ISSN 2007-9737

PumaMedNet-CXR: An Explainable Generative Artificial Intelligence for the Analysis and Classification of Chest X-Ray Images

Carlos Minutti-Martinez[1], Boris Escalante-Ramírez[1,2], Jimena Olveres-Montiel[1,2]

[1] Universidad Nacional Autónoma de México, Centro de Estudios en Computación Avanzada (CECAv-UNAM), Mexico

[2] Universidad Nacional Autónoma de México, Laboratorio Avanzado de Procesamiento de Imágenes (LaPI-UNAM), Mexico

carlos.minutti@iimas.unam.mx, {boris, jolveres}@cecav.unam.mx

Abstract. In this paper, we introduce PumaMedNet-CXR, a generative AI designed for medical image classification, with a specific emphasis on Chest X-ray (CXR) images. The model effectively corrects common defects in CXR images, offers improved explainability, enabling a deeper understanding of its decision-making process. By analyzing its latent space, we can identify and mitigate biases, ensuring a more reliable and transparent model. Notably, PumaMedNet-CXR achieves comparable performance to larger pre-trained models through transfer learning, making it a promising tool for medical image analysis. The model's highly efficient autoencoder-based architecture, along with its explainability and bias mitigation capabilities, contribute to its significant potential in advancing medical image understanding and analysis.

Keywords. Medical image analysis, autoencoder, explainable artificial intelligence, chest X-Ray.

1 Introduction

Medical image understanding is predominantly carried out by skilled medical professionals. However, the limited availability of human experts and the drawbacks of fatigue and imprecise estimation associated with manual analysis limit the effectiveness of medical image interpretation.

Convolutional Neural Networks (CNNs) have emerged as powerful tools for image understanding and have demonstrated superior performance to human experts in various image-related task [23].

Deep Learning, specifically CNNs, has shown significant advancements in object recognition, image analysis, and classification tasks. In the medical field, CNNs have found successful applications.

However, training CNNs requires a substantial amount of data and computational resources, and gathering medical image data presents significant challenges, both in terms of cost and time.

Transfer Learning (TL) addresses this challenge by fine-tuning pre-trained CNNs from large datasets like ImageNet, reducing the need for extensive medical data. Nevertheless, TL has its limitations due to differences between objects in datasets like ImageNet and medical images, such as varying shapes and image characteristics.

Furthermore, pre-trained CNNs from ImageNet come with millions of parameters, posing computational challenges, whereas a medical imaging dataset could potentially be classified more efficiently with a model pre-trained on data similar to medical images.

Additionally, large CNN models lack explainability, a crucial feature for reliable medical image analysis to ensure unbiased results. Moreover, these large models may not be practical in resource-limited areas, where financial,

ISSN 2007-9737

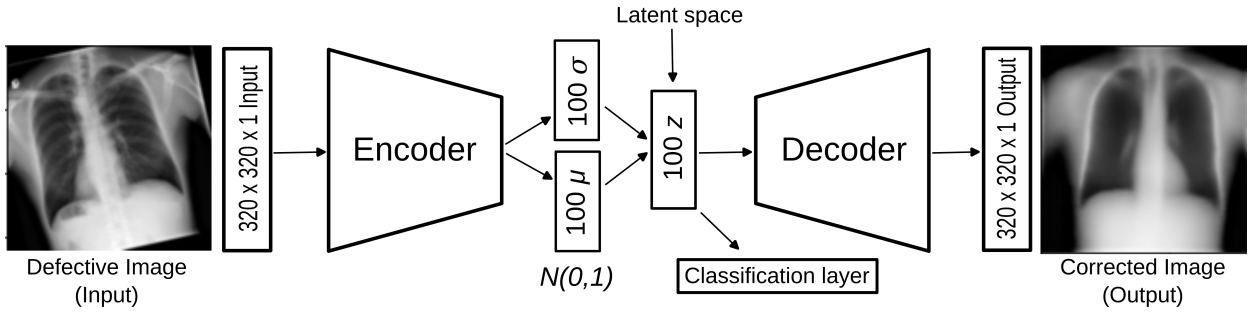

Fig. 1. Schematic diagram of the PumaMedNet-CXR model

technological, or human resources are scarce but could benefit from this technology.

For instance, ranking patients who require urgent attention could be made more accessible with smaller, more explainable models. Recent surveys underscore the significance of CNNs in medical imaging.

Suganyadevi et al. [27] review 120 medical imaging research papers with the ResNet architecture standing out for its high performance. It is also mentioned how challenges remain, such as the scarcity of properly annotated data, limited medical imaging datasets compared to general computer vision datasets, and the considerable expenses associated with teaching deep learning models, often requiring high-end GPUs.

The use of black-box models is also a major obstacle due to legal ramifications, leading to healthcare professionals' reluctance to rely on them.

Sarvamangala and Raghavendra [23] survey CNNs applications in medical image understanding of some diseases of the brain, breast, lung, colon, skin, eyes, heart and other organs, being classification and segmentation the main tasks performed.

The authors mention how CNNs are highly efficient methods of feature extraction, but black-boxes with the need of research in terms of analyzing and understanding output at every layer.

In the context of addressing the lack of large, high-quality labeled datasets, Semi-Supervised (SSL) or Unsupervised Learning (USL) methods have been explored. Solatidehkordi and Zualkernan [26] present a survey of the latest SSL methods proposed for medical image classification tasks, where Virtual Adversarial Training (VAT) is one of the must successful methods, but it keeps having the explainability problem.

Autoencoders, a type of neural network architecture, play a crucial role in USL, serving for dimensionality reduction, feature extraction, and data compression. Comprising an encoder and decoder, autoencoders map input data into a compressed representation (latent space) and then reconstruct the original or variant data from the compressed representation.

This architecture finds applications in image denoising, compression, anomaly detection (e.g., [6]), and can be a base for more complex models like Variational Autoencoders (VAEs) to generate new data samples with specific characteristics. The latent space can also be used for classification tasks, leading to supervised or semi-supervised models.

There are multiple autoencoders architectures and applications (see [5]). Some of these architectures have proven their effectiveness in various medical imaging tasks.

For example, Huang et al. [13] proposed an active learning framework called variational deep embedding-based active learning (VaDEAL) that uses a VAE with sampling strategies to improve the accuracy of diagnosing pneumonia and utilizes the latent space for classification.

Another study by Raghavendra et al. [13] employed a VAE for data imputation on Chest X-Ray (CXR) images, treating high opacity regions as missing data for lung area segmentation using a U-net (see [21]) type segmentation.

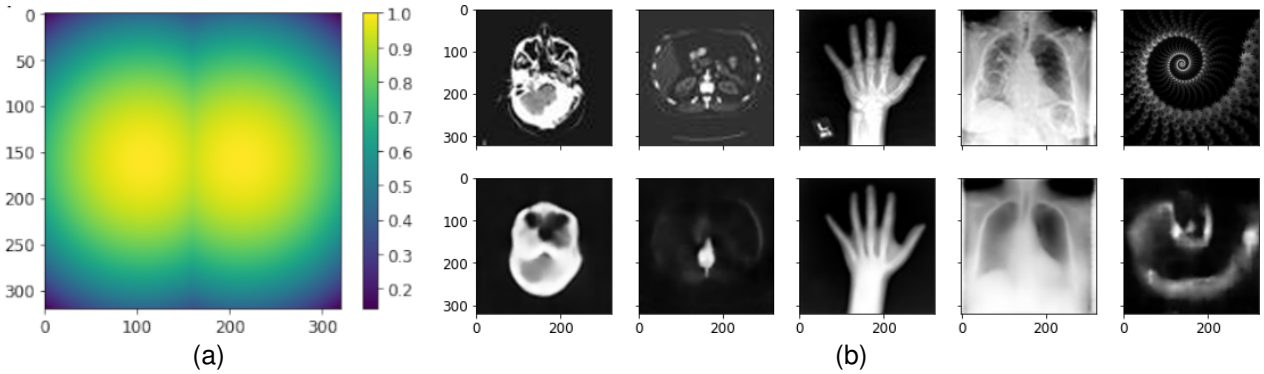

Fig. 2. (a) Weighted mask used for the loss function. Higher weights are assigned to the area of the lungs to prioritize its reconstruction. (b) Samples of original input and reconstruction for the pre-training dataset

Although CXR images are commonly available in medical datasets, their analysis has gained significant attention with the onset of COVID-19 (e.g., [3, 28, 16, 15, 19]).

Many CNN-based works for classifying the disease rely heavily on large CNN models and TL. Some of these approaches address the explainability problem by using Grad-CAM (see [24]) to detect relevant areas in the model's decision-making process and lung segmentation to mitigate biases.

However, Grad-CAM may not provide a comprehensive understanding of the model's internal workings.

The visualization is limited to highlighting areas but does not provide a explanation of how the model arrived at a particular decision, and lung segmentation may not be sufficient on its own to completely avoid biases, as critical features for decision-making could exist outside or even within the segmented lung areas, such as medical devices like pacemakers, catheters, or tubes (see [17]).

Moreover, pre-trained Large-CNN models still face computational burdens for training and prediction.

In this paper, we present the advancements of the PumaMedNet project, which aims to design a CNN architecture for medical image classification with low computational costs for transfer learning, achieving comparable accuracy to current standards while maintaining high explainability and bias detection and mitigation.

Our initial release focuses on CXR images, utilizing a denoising β-VAE as the model's backbone.

The model is trained and validated on the ChestX-ray14 medical imaging dataset, comprising 112,120 frontal-view X-ray images of 30,805 unique patients with fourteen common disease labels, obtained through NLP techniques from radiological reports.

Further validation involves transfer learning on a composite dataset of 19,362 CXR images, including COVID-19 cases not present in the ChestX-ray14 dataset.

Our results demonstrate comparable performance with pre-trained Large-CNN models like ResNet- 18 while enhancing bias mitigation and explainability by exploring the effects of variables in the latent space.

2 Methodology

The methodology of the project was divided into several stages to develop the CNN architecture based on an Autoencoder for medical image classification. The following are the key stages:

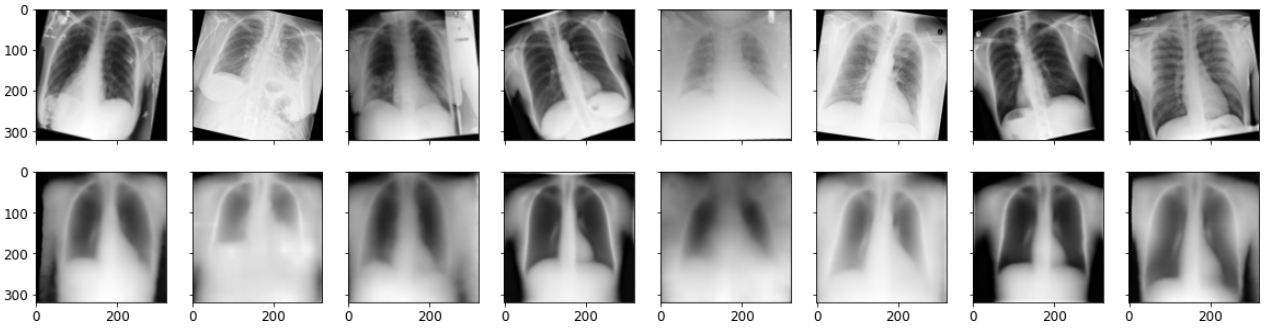

Fig. 3. Sample of input (first row) and reconstructed-corrected images (second row) for the ChestX-ray14 dataset

2.1 Model Architecture

2.1.1 Base Architecture Selection

An Autoencoder was chosen as the base structure to describe the visual characteristics of the images.

The Autoencoder allows generating a vector of latent variables (latent space) that capture essential image information, enabling explainability through the analysis of the latent space, without requiring supervised learning.

2.1.2 Evaluation of Autoencoder Architectures

Several Autoencoder architectures were explored and compared (see [22]). The VAE architecture has a continuous latent space approximation to a normal distribution.

A β-VAE [7] is an extension of the standard VAE that incorporates a hyperparameter called β. The β aims to disentangle and control the learned representations in the latent space by a penalization of the KL-divergence between the latent space and a independent normal distribution, resulting in the following characteristics:

– Disentangled Representations: β-VAEs encourage individual latent variables to capture specific features, facilitating precise control and manipulation of the generated data.

– Explainability: The disentangled representations foster more interpretable latent spaces, simplifying the comprehension and analysis of learned features.

– Bias Mitigation: Through explicit disentanglement of variation factors, β-VAEs offer potential for mitigating biases in generated data and the decision-making process of models by adjusting or deactivating factors contributing to bias.

However, the β hyperparameter introduces a trade-off between reconstruction accuracy and disentanglement, necessitating careful selection of this value.

The β-VAE model was expanded by incorporating a classification layer that employs the latent space for classification tasks.

Additionally, a denoising/corrective component was integrated by training the β-VAE with defective images as input, and measuring the error between the output and the image without added defects.

Figure 1 presents the schematic diagram of the β-VAE model, where an input image which is rotated and flipped horizontally is provided as input, and a corrected image is produced in the output.

2.1.3 Hyperparameter Optimization

We conducted a series of trial-and-error experiments to optimize the model's hyperparameters, encompassing factors like latent space size, layer count, units per layer, and activation functions. The following details highlight the final architectural characteristics:

Table 1. AUC values for different studies on the ChestX-ray14 dataset

				Baltruschat et al.			
Pathology	**Wang et al.**	**Yao et al.**	**Guendel et al.**	**ResNet-38**	**ResNet-50**	**ResNet-101**	**PumaMedNet**
Atelectasis	0.700	0.733	0.767	0.763	0.755	0.747	**0.770**
Cardiomegaly	0.810	0.856	**0.883**	0.875	0.877	0.865	0.863
Consolidation	0.703	0.711	0.745	0.749	0.742	0.734	**0.787**
Edema	0.805	0.806	0.835	0.846	0.842	0.828	**0.874**
Effusion	0.759	0.806	0.828	0.822	0.818	0.818	**0.862**
Emphysema	0.833	0.842	**0.895**	**0.895**	0.875	0.868	0.856
Fibrosis	0.786	0.743	**0.818**	0.816	0.800	0.778	0.771
Hernia	0.872	0.775	0.896	**0.937**	0.916	0.855	0.834
Infiltration	0.661	0.673	0.709	0.694	0.694	0.686	**0.710**
Mass	0.693	0.777	**0.821**	0.820	0.810	0.796	0.770
Nodule	0.669	0.718	**0.758**	0.747	0.736	0.738	0.674
Pleural Thicken.	0.684	0.724	0.761	0.763	0.742	0.739	**0.783**
Pneumonia	0.658	0.684	**0.731**	0.714	0.703	0.694	0.702
Pneumothorax	0.799	0.805	0.846	0.840	0.819	0.839	**0.861**
Average	0.745	0.761	**0.807**	0.806	0.795	0.785	0.794
No Findings	—	—	—	0.727	0.725	0.720	**0.754**

– Latent Space Size: The latent space comprises one hundred variables. Although the experiments indicated a feasible size of fifty variables, it was considered beneficial to opt for a larger latent space, useful for when transfer learning is performed across numerous classes.

– Layers: Our architecture employs six layers for the encoding algorithm and another six for decoding, utilizing ConvTranspose2d for deconvolution. Extending the number of layers for the encoder and decoder did not demonstrate enhancements in classification or reconstruction tasks. Excessively deep layers were intentionally avoided to preserve efficiency.

– Activation Functions: A range of activation functions, including ReLU, GeLU, ELU, LeakyReLU, SiLU, and the novel Smish [31], were evaluated. LeakyReLU(0.15) emerged as the most effective choice.

– Batch Normalization: Incorporating batch normalization into each encoder and decoder layer did not yield improvements due to the architecture's limited layer count. Hence, the final model omits batch normalization.

– Dropout: The dropout function introduces redundancy in the latent variables, which is an undesirable feature in the proposed model, potentially compromising explainability, so it was excluded.

– Skip-Connections: While skip-connections were explored to enhance image reconstruction, their introduction consistently affected latent space sensitivity. This reduction in explainability contradicted the model's objectives, leading to their exclusion.

– Classification Layer: This layer comprises two fully connected layers from the latent space to the classes, utilizing ReLU activation. Increasing the layer count resulted in higher classification errors.

This architecture results in a total of 1,405,753 trainable parameters, which is less than lightweight, state-of-the-art architectures tailored for mobile devices, such as MobileNetV3 Small [12], with 2,542,856 parameters.

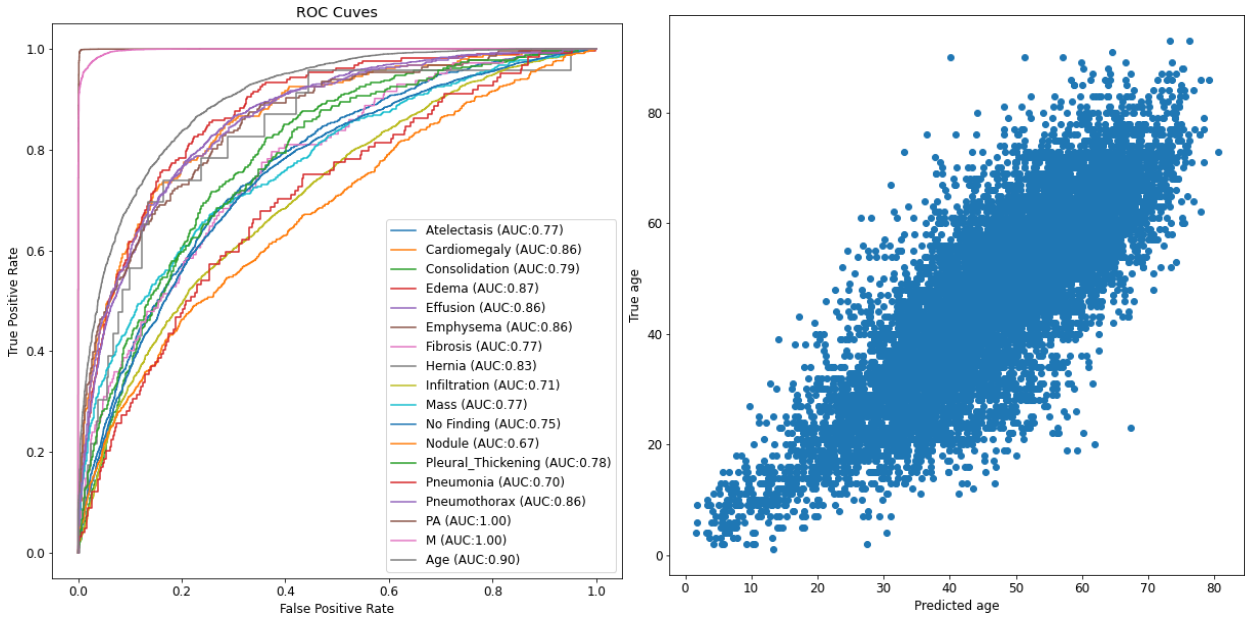

Fig. 4. ROC curves for the 15 classes of the ChestX-ray14 dataset and CXR type (AP or PA), Sex (M or F), Age (above or below the median), and predicted vs True age, reported for the patients in the test dataset

2.1.4 Loss Function Investigation

Structural Similarity Index [32] (SSIM) is an image quality assessment metric that measures the similarity between two images. It quantifies the structural information, luminance, and contrast similarities, making it a useful alternative to Mean Squared Error (MSE) as a loss function in the autoencoder architecture, which only measures pixel-wise differences.

SSIM is designed to mimic human perception of image similarity, making it more aligned with the human visual system's sensitivity to changes in structure and textures.

And its use in Medical Image Analysis as also been studied (see [18]). In addition, Bergmann et al. [6] found that it is more useful for Unsupervised Defect Segmentation, where an autoencoder is trained to reconstruct images, and defected on images can be found by differences between reconstruction and the input image.

These characteristic can be useful for the model, for a zero-shot training, where classification is possible, even for classes which are not part of the training dateset.

In addition, MSE loss can suffer from gradient saturation, especially when the autoencoder produces images that are far from the ground truth. SSIM mitigates this problem by providing a more informative loss signal during training.

2.1.5 Emphasis on Regions of Interest During Training

To enhance sensitivity towards crucial regions in CXR images, like the lungs, we incorporated a weighted mask during the autoencoder training (Figure 2). This strategy enabled the model to concentrate on clinically significant areas, thereby refining its performance.

2.1.6 Pre-Training

Building upon the methodology proposed by Singh et al. [25], who utilized weakly supervised pre-training to enhance image recognition performance, we adopted a similar approach. In our case, the model was pre-trained on three distinct datasets.

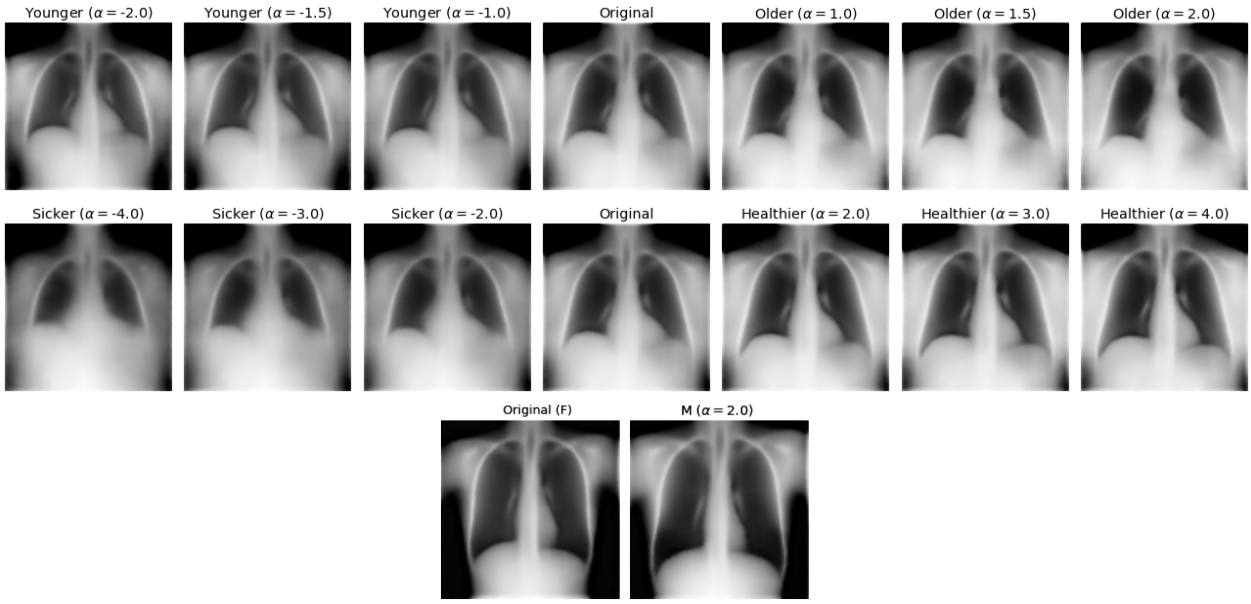

Fig. 5. Decoded images for the modification of the latent space to change the age, health, and sex of the original image. From younger to older (first row), from sicker to healthier (second row), and from female to male (third row)

These include the *Describable Textures Dataset* [8] comprising 5,640 images across 47 classes, the *Textures Classification dataset* [1] containing 8,674 images categorized into 64 classes, and the *Medical MNIST* dataset [20] comprising a substantial collection of 58,954 medical images grouped into 6 classes.

For a visual representation of this pre-training process, refer to Figure 2, which showcases example images and their corresponding reconstructions.

2.1.7 Evaluation and Result Comparison

The ChestX-ray14 dataset comprises 112,120 CXR images, an expansion of the ChestX-ray8 dataset [30], encompassing fourteen common thoracic pathologies: Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural thickening, Cardiomegaly, Nodule, Mass and Hernia.

An additional category labeled “No finding” is also included. This dataset serves as the foundation for training and validating the model.

Furthermore, to provide additional validation, TL is conducted on a composite dataset of three categories: Pneumonia, COVID-19, and Normal. These categories were sourced from various publicly accessible datasets [2, 9, 14, 29].

Duplicate images were identified using the Geeqie software [10], detecting images with a visual similarity exceeding 97% and treating them as identical. This validation dataset comprises a total of 19,362 CXR images, with 1,831 images designated for testing.

Half of these images correspond to the lung segmentation of the dataset, to introduce visual variability similar to that of the original ChestX-ray14 dataset. Comparative results were obtained against a fine-tuned pre-trained ResNet-18 model, with 11,689,512 parameters, making it 8.3 times larger than our model.

3 Results

A sample of defective inputs and the corresponding autoencoder outputs, correcting rotation and flipped images, is presented in Figure 3.

Additionally, denoising characteristics were included in the model through Gaussian Blur, Random Equalize, and Random Autocontrast applied to the input images, to be corrected at the output.

Table 1 displays a comparative analysis of various studies involving classification using the ChestX-ray14 dataset. Wang et al. [30] examined different CNN architectures (AlexNet, GoogLeNet, VGGNet-16, ResNet-50), with ResNet-50 achieving the best results. Yao *el al.* [33] employed a custom architecture, while Gundel *et .al.* [11] utilized an approach based on DenseNet121. Baltruschat *et .al.* [4] experimented with different ResNet architectures and achieved results similar to each other.

From the results, it is evident that the smallest architecture with similar performance to PumaMedNet is Baltruschat's ResNet-38 *et .al.*, which has at least 16 times as many parameters as PumaMedNet, resulting in PumaMedNet achieving better performance when considering the computational burden.

Figure 4 shows the ROC curves and AUC values for the 15 classes (14 diseases and a "No finding" category) of the ChestX-ray14 dataset. Additionally, ROC-AUC is displayed for CXR type (AP, PA), SEX (M, F), and AGE (above or below the median), which are metadata included in the dataset. The results demonstrate that the model effectively separates CXR type and sex classes and accurately predicts age.

3.1 Latent Space Interpolation

The latent space generated by the model can be used to simulate and explore how the model "understands" specific characteristics. For example, By studying the average values of the latent space for the "No finding" class versus other health conditions, it is possible to modify any image to increase or decrease its health value.

The modification is achieved through latent space manipulation, by doing $z_i^* = z_i + \alpha(z_1 - z_0)$, where z_i^* is the modified latent space of the image z_i, z_1 is the average value for the latent space for the class "No finding", and z_1 the average value for any other class.

Table 2. Classification results using PumaMedNet-CXR and ResNet-18

PumaMedNet-CXR				
	precision	**recall**	**f1-score**	**support**
COVID19	0.993	0.994	0.994	1224
NORMAL	0.919	0.951	0.935	634
PNEUMONIA	0.981	0.969	0.975	1804
ResNet-18				
	precision	**recall**	**f1-score**	**support**
COVID19	0.999	0.998	0.999	1224
NORMAL	0.911	0.951	0.931	634
PNEUMONIA	0.983	0.968	0.976	1804

Larger positive α values increase health, whereas larger negative values decrease health.

Figure 5 presents examples of health, age, and sex modifications for some images. Younger versions of the image display a more rounded thorax and better contrast compared to the versions of older patients.

Health modification mainly affects lung opacity, being higher for versions of a sicker patient. Changing from female to male results in increased thorax and heart size, as well as shoulders, while the basic structure of the lungs remains the same.

3.2 Transfer Learning

Fine-tuning the PumaMedNet-CXR model and a ResNet-18 model (which has 8.3 times more parameters), yielded very similar performance metrics, as shown in Table 2.

3.3 Explainability

Although ResNet-18 performed similar in the classification task than PumaMedNet-CXR, our model allows for a better understanding of the decision-making process done by the model.

Although ResNet-18 performed similarly to PumaMedNet-CXR in the classification task, our model provides better explainability of the decision-making process.

Figure 6 illustrates the effect of varying a latent variable that has been found to be crucial for classification.

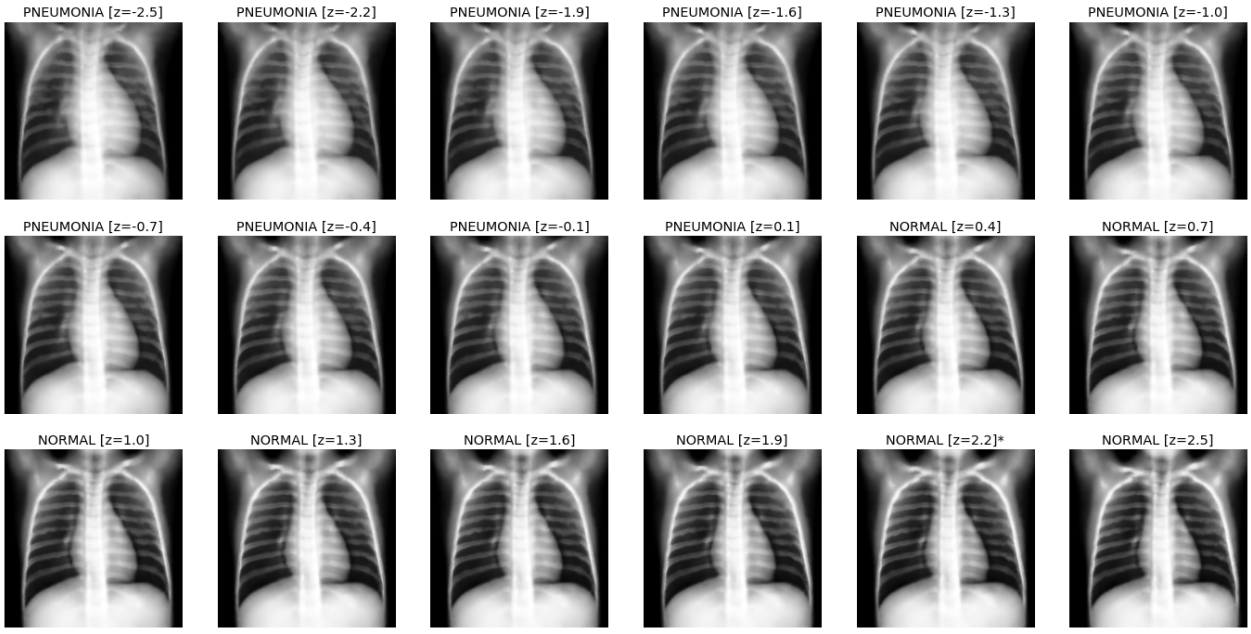

Fig. 6. Model explainability: By varying one of the latent variables most relevant for classification, it can be seen how it changes the size of the heart, resulting in different classifications

By varying its values, it changes the size of the heart, likely related to detecting whether the CXR image is AP or PA, as the AP view results in a heart magnification on the X-ray film, because in the AP view the beam enters from front to back.

Latent variable related to any bias can be ignored in the classification task, or randomly changed. resulting in a model which does not have this bias. Understanding these latent variables allows the avoidance of biases in the classification task without the need for complete model retraining or dataset modification.

4 Summary and Conclusions

In this study, we presented the PumaMedNet-CXR, an autoencoder-based CNN architecture designed for medical image classification, particularly focusing on Chest X-ray (CXR) images.

We demonstrated the effectiveness of the PumaMedNet-CXR in correcting common defects found in CXR images, such as rotation, flipping, and denoising.

The model achieved comparable performance with a ResNet-18 model, despite having significantly fewer parameters, highlighting its efficiency. Furthermore, the explainability offered by the PumaMedNet-CXR allowed us to gain insights into the decision-making process of the model and detect important latent variables relevant for classification.

Through the manipulation of the latent space, we showed how the model can simulate and explore specific characteristics, such as age, health status, and sex.

Additionally, we explored the use of transfer learning to fine-tune the model on a smaller dataset, demonstrating that the PumaMedNet-CXR can achieve similar performance to larger pre-trained models like ResNet-18 while retaining better explainability.

The explainability offered by the model is of great importance in medical image analysis, as it provides transparency in the decision-making process, helps detect potential biases, and enhances the trustworthiness of the model's predictions.

Avoiding biases is crucial in ensuring equitable healthcare outcomes for all patients. Future work will focus on extending this approach to other medical imaging modalities and exploring the model's performance on a broader range of medical conditions, while continuing to prioritize explainability and bias mitigation.

Data Availability Statement

The PumaMedNet-CXR model and weights are openly available[1].

Acknowledgments

This work was supported by the DGAPA's postdoctoral fellowship program (programa de becas posdoctorales) at UNAM and the PAPIIT–DGAPA–UNAM grant IV100420.

References

1. **Abin24 (2023).** Textures classification dataset. GitHub repository.
2. **Agchung (2023).** COVID-19 image dataset. GitHub repository.
3. **Alshmrani, G. M., Ni, Q., Jiang, R., Pervaiz, H., Elshennawy, N. M. (2023).** A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. Alexandria Engineering Journal, Vol. 64, pp. 923–935. DOI: 10.1016/j.aej.2022.10.053.
4. **Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., Saalbach, A. (2019).** Comparison of deep learning approaches for multi-label chest X-ray classification. Scientific Reports, Vol. 9, No. 1, pp. 6381. DOI: 10.1038/s41598-019-42294-8.
5. **Bank, D., Koenigstein, N., Giryes, R. (2021).** Autoencoders. DOI: 10.48550/arXiv.2003.05991.
6. **Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C. (2019).** Improving unsupervised defect segmentation by applying structural similarity to autoencoders. Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Vol. 5, pp. 372–380. DOI: 10.5220/0007364503720380.
7. **Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., Lerchner, A. (2018).** Understanding disentangling in β-vae. 31st Conference on Neural Information Processing Systems. DOI: 10.48550/arXiv.1804.03599.
8. **Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A. (2014).** Describing textures in the wild. IEEE Conference on Computer Vision and Pattern Recognition, pp. 3606–3613. DOI: 10.1109/CVPR.2014.461.
9. **Cohen, J. P., Morrison, P., Dao, L. (2020).** COVID-19 image data collection. DOI: 10.48550/arXiv.2003.11597.
10. **Geeqie (2023).** Geeqie, lightweight image viewer.
11. **Gündel, S., Grbic, S., Georgescu, B., Liu, S., Maier, A., Comaniciu, D. (2019).** Learning to recognize abnormalities in chest X-rays with location-aware dense networks. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, pp. 757–765. DOI: 10.1007/978-3-030-13469-3_88.
12. **Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., Adam, H. (2019).** Searching for MobileNetV3. IEEE/CVF International Conference on Computer Vision, pp. 1314–1324. DOI: 10.1109/ICCV.2019.00140.
13. **Huang, J., Ding, W., Zhang, J., Li, Z., Shu, T., Kuosmanen, P., Zhou, G., Zhou, C., Yu, G. (2022).** Variational deep embedding-based active learning for the diagnosis of pneumonia.

[1] https://github.com/cminuttim/PumaMedNet-CXR

Frontiers in Neurorobotics, Vol. 16. DOI: 10.3389/fnbot.2022.1059739.

14. **Kermany, D., Zhang, K., Goldbaum, M. (2018).** Labeled optical coherence tomography (OCT) and chest X-Ray images for classification. Mendeley Data. DOI: 10.17632/rscbjbr9sj.2.
15. **Kumar, S., Mallik, A. (2023).** COVID-19 detection from chest X-rays using trained output based transfer learning approach. Neural Processing Letters, Vol. 55, No. 3, pp. 2405–2428. DOI: 10.1007/s11063-022-11060-9.
16. **Kwon, H. J., Lee, S. H. (2022).** A two-step learning model for the diagnosis of coronavirus disease-19 based on chest X-ray images with 3D rotational augmentation. Applied Sciences, Vol. 12, No. 17. DOI: 10.3390/app12178668.
17. **Mathew, R. P., Alexander, T., Patel, V., Low, G. (2019).** Chest radiographs of cardiac devices (part 1): Lines, tubes, non-cardiac medical devices and materials. South African Journal of Radiology, Vol. 23, No. 1. DOI: 10.4102/sajr.v23i1.1729.
18. **Mudeng, V., Kim, M., Choe, S. (2022).** Prospects of structural similarity index for medical image analysis. Applied Sciences, Vol. 12, No. 8. DOI: 10.3390/app12083754.
19. **Nillmani, Sharma, N., Saba, L., Khanna, N. N., Kalra, M. K., Fouda, M. M., Suri, J. S. (2022).** Segmentation-based classification deep learning model embedded with explainable AI for COVID-19 detection in chest X-ray scans. Diagnostics, Vol. 12, No. 9. DOI: 10.3390/diagnostics12092132.
20. **NVIDIA (2017).** Medical image classification using the MedNIST dataset. GitHub repository, Deep Learning Institute.
21. **Ronneberger, O., Fischer, P., Brox, T. (2015).** U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
22. **Roth, K., Ibrahim, M., Akata, Z., Vincent, P., Bouchacourt, D. (2023).** Disentanglement of correlated factors via Hausdorff factorized support. International Conference on Learning Representations (ICLR). DOI: 10.48550/arXiv.2210.07347.
23. **Sarvamangala, D. R., Kulkarni, R. V. (2022).** Convolutional neural networks in medical image understanding: A survey. Evolutionary Intelligence, Vol. 15, No. 1, pp. 1–22. DOI: 10.1007/s12065-020-00540-3.
24. **Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017).** Grad-CAM: Visual explanations from deep networks via gradient-based localization. IEEE International Conference on Computer Vision (ICCV), pp. 618–626. DOI: 10.1109/ICCV.2017.74.
25. **Singh, M., Gustafson, L., Adcock, A., de Freitas-Reis, V., Gedik, B., Kosaraju, R. P., Mahajan, D., Girshick, R., Dollár, P., Maaten, L. (2022).** Revisiting weakly supervised pre-training of visual perception models. IEEE/CVF Conference on Computer Vision and Pattern Recognition. DOI: 10.1109/CVPR52688.2022.00088.
26. **Solatidehkordi, Z., Zualkernan, I. (2022).** Survey on recent trends in medical image classification using semi-supervised learning. Applied Sciences, Vol. 12, No. 23. DOI: 10.3390/app122312094.
27. **Suganyadevi, S., Seethalakshmi, V., Balasamy, K. (2022).** A review on deep learning in medical image analysis. International Journal of Multimedia Information Retrieval, Vol. 11, No. 1, pp. 19–38. DOI: 10.1007/s13735-021-00218-1.
28. **Sultana, A., Nahiduzzaman, M., Bakchy, S. C., Shahriar, S. M., Peyal, H. I., Chowdhury, M. E., Khandakar, A., Arselene-Ayari, M., Ahsan, M., Haider, J. (2023).** A real time method for distinguishing COVID-19 utilizing 2D-CNN and transfer learning. Sensors, Vol. 23, No. 9. DOI: 10.3390/s23094458.

29. **Vayá, M., Saborit-Torres, J. M., Montell-Serrano, J. A., Oliver-Garcia, E., Pertusa, A., Bustos, A., Cazorla, M., Galant, J., Barber, X., Orozco-Beltrán, D., García-García, F., Caparrós, M., González, G., Salinas, J. M. (2021).** BIMCV COVID-19+: A large annotated dataset of RX and CT images from COVID-19 patients. IEEE DataPort. DOI: 10.21227/w3aw-rv39.

30. **Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R. M. (2017).** ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3462–3471. DOI: 10.1109/cvpr.2017.369.

31. **Wang, X., Ren, H., Wang, A. (2022).** Smish: A novel activation function for deep learning methods. Electronics, Vol. 11, No. 4. DOI: 10.3390/electronics11040540.

32. **Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004).** Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600–612. DOI: 10.1109/TIP.2003.819861.

33. **Yao, L., Prosky, J., Poblenz, E., Covington, B., Lyman, K. (2018).** Weakly supervised medical diagnosis and localization from multiple resolutions. DOI: 10.13140/RG.2.2.30419.68645.

Article received on 11/06/2023; accepted on 15/09/2023.
Corresponding author is Carlos Minutti-Martinez.