (2018), which is based on a deep learning framework for the screening of patients with retinal dazzling diseases treatable and childhood pneumonia. Transfer learning is used, where a neural network is trained with a data set of optical coherence tomography images and chest X-ray images, the authors state that the performance of the proposed method is comparable to that of human specialists.

In Zech et al. (2018), it is proposed a classification system of images of pneumonia, based on deep neural networks, an analysis of trained models is made using the junction of three sets of data. The validation was performed in a set of data from different hospitals, evaluated how well the trained models manage to generalize the classification, for a pneumonia screening task.

Deep neural network architectures are used to identify glaucomatous optic neuropathy in Christopher et al. (2018) ,the performance of neural networks is analyzed and also the impact of learning by transference. The author states that, in all cases, transfer learning achieved better performance, and also a reduction in training time.

In Abiyev and Maaitah (2018), we demonstrated the feasibility of classifying chest pathologies in chest X-rays using convolutional neural networks. For comparison, back-propagation neural networks with supervised learning, neural networks with unsupervised learning, training and testing is performed on the same radiographic database.

We present the work of a model based on stochastic attention in Kermany et al. (2018), this model is able to learn which regions within a chest x-ray should be visually explored, to conclude whether the radiograph contains a specific radiological abnormality. The proposed model is a recurrent neural network that sequentially learns the entire radiograph and focuses only on information areas that are likely to contain the relevant information.

# 3 Methodology

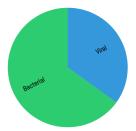
According to Shahin et al. (2004), to train a machine learning model it is necessary to divide the data into three sets (training, test and validation). According to Krawczyk (2016) the training data set is the sample of data used to fit the model, where the model sees and learns from that data.

The test dataset, however, is the data sample used to provide an unbiased assessment of a model fit in the training data set while adjusting the model hyperparameters (Krawczyk, 2016).

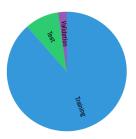
The validation data set, is the data sample used to provide an unbiased assessment of a final model, the validation data set provides the gold standard used to evaluate the model. It is only used when a model is fully trained (Esteva et al., 2017).

Ideally, the model should be evaluated on samples that were not used to construct or fit the model so as to provide an unbiased sense of model efficacy. In this case, the validation sets were used for this purpose where the implemented models were validated after being trained Fig. 1.

Fig. 1a shows the number of samples selected for training, testing and validation for both cases, Bacterial Pneumonia and Viral Pneumonia. Fig. 1b shows the total amount of samples for training, testing and validation for cases of Bacterial Pneumonia and Viral Pneumonia.



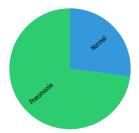
(a) 2488 images of Bacterial Pneumonia for training, 242 for testing and 50 for validation. 1295 images of viral pneumonia for training, 148 for test and 50 for validation.

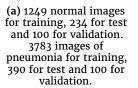


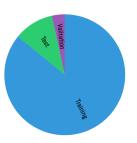
(b) Dividing the dataset into training, testing and validation. 3783 images for training, 390 images for test and 100 images for validation

**Figure 1:** Division of the dataset for the tests of Bacterial Pneumonia and Viral Pneumonia.

In the Fig. 2 the division used to carry out the tests is represented, the comparison proposed in this paper takes into account the structure described in Fig. 2. Thus, the pre-trained networks use the same structure.







(b) Dividing the dataset into training, testing and validation. 5032 images for training, 624 images for test and 200 images for validation.

**Figure 2:** Division of the dataset for the tests of pneumonia and normal.

Neural networks were implemented using the Keras deep learning API (Gulli and Pal, 2017), written in Python and capable of running on top of TensorFlow (Tang, 2016), CNTK (Seide and Agarwal, 2016), or Theano (Bergstra et al., 2010). The networks were processed using an NVIDIA Quadro P6000 24 GB video card, which has 3840 CUDA cores and an Intel Core i7 processor with 12 Gigabytes of RAM, though, most of the processing effort is done per video card, since CNN can run on GPUs, if available. The Table 1 shows the time for training the NasNetLarge x Xception networks, it is possible to notice that the training times are relatively low, however the Xception is up to 55% faster than the NasNetLarge, approximately.

### 3.1 Transfer Learning

Transfer learning is used to provide a pre-trained structure in a knowledge base that can be from the same or another domain, taking advantage of the knowledge acquired to solve new problems more quickly and effectively (Weiss et al., 2016, Lu et al., 2015).

In the state of the art the transfer of learning to solve problems is much more present, some examples are: Abidin et al. (2018), Douarre et al. (2018), Khatami et al. (2018), Baltruschat et al. (2018), Chen et al. (2018). The technique consists in using a pre-trained model with distinct classes of the problem to be solved (Wu et al., 2018), this becomes an advantage in the use of small data sets (Shallu and Mehra, 2018), because there is a difficulty in getting large sets of data for specific problems (Ramalingam and Garzia, 2018).

In the transfer of learning the initial layers and intermediates remain, being that the final layer is subsisted and trained again (Ramalingam and Garzia, 2018). In the Fig. 3 it is possible to visualize the flowchart in which it represents the transfer of learning.

For the training of neural networks, all weights are

- 1	1 . 1		
Reference paper	Neural network	Training time	
**	**	Pneumonia X Normal	Viral X Bacterial
This paper	NasNetLarge	17.16 min	09.69 min
This paper	Xception	07.09 min	04.24 min
Kermany (2018)	Inception V3	Null	Null

Table 1: Training time, between image loading and network training. All networks had 100 epochs.

defined as non-trainable, since they were trained with the Imagenet data set. In this way, the last layer of the networks is removed and four dense layers are added, with the latter having the same number of neurons as the number of classes to be classified. The Sofitmax function is used to activate the last layer of the networks Fig. 3.

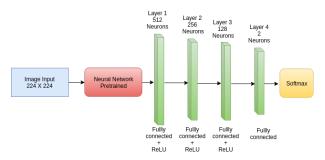


Figure 3: Transfer learning Architecture.

The hyperparameters for each of the networks can be analyzed in Table 2, the epoch parameter used for both networks for both Pneumoia X Normal and Viral X Bacterial classification was 100, this means the number of times the dataset is analyzed in each layer of the network, with this parameter the network Xception had an advantage over the NasNetLarge with regard to training time taking half the time to be trained Table 1. The Batch Size parameter (hyperparameter that defines the number of samples to work before updating the internal parameters of the model) is the 300 used for NasNetLarge and 200 for Xcepion.

A CNN is described as a sequence of layers, an example is shown in the Fig. 3, which is composed of three main layers, the convolutional layer, the pooling layer and the fully connected layer (Saraiva. et al., 2020), in fully connected layers, the use of ReLU functions demonstrated in the Eqs. (1) and (2) are common. These layers, when placed in sequence, form an architecture of a CNN (Salamon and Bello, 2017). In the output layer, the softmax function Eq. (3) is used, which is a generalization of the logistic function for various dimensions (Zhao, 2017)

$$f(x) = x^+ = \max(0, x) \tag{1}$$

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \ge 0 \end{cases}$$
 (2)

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$
 (3)

Choosing the optimization algorithm for a deep learning model can mean the difference between good results in minutes, hours, and days (Bottou et al., 2018). This is done using ADAM (Kingma and Ba, 2014) and SGD (Metz et al., 2018) optimizers, where they are quite present in the literature (Anil et al., 2018, Zoph and Le, 2016, Bello et al., 2017), Adam is used in NasNetLarge with a learning rate of 0.001, since for Xception the SGD is used with a learning rate of 0.002.

#### 3.2 NasNet Large

Zoph et al. (2017) proposed a learning model for image recognition in the Cifrar-10 and imageNet data sets, he defends the contribution of his work as a design of a search space in which he calls the "nasnet search space", which according to the author improves transferability to the network contains about 88.9 million parameters (Arend et al., 2018, Bressan et al., 2018).

### 3.3 Xception

The Xception network, proposed by Chollet (2016), is a derivation of the Inception V3 network (Szegedy et al., 2015) of google, that in the classification of the dataset imagenet obtained an improvement in the precision. A imageNet is a data set with 15 million labeled images (He et al., 2018). Xception has 22855952 parameters, which represents a reduction in quantity compared to Inception V3 (Chollet, 2016).

# 4 Description of the dataset

The set of images contains 5856 X-ray images (JPEG) and 3 categories (Viral Pneumonia, Bacterial Pneumonia and Normal) provided by the Kermany, Zhang and Goldbaum (2018). You can see in the Fig. 5, where they are divided into Normal X pneumonia (Fig. 5a and b) and Viral Pneumonia, Bacterial and Normal Pneumonia (Fig. 5c).

Chest x-ray images (anteroposterior) were selected from pediatric patients aged one to five years. The images come from the Guangzhou Women and Children Medical Center. All chest X-ray images were performed as part of the routine clinical care of the patients (Kermany, Zhang and Goldbaum, 2018).

The dataset still has quality control, where garbled,

Table 2: Training hyperparameters of NasNetLarge and Xception.

Neural Network	Initial Learning Rate	Optimiser	Batch Size	Max Epochs
NasNetLarge	0.001	Adam	300	100
Xception	0.002	SGD	200	100

low-quality images have been removed. The diagnosis was classified by two specialist physicians and checked by a third expert in order to extinguish the errors (Kermany, Zhang and Goldbaum, 2018).

The dataset consists of 5856 images, 1583 images of normal patients and 2780 of patients with bacterial pneumonia and 1493 of patients with viral pneumonia.

Pneumonia causes a pulmonary consolidation, this means that the pulmonary alveoli are full of inflammatory fluid, this liquid replaces the air in the alveoli, so that the affected part of the lung does not contain air (Iorio et al., 2018). In the radiographic, pulmonary consolidation corresponds to an opacity (whitish area).

The identification of the existence of pneumonia is based on the opacities of the radiography, ie, on the radiograph you can see the darker part near the spine that corresponds to the bronchi (Kunz et al., 2018). The air contained in the bronchi gives this color to the radiograph, while the outside of the lung is lighter (opaque) because the alveoli are filled with fluid Figs. 4 and 5.





(a) Normal image

(b) Image of pneumonia

**Figure 4:** Example of radiographic images divided into normal and with pneumonia.

Source: Kermany, Zhang and Goldbaum (2018), Kermany et al. (2018)

#### 5 Metrics of the evaluation

A statistical tool is the confusion matrix that provides the basis for describing classification accuracy and characterizing errors, helping to refine the ranking (Saraiva et al., 2018). The confusion matrix is formed by an array of squares of numbers arranged in rows and columns that express the number of sample units of a particular category, inferred by a rule of decision, compared to the category current verified in the field.

The measures derived from the confusion matrix are: the total accuracy being that chosen by the present work, accuracy of individual class, producer precision, user precision and Kappa index, among others. The

total accuracy is calculated by dividing the sum of the main diagonal of the error matrix  $x_{ii}$ , by the total number of samples collected n. According to the Eq. (4).

$$T = \frac{\sum_{i=1}^{a} \mathsf{x}_{ii}}{n} \tag{4}$$

Receiver Operating Characteristic Curve (ROC curve) is a measure of performance for classification problems in various boundary settings. The ROC is a probability curve and the ROC curve represents the degree or measure of separability. It informs how much model is able to distinguish between classes. the ROC curve is a curve that is drawn using the true positive rate and the false positive rate. The ROC curve is a complete sensitivity/specificity report.

In a ROC curve, the true positive rate (Sensitivity) is plotted against the false positive rate (specificity of 100) for different cut-off points of a parameter. Each point in the ROC curve represents a sensitivity / specificity pair corresponding to a given decision threshold.

To evaluate the performance of the classifiers of the present study, the confusion matrix, accuracy, along with the measures given by the ROC curve are used: sensitivity, specificity.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
 (5)

$$Sensitivity = \frac{TP}{TP + FN}$$
 (6)

$$Specificity = \frac{TN}{TN + FP}$$
 (7)

To exemplify a classification (Pneumonia X Normal), where true positive (TP) is the number of pneumonia samples that are correctly classified and true negative (TN) is the number of normal samples that are classified correctly. The false positive (FP) are Normal samples classified as wrong Pneumonia, and false negative (FN) are Pneumonia samples classified as Normal.

#### 6 Results

In this section we will present the performance results of neural networks. Accuracy data, ROC curve and matrix of confusion will be presented. Satisfactory results were obtained, compared to the paper of Kermany (2018). Two pre-trained neural networks



(a) Source: (Kermany et al., 2018)

Figure 5: Example of radiographic images divided into normal and with pneumonia. Source: (Kermany, Zhang and Goldbaum, 2018) and (Kermany et al., 2018)

**Table 3:** Test accuracy for the NasNetLarge, Xception and Inception V3.

Reference paper	Neural network	Accuracy test	
**	**	Pneumonia X Normal	Viral X Bacterial
This paper	NasNetLarge	95.35%	91.79 %
This paper	Xception	93.59%	91.03 %
(Kermany, 2018)	Inception V3	92.80 %	90.70 %

were implemented and the learning transfer was used, the networks used were: Xception and NasNetLarge. and showed an improvement in accuracy.

Data were divided into three groups: Training, testing and validation. On the Table 3 it is possible to visualize the accuracy that each network obtained, compared to Inception V3 implemented by the Kermany (2018).

As described above, a portion of the dataset was separated for validation. (Bacterial pneumonia X viral pneumonia) were separated 50 images of each, already in (Normal X pneumonia) were separated 100 images of each. On the Table 4 it is possible to visualize the results obtained by the networks, and the best results were attributed to the comparison of (Pneumonia The networks performed less than X Normal). the test in cases of (Pneumonia Bacteria X Viral Pneumonia). Thus, the network that had the best capacity to generalize was the NasNetLarge with 96.5% in (Pneumonia X normal) and 69.00% in (Pneumonia Bacteria X Viral Pneumonia). The two networks did well in generalization when the case was only detecting pneumonia. In the Figs. 6 to 9 it is possible to visualize some more results, represented by the matrices of confusion, curve ROC and Recall curve.

In Fig. 6 it is possible to see the confusion matrices and ROC curve for the classification of (Normal X Pneumonia) with the Xception network, with a good performance for the test and validation sets, with 99% acurracy in the validation, demonstrating that the model is able to generalize well, since in the test set it obtained a slightly lower percentage. In Fig. 7d it is possible to see that the validation performed considerably less, with 60% accuracy, this is due to the fact that the model has a greater difficulty in distinguishing differences between the classes of Viral

and Bacterial Pneumonia . In the Figs. 8 and 9 the results of the NasNetLarge network are shown, which had the same difficulty in distinguishing between the classes of viral and bacterial pneumonia, but with a slightly better performance, 69% accuracy .

In the Table 5 it is possible to visualize measures of Sensitivity and Specificity, which are metrics chosen to evaluate the performance of the classifiers, with this one can observe the comparative of the same, related to the classification of (Pneumonia X Normal) and (Pneumonia Bacterial X Viral Pneumonia). It is possible to see how the networks performed well in the tests and validation, and also how the proposed method performed better than Kermany (2018).

Table 4. Validation accuracy for the NasiverLarge and Acception networks			
Reference paper	Neural network	Accuracy validation	
**	**	Pneumonia X Normal	Viral X Bacterial
This paper	NasNetLarge	96.50%	69.00%
This paper	Xception	99.00%	60.00%

Table 4: Validation accuracy for the NasNetLarge and Xception networks.

**Table 5:** Specificity and sensitivity between the validation and test data sets, comparisons: (Pneumonia X Normal) and (Pneumonia Bacterial X Viral Pneumonia).

viiai riiedillollia).				
Network	Pneumonia X Normal			
Test				
Sensitivity Specificity				
NasNetLarge	95.23%	95.55%		
Xception	92.68%	95.32%		
Inception V3	93.2%	90.1%		
Validation				
**	Sensitivity	Specificity		
NasNetLarge	97.93%	95.14%		
Xception	100.00%	98.03%		
Inception V <sub>3</sub>	Null	Null		
**	Bacterial X Viral			
Test				
**	Sensitivity	Specificity		
NasNetLarge	91.01%	93.28%		
Xception	91.90%	89.51%		
Inception V <sub>3</sub>	88.60%	90.90%		
Validation		90.90%		
		90.90% Specificity		
Validation	1			
Validation **	n Sensitivity	Specificity		
	Test ** NasNetLarge Xception Inception V3 Validation ** NasNetLarge Xception Inception V3  ** Test ** NasNetLarge	Test           **         Sensitivity           NasNetLarge         95.23%           Xception         92.68%           Inception V3         93.2%           Validation         **           Sensitivity           NasNetLarge         97.93%           Xception         100.00%           Inception V3         Null           **         Bacteria           Test         **           Sensitivity           NasNetLarge         91.01%		

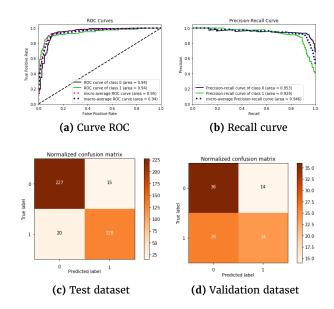
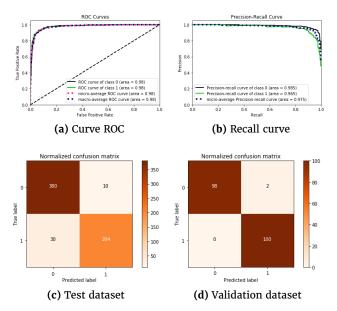
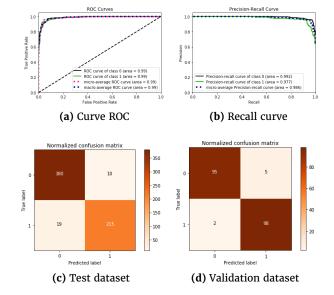


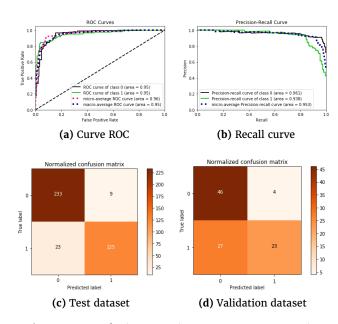
Figure 7: Confusion matrix, ROC curve in the comparison of Bacterial Pneumonia(0) X Viral Pneumonia(1) and confusion matrix of the validation of the neural network **Xception**.



**Figure 6:** Confusion matrix, ROC curve in the comparison of Normal(0) X Pneumonia(1) and confounding matrix of the validation of the neural network **Xception**.



**Figure 8:** Confusion matrix, ROC curve in the comparison of Normal(0) X Pneumonia(1) and confusion matrix of the validation of the neural network **NasNetLarge**.



**Figure 9:** Confusion matrix, ROC curve comparing Bacterial Pneumonia(0) X Viral Pneumonia(1) and confusion matrix of neural network validation **NasNetLarge**.

## 7 Conclusion

In this work, a comparison was made between the pre-trained neural networks Xception, Inception V3 and NasNetLarge with learning transfer for the classification of thoracic x-ray images in the detection of pneumonia. The use of pre-trained neural networks with learning transfer has been shown to be efficient in classification, as discussed in this work, the state of the art is optimistic about the use of this technique.

In the literature, applications for classification of pneumonia are presented that use approaches such as network construction, as is the case of the works Saraiva et al. (2019), Andika et al. (2019) and transfer of learning as an extractor of features such as the work of Togacar et al. (2020) that uses a combination of three pre-trunked networks, alexnet, VGG16 and VGG19 to make the classification. The main contribution of this work is in the detailed analysis of performance of the Xception and NasNetLarge networks in the classification of three classes, separated in pairs, classification of viral x bacterial pneumonia and classification between pneumonia and normal. Fundamentals that strengthen the use of transfer of learning for the classification of chest x-ray images of pathologies such as pneumonia are presented during the work, but with the potential to extend other classes.

The neural network NasNetlarge was the one that obtained the best result, even in the validation of (Bacterial Pneumonia X Viral Pneumonia) it demonstrated better than the Xception. However, the networks NasNetLarge and Xception had a good result in the generalization question in the comparison of

(Pneumonia X Normal), having a result in the validation similar to that of the test, thus demonstrating its capacity. The study of machine learning techniques in assisting medicine is promising, with the possibility of improving diagnoses in hospitals.

#### ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro P6000 GPU used for this research.

## References

Abidin, A. Z., Deng, B., DSouza, A. M., Nagarajan, M. B., Coan, P. and Wismüller, A. (2018). Deep transfer learning for characterizing chondrocyte patterns in phase contrast x-ray computed tomography images of the human patellar cartilage, *Computers in Biology and Medicine* **95**: 24 – 33. https://doi.org/10.1016/j.compbiomed.2018.01.008.

Abiyev, R. H. and Maaitah, M. K. S. (2018). Deep convolutional neural networks for chest diseases detection, *Journal of healthcare engineering* **2018**. https://doi.org/10.1155/2018/4168538.

Andika, L. A., Pratiwi, H. and Handajani, S. S. (2019). Klasifikasi penyakit pneumonia menggunakan metode convolutional neural network dengan optimasi adaptive momentum. https://doi.org/10.29244/ijsa.v3i3.560.

Anil, R., Pereyra, G., Passos, A., Ormandi, R., Dahl, G. E. and Hinton, G. E. (2018). Large scale distributed neural network training through online distillation, arXiv preprint arXiv:1804.03235. Available at https://arxiv.org/abs/1804.03235.

Arend, L., Han, Y., Schrimpf, M., Bashivan, P., Kar, K., Poggio, T., DiCarlo, J. J. and Boix, X. (2018). Single units in a deep neural network functionally correspond with neurons in the brain: preliminary results, *Technical report*, Center for Brains, Minds and Machines (CBMM). Available at https://cbmm.mit.edu/sites/default/files/publications/CBMM-Memo-093.pdf.

Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T. and Saalbach, A. (2018). Comparison of deep learning approaches for multi-label chest x-ray classification, CoRR abs/1803.02315. https://doi.org/10.1038/s41598-019-42294-8.

Bello, I., Zoph, B., Vasudevan, V. and Le, Q. V. (2017). Neural optimizer search with reinforcement learning, Proceedings of the 34th International Conference on Machine Learning-Volume 70, JMLR. org, pp. 459–468. Available at https://dl.acm.org/doi/pdf/10.5555/3305381.3305429.

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D. and Bengio, Y. (2010). Theano: a cpu and gpu math

- expression compiler, Proceedings of the Python for scientific computing conference (SciPy), Vol. 4, Austin, TX. Available at http://www.iro.umontreal.ca/~lisa/pointeurs/theano\_scipy2010.pdf.
- Bottou, L., Curtis, F. E. and Nocedal, J. (2018). Optimization methods for large-scale machine learning, *Siam Review* **60**(2): 223-311. https://doi.org/10.1137/16M1080173.
- Bressan, R. S., Alves, D. H., Valerio, L. M., Bugatti, P. H. and Saito, P. T. (2018). Doctor: the role of deep features in content-based mammographic image retrieval, 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), IEEE, pp. 158–163. https://doi.org/10.1109/CBMS.2018.00035.
- Chen, C., Dou, Q., Chen, H. and Heng, P. (2018). Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation, *CoRR* abs/1806.00600. https://doi.org/10.1007/978-3-030-00919-9\_17.
- Chollet, F. (2016). Xception: Deep learning with depthwise separable convolutions, *CoRR* **abs/1610.02357**. https://doi.org/10.1109/CVPR.2017.195.
- Christopher, M., Belghith, A., Bowd, C., Proudfoot, J. A., Goldbaum, M. H., Weinreb, R. N., Girkin, C. A., Liebmann, J. M. and Zangwill, L. M. (2018). Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs, *Scientific reports* 8(1): 16685. https://doi.org/10.1038/s41598-018-35044-9.
- Douarre, C., Schielein, R., Frindel, C., Gerth, S. and Rousseau, D. (2018). Transfer learning from synthetic data applied to soil—root segmentation in x-ray tomography images, *Journal of Imaging* 4(5). https://doi.org/10.3390/jimaging4050065.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks, *Nature* **542**(7639): 115. https://doi.org/10.1038/nature21056.
- Gulli, A. and Pal, S. (2017). Deep Learning with Keras, Packt Publishing Ltd.
- He, K., Girshick, R. B. and Dollár, P. (2018). Rethinking imagenet pre-training, CoRR abs/1811.08883. https://doi.org/10.1109/ICCV.2019.00502.
- Iorio, G., Capasso, M., Prisco, S., De Luca, G., Mancusi, C., Laganà, B., Piscopo, M. A. and Comune, V. (2018). Lung ultrasound findings undetectable by chest radiography in children with community-acquired pneumonia, *Ultrasound in medicine & biology*. http://dx.doi.org/10.1016/j.ultrasmedbio.2018.04.007.
- Kermany, D. S. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning,

- Cell 172(5): 1122 1131.e9. https://doi.org/10.1016/j.cell.2018.02.010.
- Kermany, D. S. et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning, *Cell* 172(5): 1122-1131. https://doi.org/10.1016/j.cell.2018.02.010.
- Kermany, D., Zhang, K. and Goldbaum, M. (2018). Large dataset of labeled optical coherence tomography (oct) and chest x-ray images, *Mendeley Data V*<sub>3</sub>. http://dx.doi.org/10.17632/rscbjbr9sj.3.
- Khatami, A., Babaie, M., Tizhoosh, H., Khosravi, A., Nguyen, T. and Nahavandi, S. (2018). A sequential search-space shrinking using cnn transfer learning and a radon projection pool for medical image retrieval, *Expert Systems with Applications* **100**: 224 233. https://doi.org/10.1016/j.eswa.2018.01.056.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980. Available at https://arxiv.org/abs/1412.6980.
- Koçer, S. and Tümer, A. E. (2017). Classifying neuromuscular diseases using artificial neural networks with applied autoregressive and cepstral analysis, *Neural Computing and Applications* **28**(1): 945-952. https://doi.org/10.1007/s00521-016-2383-8.
- Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions, *Progress in Artificial Intelligence* 5(4): 221–232. https://doi.org/10.1007/s13748-016-0094-0.
- Kunz, W. G., Patzig, M., Crispin, A., Stahl, R., Reiser, M. F. and Notohamiprodjo, M. (2018). The value of supine chest x-ray in the diagnosis of pneumonia in the basal lung zones, Academic radiology. https: //doi.org/10.1016/j.acra.2018.01.027.
- Kurt, I. S., Unluer, E. E., Evrin, T., Katipoglu, B. and Eser, U. (2018). Urine dipstick of sputum for the rapid diagnosis of community acquired pneumonia, *Journal of the National Medical Association*. https://doi.org/10.1016/j.jnma.2018.03.008.
- Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S. and Zhang, G. (2015). Transfer learning using computational intelligence: A survey, *Knowledge-Based Systems* 80: 14 23. https://doi.org/10.1016/j.knosys.2015.01.010.
- Malmir, B., Amini, M. and Chang, S. I. (2017). A medical decision support system for disease diagnosis under uncertainty, Expert Systems with Applications 88: 95–108. https://doi.org/10.1016/j.eswa.2017.06.031.
- Manogaran, G., Varatharajan, R. and Priyan, M. (2018). Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system, *Multimedia tools and applications* **77**(4): 4379–4399. https://doi.org/10.1007/s11042-017-5515-y.

- Metz, L., Maheswaranathan, N., Nixon, J., Freeman, C. D. and Sohl-Dickstein, J. (2018). Learned optimizers that outperform sgd on wall-clock and validation loss, arXiv preprint arXiv:1810.10180. Available at http://metalearning.ml/2018/papers/metalearn2018\_paper38.pdf.
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K. et al. (2017). Chexnet: Radiologist–level pneumonia detection on chest x-rays with deep learning, arXiv preprint arXiv:1711.05225. Available at https://arxiv.org/abs/1711.05225v3.
- Ramalingam, S. and Garzia, F. (2018). Facial expression recognition using transfer learning, 2018 International Carnahan Conference on Security Technology (ICCST), pp. 1–5. https://doi.org/10.1109/CCST.2018.8585504.
- Salamon, J. and Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification, IEEE Signal Processing Letters 24(3): 279–283. https://doi.org/10.1109/LSP.2017.2657381.
- Saraiva, A. A., Ferreira, N. M., de Sousa, L. L., Costa, N. J. C., Sousa, J. V. M., Santos, D. B. S., Valente, A. and Soares, S. (2019). Classification of images of childhood pneumonia using convolutional neural networks, *BIOIMAGING*. https://doi.org/10.5220/0007404301120119.
- Saraiva., A. A., Santos., D. B. S., Pedro., P., Sousa., J. V. M., Ferreira., N. M. F., Neto., J. E. S. B., Soares., S. and Valente., A. (2020). Classification of optical coherence tomography using convolutional neural networks, *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies Volume 3: BIOINFORMATICS*,, INSTICC, SciTePress, pp. 168–175. https://doi.org/10.5220/0009091001680175.
- Saraiva, A., Melo, R., Filipe, V., Sousa, J., Ferreira, N. F. and Valente, A. (2018). Mobile multirobot manipulation by image recognition, International Journal of Systems Applications, Engineering & Development 12. Available at https://www.naun.org/main/UPress/saed/2018/a222014-aay.pdf.
- Seide, F. and Agarwal, A. (2016). Cntk: Microsoft's open-source deep-learning toolkit, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 2135–2135. https://doi.org/10.1145/2939672.2945397.
- Shahin, M. A., Maier, H. R. and Jaksa, M. B. (2004). Data division for developing neural networks applied to geotechnical engineering, *Journal of Computing in Civil Engineering* **18**(2): 105–114. https://doi.org/10.1061/(ASCE)0887-3801(2004)18:2(105).
- Shallu and Mehra, R. (2018). Breast cancer histology images classification: Training from scratch or transfer learning?, *ICT Express* **4**(4): 247 254. https://doi.org/10.1016/j.icte.2018.10.007.

- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2015). Rethinking the inception architecture for computer vision, *CoRR* **abs/1512.00567**. https://doi.org/10.1109/CVPR.2016.308.
- Tang, Y. (2016). Tf. learn: Tensorflow's high-level module for distributed machine learning, *arXiv* preprint arXiv:1612.04251. Available at https://arxiv.org/abs/1612.04251v1.
- Toğaçar, M., Ergen, B., Cömert, Z. and Özyurt, F. (2020). A deep feature learning model for pneumonia detection applying a combination of mrmr feature selection and machine learning models, *IRBM* **41**(4): 212 222. https://doi.org/10.1016/j.irbm.2019.10.006.
- Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y. and Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines, *NeuroImage* 145: 166-179. https://doi.org/10.1016/j.neuroimage. 2016.10.038.
- Weiss, K., Khoshgoftaar, T. M. and Wang, D. (2016). A survey of transfer learning, *Journal of Big data* **3**(1): 9. https://doi.org/10.1186/s40537-016-0043-6.
- World, Health, O. (2016). World, health, organization pneumonia. Available at https://www.who.int/en/news-room/fact-sheets/detail/pneumonia.
- Wu, Y., Qin, X., Pan, Y. and Yuan, C. (2018). Convolution neural network based transfer learning for classification of flowers, 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP), pp. 562–566. https://doi.org/10.1109/SIPROCESS.2018.8600536.
- Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J. and Oermann, E. K. (2018). Confounding variables can degrade generalization performance of radiological deep learning models, arXiv preprint arXiv:1807.00431 . https://doi.org/10.1371/journal.pmed.1002683.
- Zhao, W. (2017). Research on the deep learning of the small sample data based on transfer learning, AIP Conference Proceedings, Vol. 1864, AIP Publishing LLC, p. 020018. https://doi.org/10.1063/1.4992835.
- Zoph, B. and Le, Q. V. (2016). Neural architecture search with reinforcement learning, arXiv preprint arXiv:1611.01578 . Available at https://arxiv.org/abs/1611.01578v2.
- Zoph, B., Vasudevan, V., Shlens, J. and Le, Q. V. (2017). Learning transferable architectures for scalable image recognition, *CoRR* abs/1707.07012. https://doi.org/10.1109/CVPR.2018.00907.