xRayAID
Detecting Pneumonia Using Artificial Intelligence

Vinicius Trevisan[1],
Daniele U. M. Rodrigues[1], **Edmar R. S. Rezende**[1]

[1]Pontifical Catholic University of Campinas
Campinas – SP – Brazil

vinicius_trev@hotmail.com,

{danielerodrigues,edmar_rezende}@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 tool that uses machine learning to assist doctors in diagnosis of pneumonia on frontal chest radiographs. That was done by using a modified DenseNet-121 neural 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 state medical universities and the knowledge of newly graduated doctors, discovered that more than 50% of the enrolled failed. 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.

Chest 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 chest 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, aiming to help different kinds of physicians to diagnose pneumonia with precision and accuracy. The solution uses a slightly modified DenseNet-121 architecture trained on the Radiological Society of North America (RSNA) public dataset, 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

[Rajpurkar et al. 2017] presented a 121-layer convolutional neural network called CheXNet able to detect pneumonia on frontal chest radiography trained on the ChestX-ray-14 dataset. ChestX-ray14 is a medical imaging dataset which comprises 112,120 frontal-view X-ray images of 30,805 (collected from the year of 1992 to 2015) unique patients with the text-mined fourteen common disease labels, mined from the text radiological reports. The developed application exhibits, in the end, a heat-map that indicates the possible regions affected by the pathology. This approach achieves an F1-Score of 0.435 (with a confidence interval of 95% between [0.387, 0.481]) and AUROC of 76.8%, 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]).

[Han et al. 2021] proposed a new way to classify chest X-rays applying radiomic features and contrastive learning using a Residual Attention Network (ResNetAttention) as a backbone. It uses the RSNA dataset with a data distribution between training and validation phases of 75% and 25%, respectively. Their evaluation results achieved an accuracy of 85.4% and AUROC of 87.7%. The main disadvantage of this method is that it uses bounding box annotations during the training phase, requiring manual and time-consuming steps to annotate large datasets, such as RSNA.

[Stephen et al. 2019] also proposed a deep convolutional neural network to detect pneumonia. The developed model was subdivided into two major parts: the 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. Some data augmentation techniques as vertical rotation and horizontal zoom were used to to increase the amount of training data. The network hyperparameters also needed some fine-tuning to achieve the accuracy of 93% at the validation data. The main limitation of their work is that the dataset includes only 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

In this section, we present the details about the machine learning classifier trained to detect pneumonia and the architecture of Web Application developed to assist doctors in diagnosis of pneumonia using frontal chest radiographs.

3.1. Machine Learning

The neural network used as the machine learning classifier for the pneumonia detection is based on the CheXNet [Rajpurkar et al. 2017]. CheXNet is a 121-layer Dense Convolutional Network (DenseNet) [Huang et al. 2017] trained on the ChestX-ray 14 dataset. DenseNets improve flow of information and gradients through the network, making the optimization of deep networks tractable.

The last layer of CheXNet works as the output layer of the neural network performing the classification task, while the previous layers can be seen as learned feature extraction layers. In the proposed architecture, we use CheXNet (Figure 1) as a feature extractor, replacing the final fully connected layer by a stack of dropout layers and fully connected layers with Leaky Rectified Linear Unit (Leaky ReLU) and sigmoid activation functions, resulting in a new architecture shown in Figure 2. The dropout layers are used to prevent overfitting during training. The fully connected layers are used to reduce the output size from 14 (CheXNet number of predicted classes) to 1 (probability of pneumonia). Since the training batch size is hardware dependent, we represent it as a question mark (?) in Figure 1 and Figure 2.

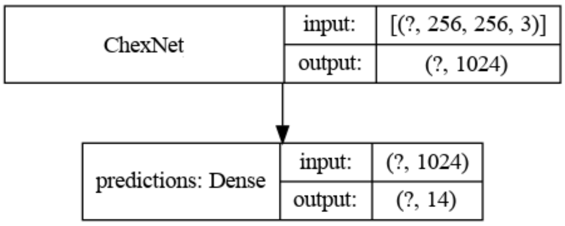

Figure 1. CheXNet architecture.

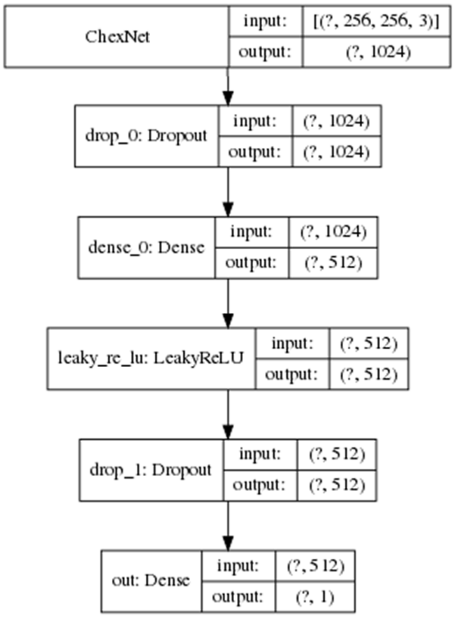

Figure 2. Proposed architecture.

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] as a Kaggle challenge in 2018. We select this dataset instead of the ChestX-ray14 because it only considers the class studied in this paper and has almost 20 times more pneumonia examples in total. This dataset consists of 26,684 labeled examples with a resolution of 1024x1024 pixels in an 8-bit grayscale palette, 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 image samples in the dataset went through a pre-processing phase consisting of conversion, from 8-bit grayscale images to 3-channel RGB images, and resizing, from 1024x1024 pixels to 256x256 pixels. Both pre-processing steps are required to adapt image samples to the format used by the CheXNet input layer.

Furthermore, data augmentation techniques were used to increase the amount of training data, resulting in more than 130.000 image samples. Data augmentation acts as a regularizer and helps reduce overfitting when training a machine learning model.

The data augmentation was performed in 5 possible random iterations per image, as shown in Figure 3. 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 $+10°$;

To evaluate the performance of proposed models we used a 5-fold cross-validation, randomly partitioning the samples into five disjoint sets of equal size, selecting

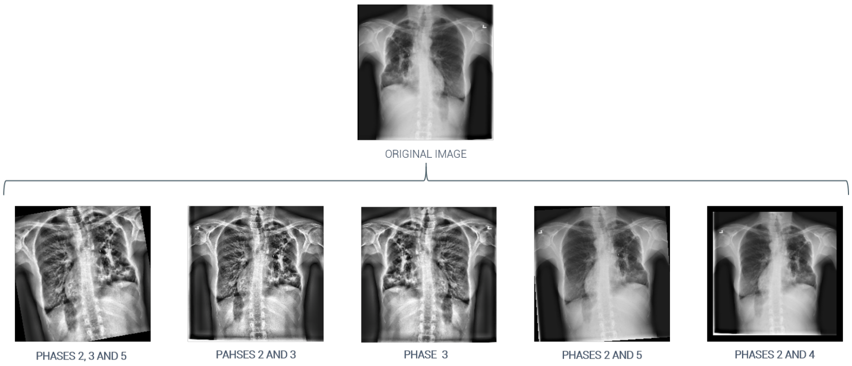

Figure 3. Data augmentation phases exemplified.

one as a testing set and combining the remaining four to form a training set. We conducted five such runs using each partition as the testing set and reported accuracy, loss and area under the receiving operating characteristic curve (AUROC) by fold.

The neural network was trained for 16 epochs per fold, using a batch size of 16 images. The hardware used for training consisted of an Intel Core i7 3930K CPU with 32 GB of RAM and a Nvidia RTX 2070 GPU.

To simplify and facilitate the visualization of the results by the end-user, an algorithm was developed to plot a heatmap that highlights the possible regions of the image 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.

3.2. Web Application

The Web Application developed in this project encompasses two large blocks, as can be seen in the architecture diagram displayed in Figure 4.

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.

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.

All user data (username, password, classifications metadata and annotations) is 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 developed application was embedded on an Nvidia Jetson Nano development

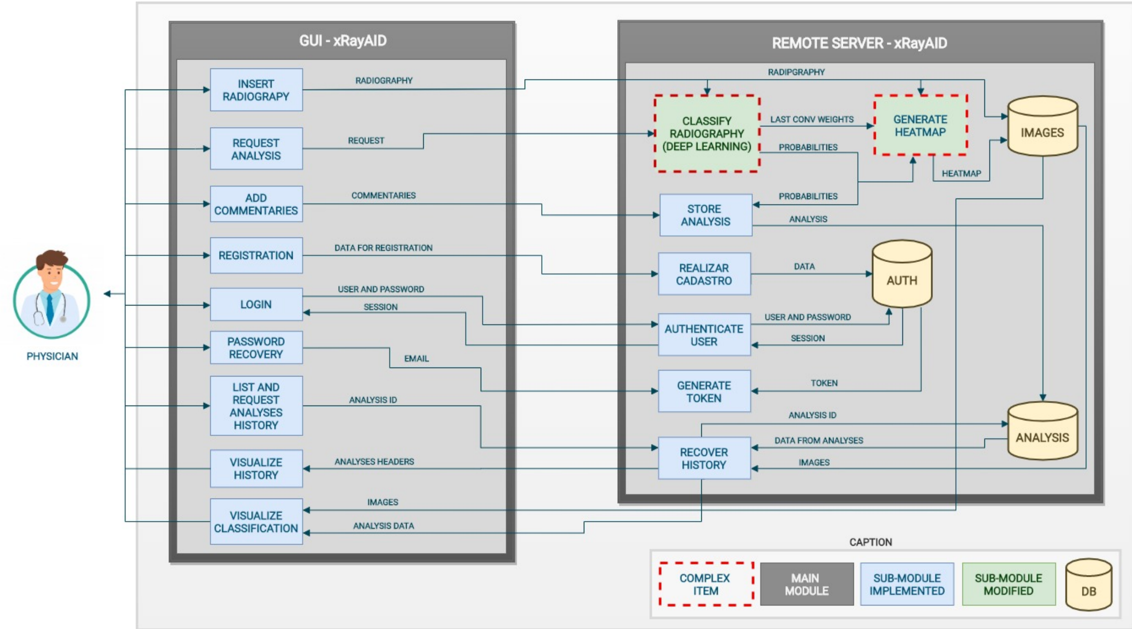

Figure 4. Project Architecture Diagram.

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 with a batch size of only one image. 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

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 AUROC curve in Figure 7, that last one indicating how much the model is capable of distinguishing between radiography with disease and no disease.

The average score of the 5-fold cross-validation was an accuracy of 85.3% ± 2.9%, loss of 33.3% ± 5.9% and AUROC curve of 87.0% ± 0.5%.

Table 2. Model Parameters Statistics.

	Parameters
Total	7.562.817
Trainable	7.479.169
Non-trainable	83.648

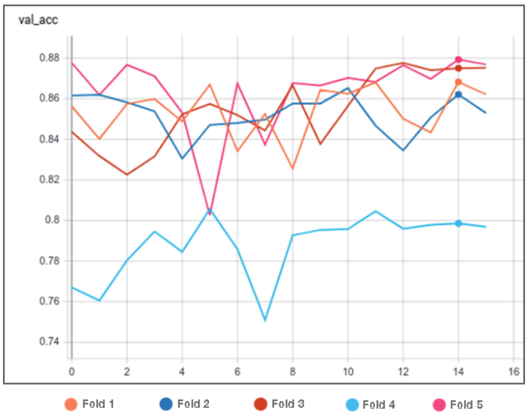

Figure 5. Validation Accuracy.

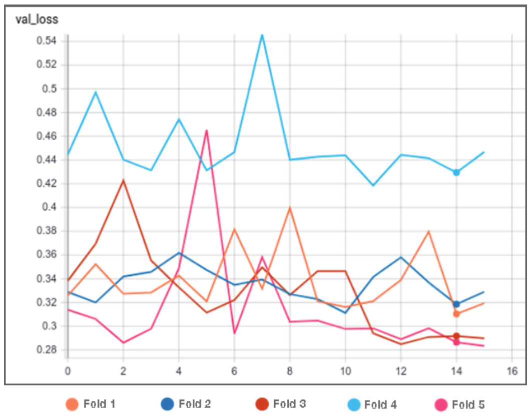

Figure 6. Validation Loss.

The best result was obtained on fold number 5 at the 15th epoch, with a validation accuracy of 87.9%, validation loss of 28.7% and validation AUROC curve of 86.9%. For this reason, this trained model was selected to be used by the Web Application.

Compared to the related works, the proposed system achieved good results as seen in Table 3. Despite the fact that [Han et al. 2021] obtained a better result with the ResNetAttRadi architecture, this approach have the drawback of requiring bounding box annotations, as explained in Section 2 .

Table 3. Comparison With Related Works.

Paper	Dataset	AUROC	Core Network
[Rajpurkar et al. 2017]	ChestX-ray14	0.768	DenseNet
[Han et al. 2021]	RSNA	0.795	ResNet
[Han et al. 2021]	RSNA	0.864	ResNetAtt
[Han et al. 2021]	RSNA	0.848	ResNetRadi
[Han et al. 2021]	RSNA	0.877	ResNetAttRadi
xRayAID	**RSNA**	**0.870**	**DenseNet**

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.

To validate the proposed solution with the end-user, 26 non anonymized radiographs 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.

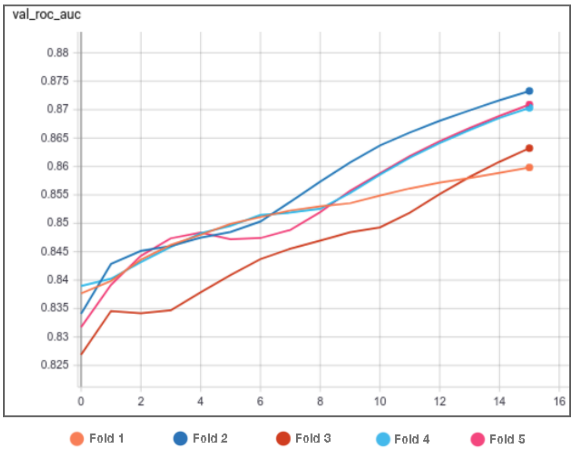

Figure 7. Validation AUROC.

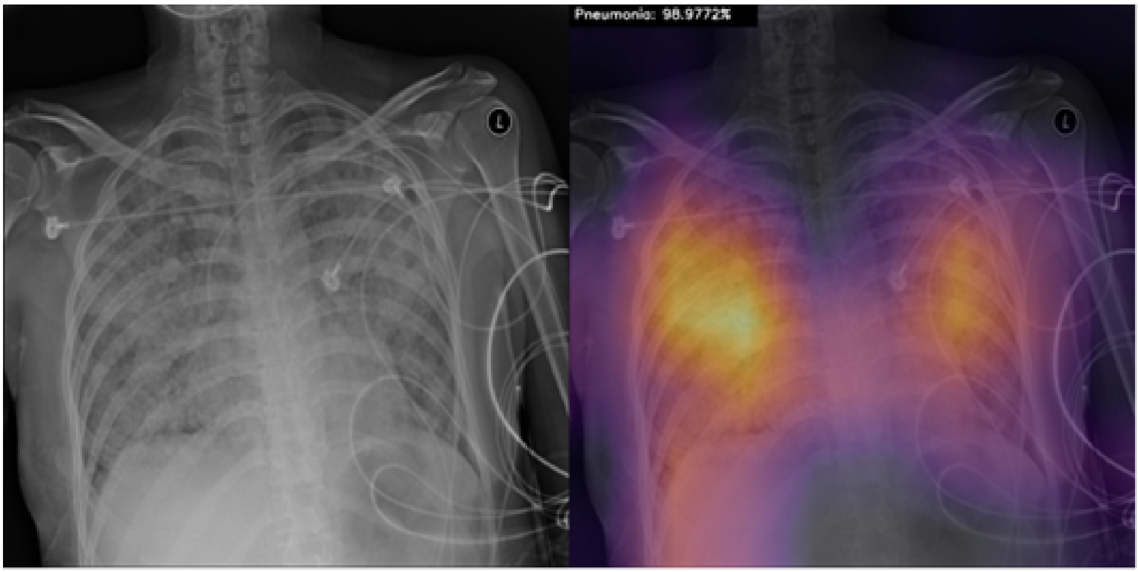

Figure 8. Classification and Heatmap Example.

The project website[1] presents a homepage describing the project, its objectives and functionalities, as seen in Figure 9. That page is responsible for user authentication, giving access to the system after previous registration. Also on that page, the administrators can access a webmail platform used to communicate with users.

The authentication uses *bcrypt* algorithm, since it is considered one of the most secure password-hashing functions, and is protected by Google's *Captcha*. 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 that page, there is a lateral menu bar responsible for redirecting the user to the desired functionality. There are current 3 options: homepage, new classification and history pages. That last one is responsible for showing all the records already made at the system by the user.

[1]https://xrayaid.com.br

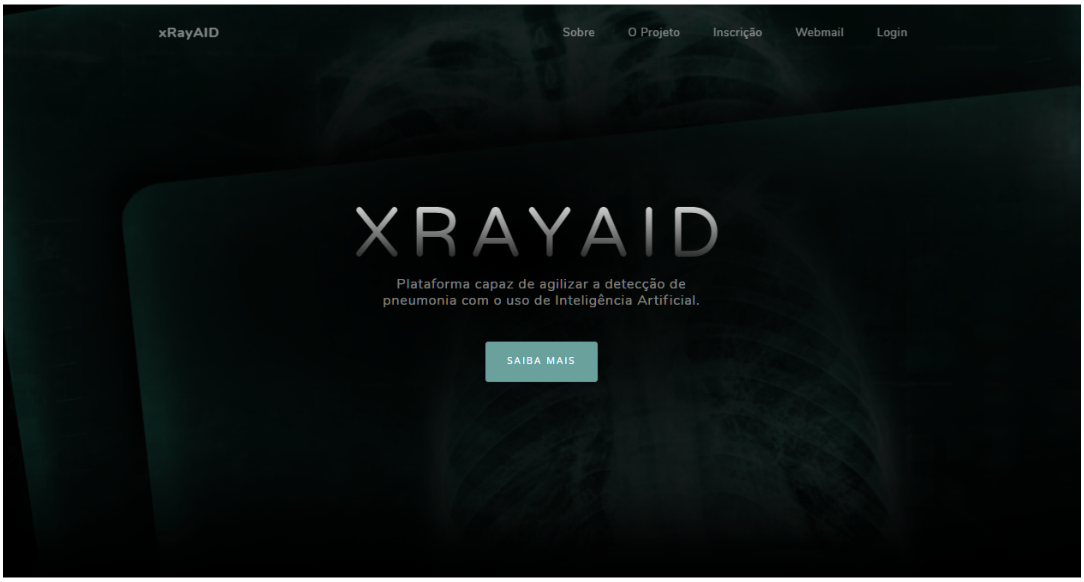

Figure 9. Website Homepage.

By selecting the option to perform a new classification, a new page is displayed. The system support images in PNG or JPEG formats only, at a maximum of 10 megabytes per image, without resolution limits. At that same page, after the server finishes its processing, it automatically reloads the page displaying the results view, as shown in Figure 10.

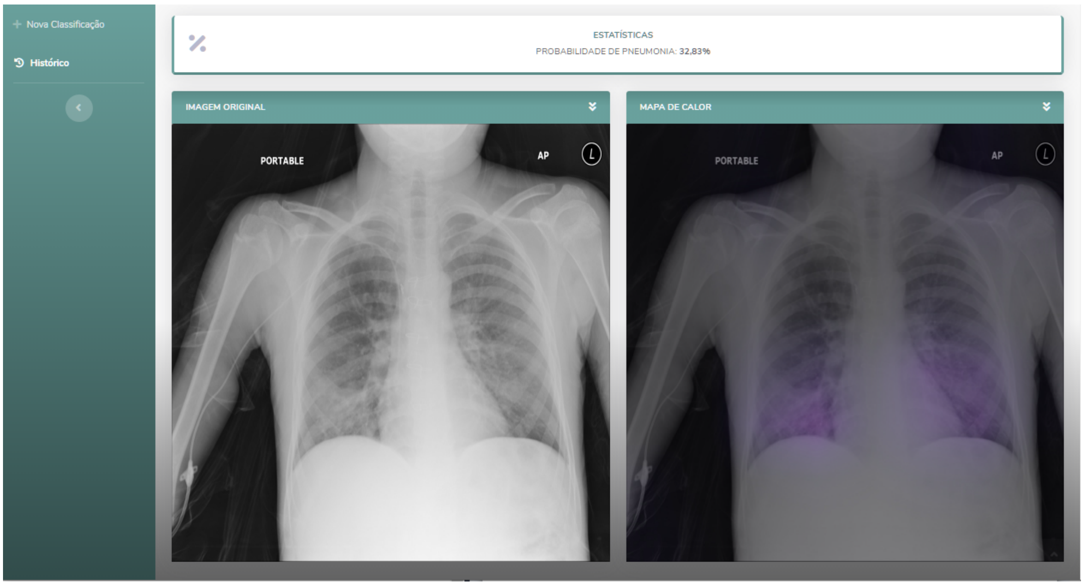

Figure 10. Results View.

On the view page, it is possible to visualize the pneumonia probability as a percentage at the top. 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 11. The user can also write relevant annotations and provide us some feedback about the produced results, as demonstrated in Figure 12.

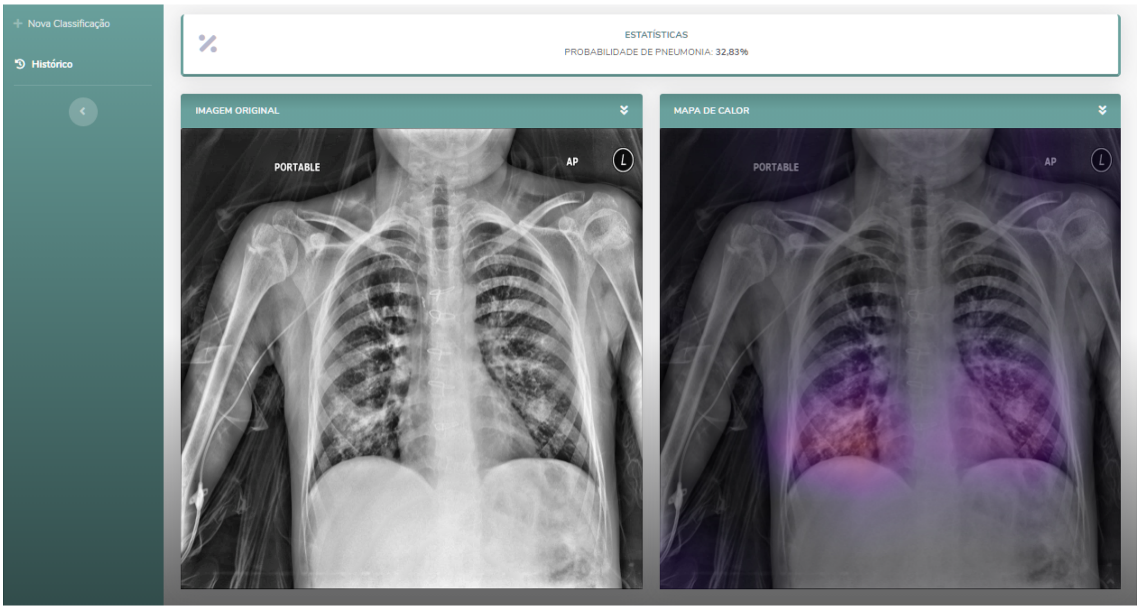

Figure 11. Results View with Filters.

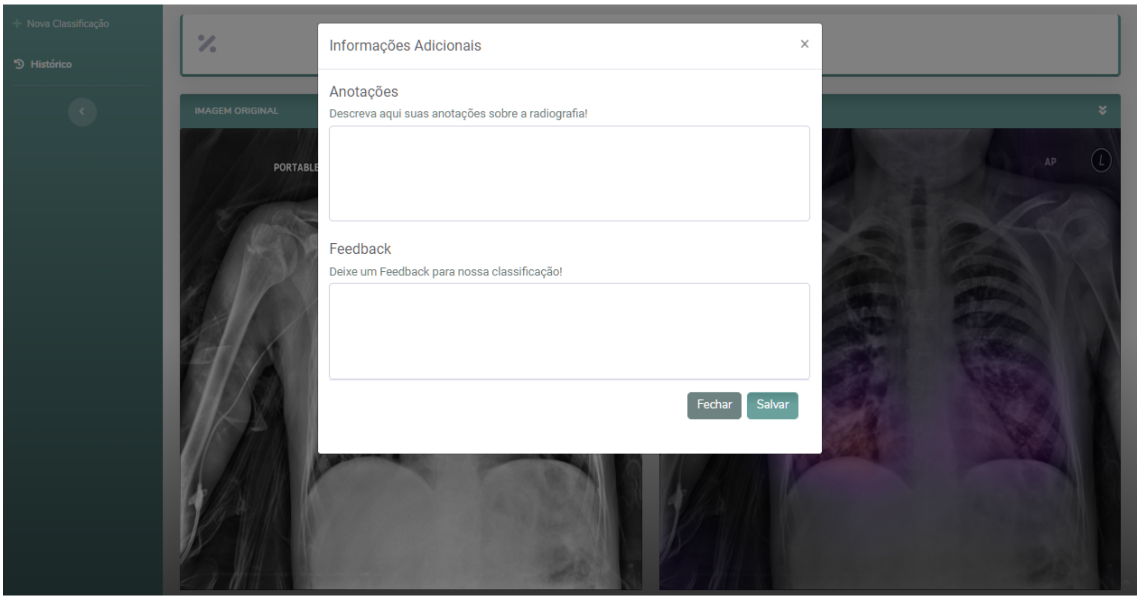

Figure 12. Annotations Form.

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 using keywords, as seen in Figure 13. To visualize the desired classification it is just necessary to click on the corresponding green button "Visualize". After that, the request is processed and displayed as the same view page described in Figure 10. The history page also allows undesired classifications deletion via the trashcan icon.

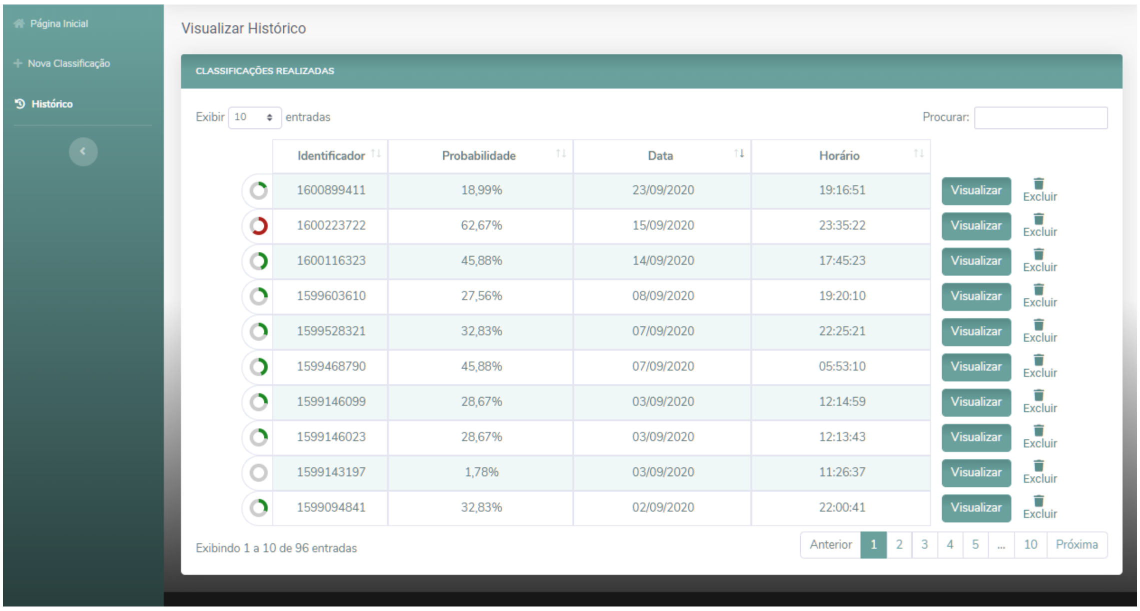

Figure 13. History Page.

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.

Regarding further works on the system, the dataset used to train the machine learning model could be improved including more kinds of radiography, since different X-ray machines produce different results. Moreover, the annotations and feedbacks provided by the users can also be used to fine-tune the trained model, improving the heatmap localization overlay and the classification AUROC.

For the Web Application, some user interface (UI) modifications can be done, especially at userpage and history listing, adding more functionalities to enhance the tool. It would be useful to implement functionalities like remember-me and user-info.

Finally, as we are living in a pandemic world, the proposed network could be trained with a COVID-19 dataset as [Tsai et al. 2021] to evaluate its performance in chest X-ray images.

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.

Han, Y., Chen, C., Tewfik, A. H., Ding, Y., and Peng, Y. (2021). Pneumonia detection on chest x-ray using radiomic features and contrastive learning. *arXiv preprint arXiv:2101.04269*.

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.

Tsai, E. B., Simpson, S., Lungren, M., Hershman, M., Roshkovan, L., Colak, E., Erickson, B. J., Shih, G., Stein, A., Kalpathy-Cramer, J., et al. (2021). The rsna international covid-19 open annotated radiology database (ricord). *Radiology*, page 203957.