

xRayAID

Detecting Pneumonia using Artificial Intelligence

Vinicius Trevisan¹,
Edmar R. S. Rezende¹, Daniele U. M. Rodrigues¹

¹Pontifical University Catholic of Campinas
Campinas – SP – Brazil

vinicius_trev@hotmail.com,

{edmar_rezende,danielerodrigues}@puc-campinas.edu.br

Abstract. *Pneumonia is a type of acute respiratory infection that impacts people's lives in several ways, demanding an accurate and fast diagnosis. High death rates, massive socioeconomic impacts, and a significant gap between the number of available doctors based on its geographic location are some of the problems surrounding this topic. The xRayAID is a classification tool capable of receiving frontal thorax radiography to speed up the diagnosis of pneumonia using machine learning. That was done by using a slightly modified DenseNet-121 network architecture trained on the Radiological Society of North America (RSNA) public dataset. The results showed that this tool is able to help doctors to identify pneumonia scenarios, achieving a validation accuracy of 87.9%.*

1. Introduction

In 2012, the World Health Organization (WHO) stated that pneumonia was the pathology that killed the most number of children in the world, being at least 1.2 million per year, surpassing the deaths caused by HIV, malaria and tuberculosis summed. From those deaths, more than 99% were recorded in underdeveloped and developing countries, like Brazil, making them improve their techniques of combat and prevention of this disease [Fundação Oswaldo Cruz 2020].

In Brazil, data from the DATASUS government platform showed that only in 2019, more than 630 thousand hospitalizations occurred due to pneumonia, resulting in more than 60 thousand deaths, giving costs of approximately 2 million dollars to the public funds. This disease can be fatal, especially for people over 60 years, since more than 49 thousand deaths occurred on older people (81,6% of the deaths in 2019). Between 2015 and 2019, nearly 288 thousand deaths occurred, an average of 57,6 thousand cases per year, or 6 per hour [DATASUS 2020].

One exam applied in 2016 by São Paulo's regional medical council (CREMESP), focused on evaluating the knowledge of newly graduated doctors, discovered that more than 50% of them failed the exam. This test stated that 80% of the participants were unable to interpret one radiography and took wrong actions when conducting treatment for elderly persons [CREMESP 2020], something that is a routine at a clinic on a daily basis. In addition, 75% of the participants were also unable to identify the main characteristics of respiratory diseases. This exam showed that most of the newly graduated professionals were unable to diagnose pneumonia efficiently.

Another interesting factor is that in 2018, the rate of physicians in the entire Brazil per thousand persons was 2.18, against 5.07 in the capitals and 1.28 in the central regions of the country. Those numbers showed that there is a visible disparity of physicians between the different places of the country, especially in North and Northeast, with rates of 1.16 and 1.41 health professionals per thousand persons [Scheffer et al. 2018]. That difference strikes directly the availability of physicians in the country, affecting right in the health sector.

Thorax X-rays are one of the most efficient ways to diagnose pneumonia [Organization et al. 2001], these exams are essentials at epidemiological studies and clinics, being available on a large scale due to the low-cost pieces of equipment needed to identify it. However, diagnose pneumonia in thorax radiography can be a complex task, at the same time that as soon as the diagnose is achieved, bigger are the chances of success on the treatment therapies [Guzzetta et al. 1983].

This paper describes a platform that will work as an auxiliary tool, being capable to help different kinds of physicians, especially at public health care, to diagnose pneumonia with precision and accuracy. The solution uses a slightly modified DenseNet-121 architecture, achieving a validation accuracy of 87.9%.

The remaining of this paper is structured as follows: in Section 2, we present the related work; in Section 3, we describe our methods to achieve the expected results; in Section 4, we present the results and findings; finally, we draw our conclusions and possible further works in Section 5.

2. Related Work

The main work that inspired this project was the [Rajpurkar et al. 2017]. It consists of a deep artificial convolutional neural network DenseNet-121 capable of detecting pneumonia on frontal thorax radiography trained on the ChestX-ray-14 dataset. That dataset consists of 112,120 frontal radiography of 30,805 different patients. This dataset wasn't used in this project because it lacks sufficient examples of pneumonia since it includes 13 other pathologies. The developed application exhibits, in the end, a heat-map that indicates the possible regions affected by the pathology, something similar to what was done in this project. This approach achieves an F1-Score of 0.435 (with a confidence interval of 95% between [0.387, 0.481]), overcoming the average of four specialists in the area, as they obtained an average F1-Score of 0.387 (with a confidence interval of 95% between [0.330, 0.442]).

The dataset used to train the AI model developed in this paper was the RSNA [RSNA 2020], that was released as a Kaggle challenge in 2018. Some metrics will be described in Section 3, but that dataset was selected because it has more relevant examples to be used for training and validation compared to the ChestX-ray-14.

The [Stephen et al. 2019] work also consists of a deep convolutional neural network to detect pneumonia. The developed model was subdivided into two major parts, the first one, also called feature extractor, and the classifier (using sigmoid as activation function). The dataset used contains 5,856 labeled posteroanterior (PA) radiography from children between 1 and 5 years old, divided into 63.5% for training and 36.5% for validation. As the database used was relatively small, some data-augmentation techniques

were also used as vertical rotation, horizontal zoom and others. The network hyperparameters also needed some fine-tuning to achieve the accuracy of 93% at the validation data. The main problem of the described approach were that the dataset lacks examples that generalize the problem, since it only includes pediatric X-Rays.

Finally, given the current pandemic scenario caused by Sars-CoV-2, some recent researches came as a way to apply AI algorithms to diagnose that disease using only radiography, since the condition affects mainly epithelial cells that belong to the respiratory system. That's the case of the work developed by [Sethy and Behera 2020]. With limited test kits for the recent disease, a tool to diagnose COVID-19 was rapidly developed by the use of transfer learning on a ResNet-50 model in combination with SVM to extract deep features for the model. That project achieved an F1-Score of 95,52% with only 183 examples.

Despite all these contributions, none of them merged the machine learning techniques into a system designed for the end-user. That's why, based on the described ideas, we create a system synthesizing all the concepts to reduce human biases; because respiratory infections, associated with the described problems, pose a chronic obstacle, with considerable mortality rates.

Such factors, justify the rationalization of a system that links medicine to computing, being the main goal of this project assist doctors in decisions making regarding a diagnosis of pneumonia. The idea is to propose a tool and a new way to carry out diagnostics in the health field, mainly in the areas of public health, hospitals, clinics and emergency care, using machine learning algorithms.

3. Methodology

The developed project encompasses two large blocks and has a single user, defined as a doctor, as can be seen in the architecture diagram displayed in Figure 1.

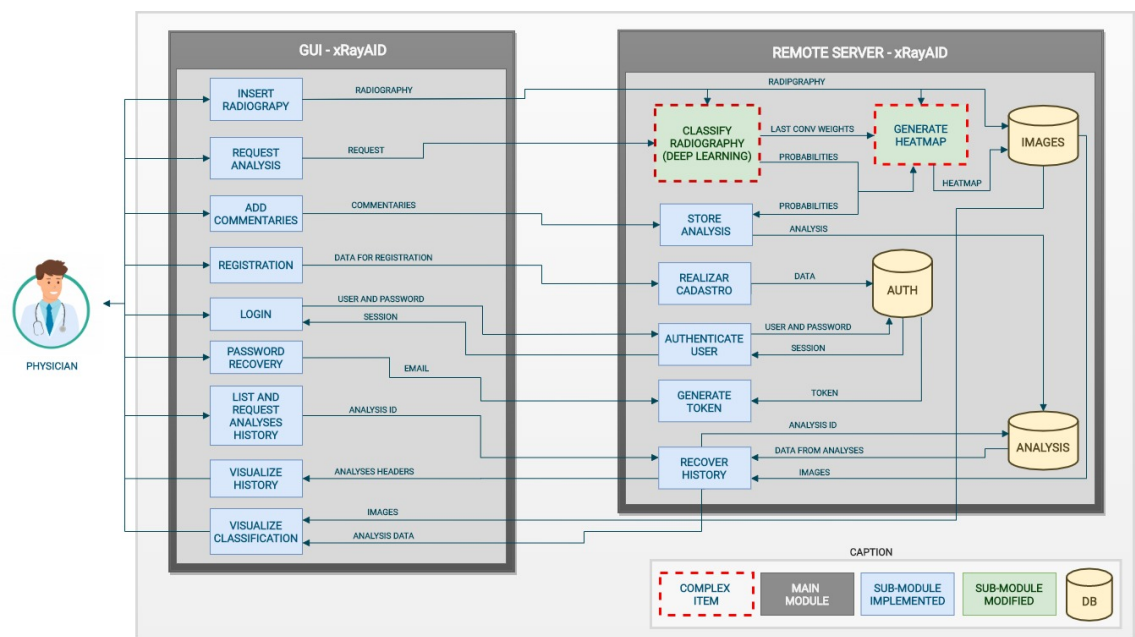


Figure 1. Project Architecture Diagram.

For its development, several programming tools were used. In the machine learning part, the Python programming language was used, considering mainly Keras, TensorFlow, NumPy and OpenCV frameworks. For the website, technologies such as HTML, CSS, JavaScript and PHP were used. All communication performed by the Python application and the website was done through a Flask API, thinking mainly at a possible expandability and distribution of the tool.

All user data (username, password, classifications metadata and annotations) was stored on MongoDB instances, being the password encrypted with the BCrypt algorithm. The images received for classification and its results were stored on a file structure as PNG or JPEG files.

The algorithm used for supervised machine learning is based on a DenseNet-121 architecture [Huang et al. 2017], defined as CheXNet [Rajpurkar et al. 2017]. In order to meet the needs of the project, some modifications in the original model, as seen in Figure 2, were necessary, resulting in a new model as shown in Figure 3.

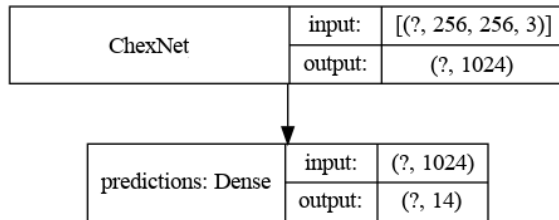


Figure 2. Original Model.

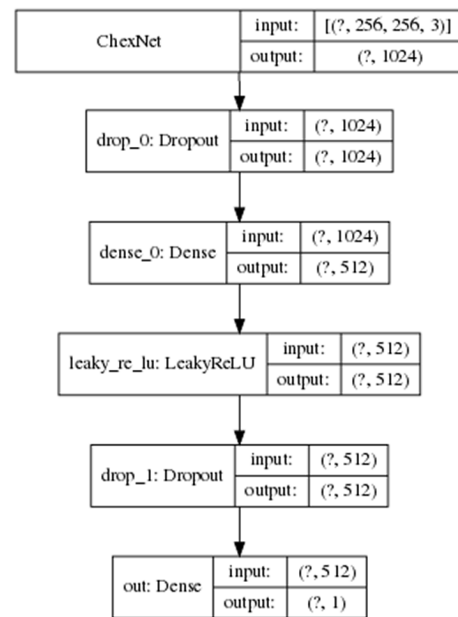


Figure 3. Modified Model.

Regarding the information used for training the AI model, a public dataset was used, made available by the Radiological Society of North America [RSNA 2020]. This dataset has a total of 26,684 labeled examples, segmented into 6,012 and 20,672 images that present and do not present, respectively, the pathology to be classified, representing a rate of 0.29 images with pneumonia for one without it.

All examples used to extract characteristics went through a pre-processing and data augmentation phase. Such a phase can be performed in 5 possible random iterations grouped per image, as shown in Figure 4. That way, we obtained more than 130,000 examples. The phases are defined as:

1. Maintain original image;
2. Horizontal mirroring of the radiography;
3. Conversion of the RGB image to CIELAB color-space, applying automatic filters to improve contrast;
4. Zoom with random factors between 0.8x and 1.2x;
5. Rotation with random factors between -10° and $+10^\circ$;

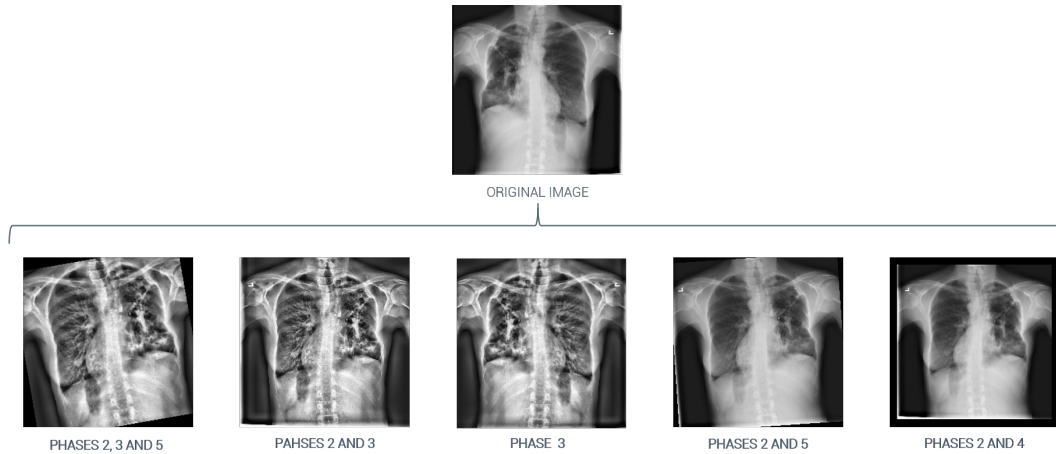


Figure 4. Data Augmentation Phases Exemplified.

For the model training, 5-folds of the entire dataset were made respecting, always, the proportion of 80% for training and 20% for validation. The parameters used in this step consist of an input format of (224, 224, 3), 16 epochs per fold and batch size of 16 images. The hardware used for this step consisted of 32 GB of RAM, an Intel Core i7 3930K CPU and a Nvidia RTX 2070 GPU. In order to obtain the best possible result, the best generation for each of the five folds was saved and later used for cross-validation. That way, we defined which weights best fit the problem.

In order to evaluate the training of the developed AI model, metrics of accuracy, loss and area under the receiving operating characteristic curve (AUC - ROC) were evaluated.

To simplify and facilitate the visualization of the results by the end-user, an algorithm was developed capable of plotting a heat-map that displays the possible regions affected by the pathology. In this case, as the tone becomes yellowish, greater is the probability of the disease in that specific area. The development of this functionality used the following flow: from the feature maps of the last convolution layer of the model, the function ReLU (Rectifier Linear Unit) was applied, transforming all negative values into null; therefore, these values were multiplied by the probability of pneumonia, colored and overlaid on the original image.

For the website aspect, the methodology used was to start with simpler functionalities, expanding it to complex tasks. The development began implementing the home page, login and register functions, describing mainly what was the purpose of the tool. Later on, we expand the system, including a user-page with a tutorial on how to classify and visualize the results, integrating it with the Flask API developed before.

The developed application was embedded on an Nvidia Jetson Nano development board. For that, we used Debian *systemd* service manager to start and control the API initialization, handling possible exceptions. The website was deployed using *Nginx* with encrypted traffic via HTTPS. We also have a dedicated email server, installed with *postfix* and *dovecot*.

4. Results

The number of parameters for the developed machine learning model is being detailed in Table 1 for each layer of the model. In Table 2, we described the total, trainable and non-trainable number of parameters. With that number of parameters and the configurations described in previous section, each epoch of training took approximately 20 minutes.

Table 1. Model Parameters by Layer.

Layer Type	Output Shape	Parameters
ChexNet	(?, 1024)	7.037.504
Dropout 0	(?, 1024)	0
Dense 0	(?, 512)	524.800
Leaky ReLU	(?, 512)	0
Dropout 1	(?, 512)	0
Output	(?, 1)	513

Table 2. Model Parameters Statistics.

	Parameters
Total	7.562.817
Trainable	7.479.169
Non-trainable	83.648

After the training step with the five folds described previously, we generate three validation graphics indicating the accuracy in Figure 5, loss in Figure 6 and the AUC - ROC curve in Figure 7, that last one indicating how much the model is capable of distinguishing between radiography with disease and no disease.

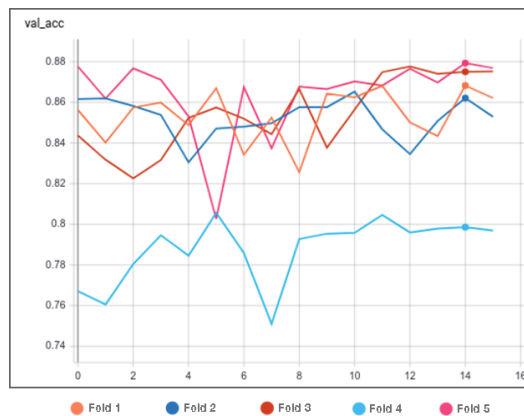


Figure 5. Validation Accuracy.

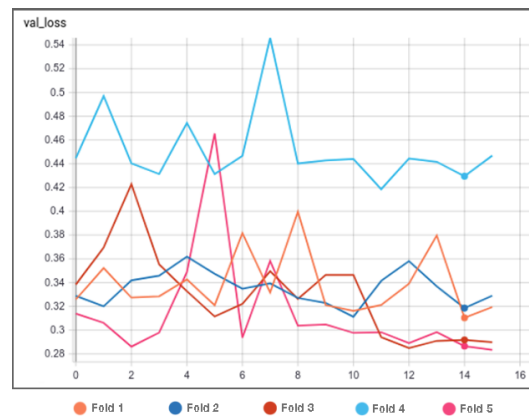


Figure 6. Validation Loss.

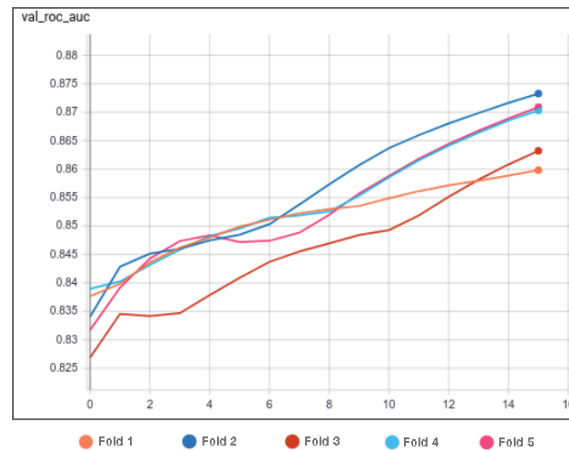


Figure 7. Validation AUC - ROC.

From those graphics, the best combination encountered as a way to obtain the best-trained model to fit our image samples was the one found on fold number 5 at the 15th epoch. With a validation accuracy of 87.9%, validation loss of 28.7% and validation AUC - ROC curve of 86.9%; that was the selected weights for the final application.

The average score of the 5-fold cross-validation was an accuracy of $85.3\% \pm 2.9\%$, loss of $33.3\% \pm 5.9\%$ and AUC - ROC curve of $87.0\% \pm 0.5\%$.

Just to exemplify, Figure 8 shows one result after the classification step and the heatmap overlay. In the example, the original image and the classified one are displayed side-by-side with the pneumonia probability at the top of the second image.

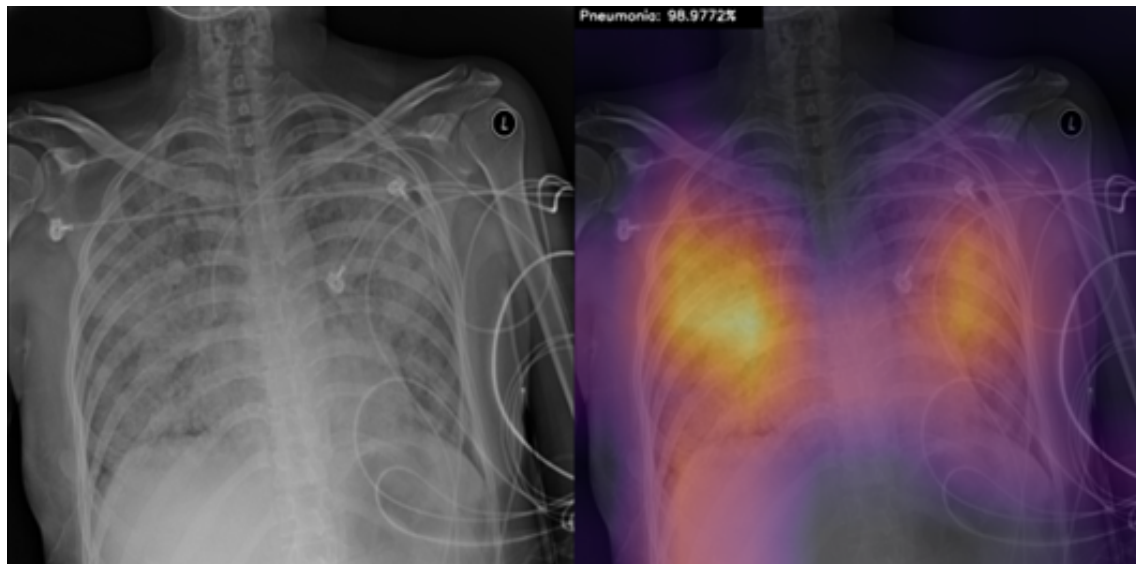


Figure 8. Classification and Heatmap Example.

To validate the proposed solution with the end-user, 26 non anonymized radiography from a local clinic was used. The classification results were analyzed by two physicians, and both of them returned positive feedback about the classifications, pointing only some misplaced regions on the heatmap pattern.

As for the developed website¹, it begins with a homepage describing the project, its objectives and functionalities, as seen in Figure 9. That page is responsible to authenticate the user, giving him access to use the system after previous registration. Also on that page, the administrators can access a webmail server used to communicate with customers.

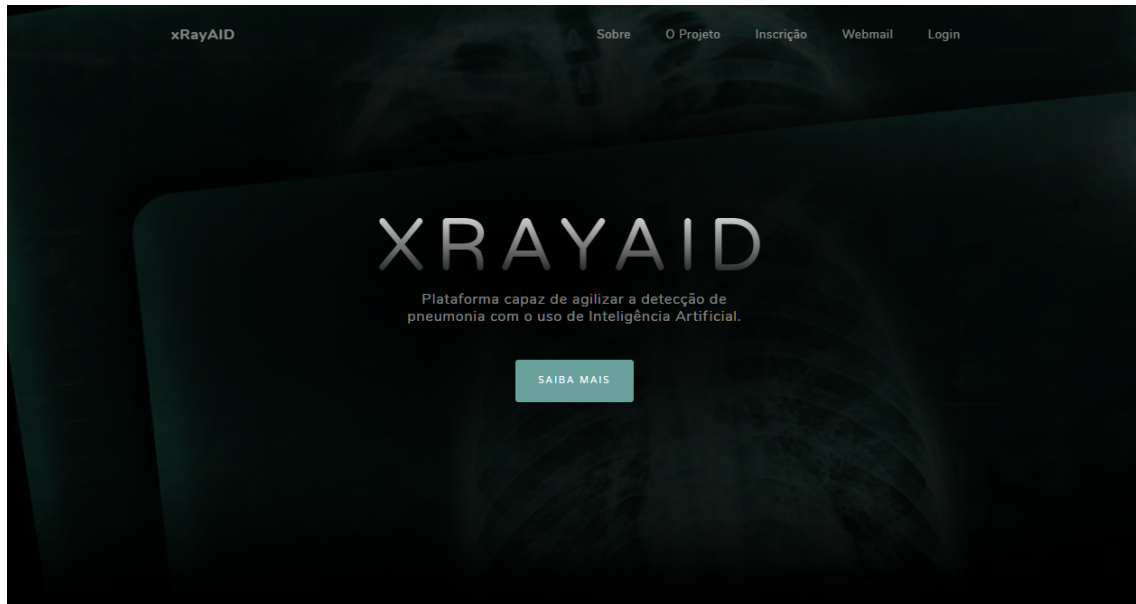


Figure 9. Website Homepage.

The authentication was done via *bcrypt* algorithm since it is considered one of the most secure password-hashing functions. Figure 10 and Figure 11 demonstrates the registration and login page, respectively. Both are protected using Google's *Captcha*.

Figure 10. Registration Form.

Figure 11. Login Form.

For each user, after authentication, there is a user page that exhibits some news about the project and a how-to video demonstrating its capabilities and functionalities. On

¹<https://xrayaid.com.br>

that page, there is a lateral menu bar responsible to redirect the user to the desired functionality. There are current 3 options: homepage, new classification and history pages, that last one responsible to show all the records already made at the system by the user.

By selecting the option to perform a new classification, a page similar to the Figure 12 is displayed. The system support images in PNG or JPEG formats only, at a maximum of 10 megabytes per image. At that same page, after the server finishes its processing, the page is automatically reloaded displaying the results view, similar to the Figure 13.

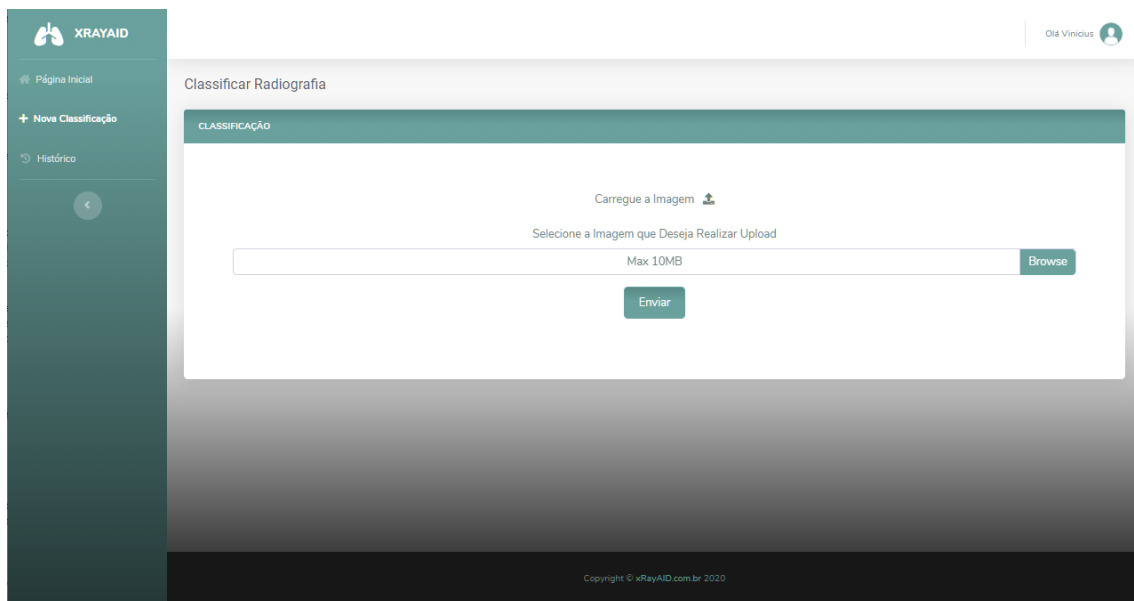


Figure 12. New Classification Page.

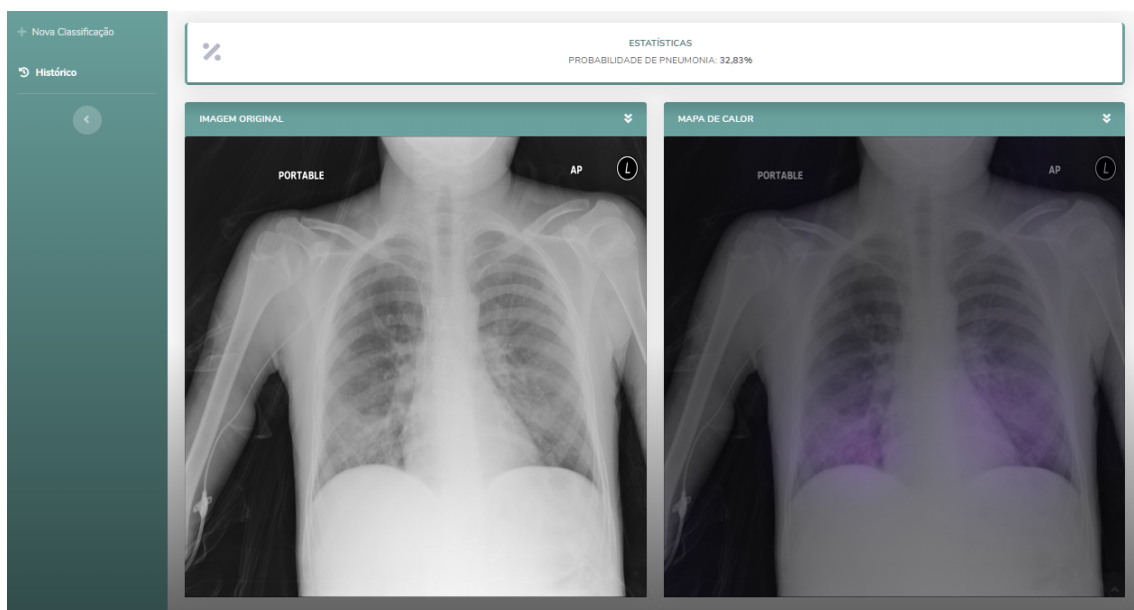


Figure 13. Results View.

On the view page, it is possible to visualize, at the top, the pneumonia probability as a percentage. That page also shows the original image and the heatmap side-by-side. That view allows the user to zoom-in the images independently. It also can apply filters to improve contrast and brightness on both images, as exemplified in Figure 14. The user can also write relevant annotations and provide us some feedback about the produced results, as demonstrated in Figure 15.

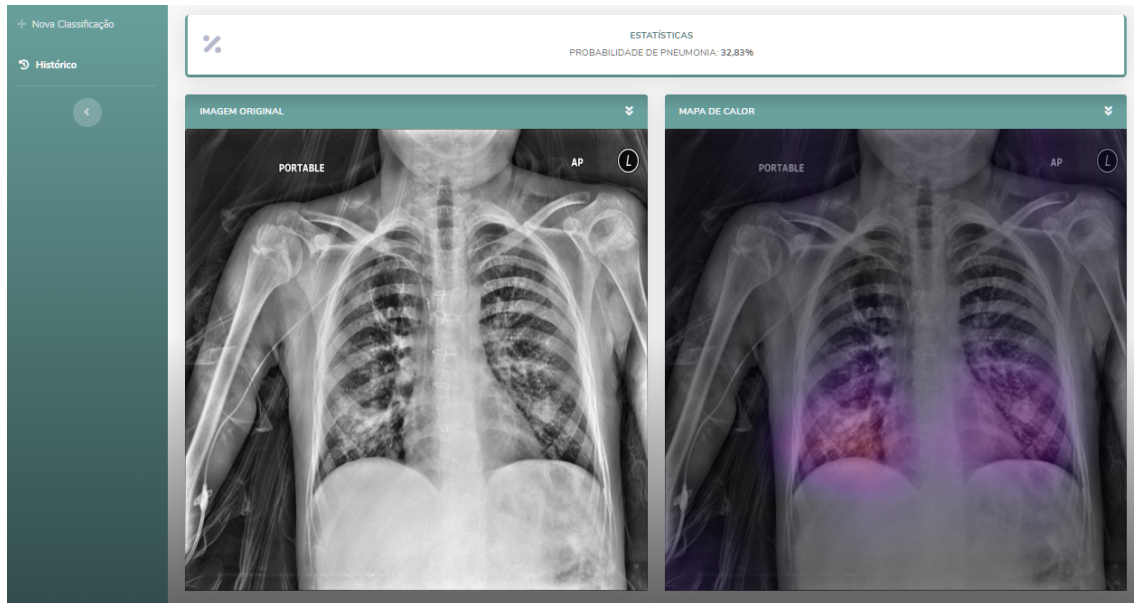


Figure 14. Results View with Filters.

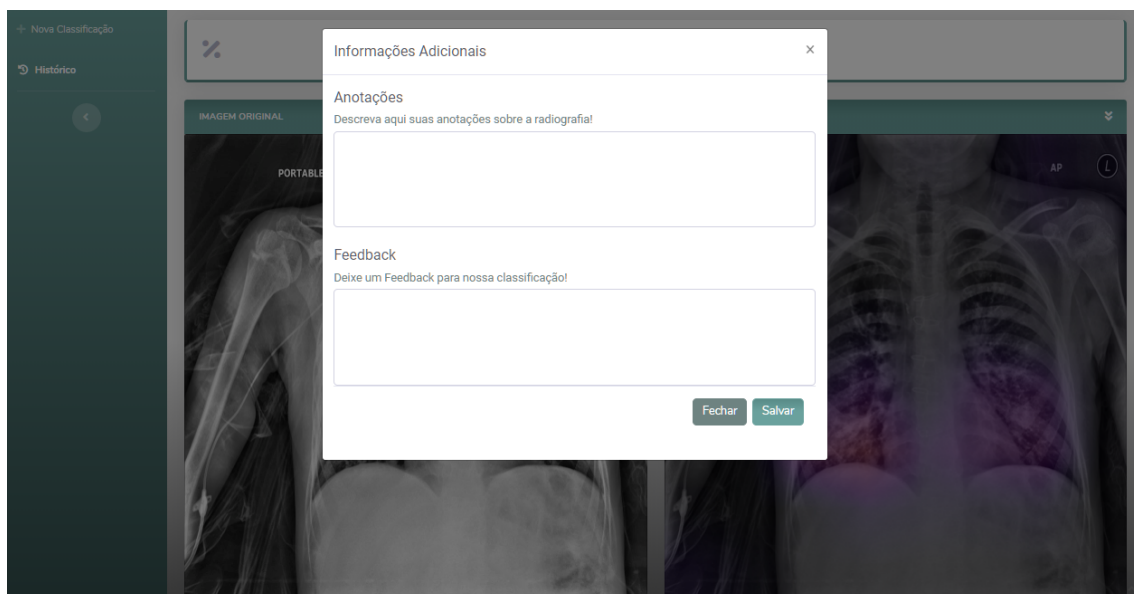


Figure 15. Annotations Form.

The last developed feature was the ability to recover all the classifications already done by the user, listing them in a user-friendly table capable of applying filters and custom search for keywords, as seen in Figure 16. To visualize the desired classification it is just necessary to click on the corresponding green button "Visualize". After that, the request is loaded and displayed as the same view page described in Figure 13. The history page can also delete undesired classifications via the trashcan icon.

Visualizar Histórico

CLASSIFICAÇÕES REALIZADAS

Exibir 10 entradas Procurar:

	Identificador	Probabilidade	Data	Horário	
	1600899411	18,99%	23/09/2020	19:16:51	
	1600223722	62,67%	15/09/2020	23:35:22	
	1600116323	45,88%	14/09/2020	17:45:23	
	1599603610	27,56%	08/09/2020	19:20:10	
	1599528321	32,83%	07/09/2020	22:25:21	
	1599468790	45,88%	07/09/2020	05:53:10	
	1599146099	28,67%	03/09/2020	12:14:59	
	1599146023	28,67%	03/09/2020	12:13:43	
	1599143197	1,78%	03/09/2020	11:26:37	
	1599094841	32,83%	02/09/2020	22:00:41	

Exibindo 1 a 10 de 96 entradas

Anterior 1 2 3 4 5 ... 10 Próxima

Figure 16. History Page.

As a way to keep the machine learning core on the edge, simplifying the communication between the client and the server and keeping the data safe, all the proposed solution was embedded in a Nvidia Jetson Nano 4GB development board. That board was able to handle all the proposed system and took less than 30 seconds to process one classification request.

5. Conclusions

It is noticeable that pneumonia can impact all kinds of persons, and it can be fatal. Previous works have shown that this disease can be diagnosed with computer vision assistance, as described in this paper.

The results showed that the developed system can predict and detect the possible pneumonia region on x-Ray images, acting as an auxiliary tool for physicians to diagnose pneumonia. That's achieves all the objectives proposed for this work.

As for further works on the system, the machine learning model needs some improvements since it isn't generic enough to handle all kinds of radiography that exist nowadays, as different x-Ray machines produce different results. It also needs some fine-tuning to improve the heatmap localization overlay. For the website, some user interface (UI) modifications can be done, especially at userpage and history listing, adding more functionalities to enhance the tool. It would also be useful to implement remember-me and user-info panel functions.

References

- CREMESP (2016 (accessed March 7, 2020)). Exame do CREMESP reprova mais da metade dos médicos recém-formados. <https://www.cremesp.org.br/pdfs/releasefinal2examecremesp2016.pdf>.
- DATASUS (2020 (accessed March 7, 2020)). Tecnologia da Informação a Serviço do SUS. <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/nruf.def>.
- Fundação Oswaldo Cruz (2013 (accessed March 1, 2020)). Pneumonia. <https://agencia.fiocruz.br/pneumonias>.
- Guzzetta, P., Toews, G. B., Robertson, K. J., and Pierce, A. K. (1983). Rapid diagnosis of community-acquired bacterial pneumonia. *American Review of Respiratory Disease*, 128(3):461–464.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Organization, W. H. et al. (2001). Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children. Technical report, World Health Organization.
- 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*.
- RSNA (2018 (accessed March 25, 2020)). RSNA Pneumonia Detection Challenge. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/overview>.
- Scheffer, M., Biancarelli, A., and Cassenote, A. (2018). Demografia médica no brasil 2015 [internet]. *São Paulo: Departamento de Medicina Preventiva, Faculdade de Medicina da USP*, page 284.
- Sethy, P. K. and Behera, S. K. (2020). Detection of coronavirus disease (covid-19) based on deep features. *Preprints*, 2020030300:2020.
- Stephen, O., Sain, M., Maduh, U. J., and Jeong, D.-U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering*, 2019.