Preliminary Study on the Impact of Attention Mechanisms for Medical Image Classification

Tiago Gonçalves (FEUP/INESC TEC), Jaime S. Cardoso (FEUP/INESC TEC)

Introduction

The democratised access to data and the increase of the availability of computational power allowed deep learning methodologies to achieve nearly-human performances in several areas of science, business and government.

Given the high predictive performance rates of convolutional neural networks in computer vision tasks (e.g., natural image recognition), the application of deep learning algorithms in medical image classification occurred almost naturally.

Explainable Artificial Intelligence

The complexity and black-box behaviour of deep learning algorithms motivated the development of the topic of explainable artificial intelligence.

This framework can be seen as a three-stage process [1]: pre-model, in-model and post-model.

In healthcare applications, it is fundamental to assess the quality of these explanations for the sake of transparency, ethics and fairness [2].

Attention Mechanisms

According to psychology, humans tend to selectively concentrate on a specific part of the information. For instance, in artificial intelligence systems, some parts of the inputs may be more relevant than others [3]. This rationale motivated the development of attention mechanisms for deep neural networks.

Several attention mechanisms were proposed for biomedical applications [4, 5], and their impact on the interpretability is assessed through an analysis of the saliency maps (e.g., Grad-CAM [6]).

Implementation and Results

We adapted the attention mechanism described in [4] (MLDAM) for three backbones (VGG-16 [7], ResNet-50 [8] and DenseNet-121 [9]) and proposed four use-cases: baseline, baseline with data augmentation, baseline and MLDAM, baseline and MLDAM with data augmentation.

We computed the accuracy, precision, recall and F1-score for the test set of each database and generated saliency maps [10] for the positive and negative samples of the test set that were correctly predicted by all the use-cases.

Code and supplementary results are publicly available in: https://github.com/TiagoFilipeSousaGoncalves/attention-mechanisms-healthcare

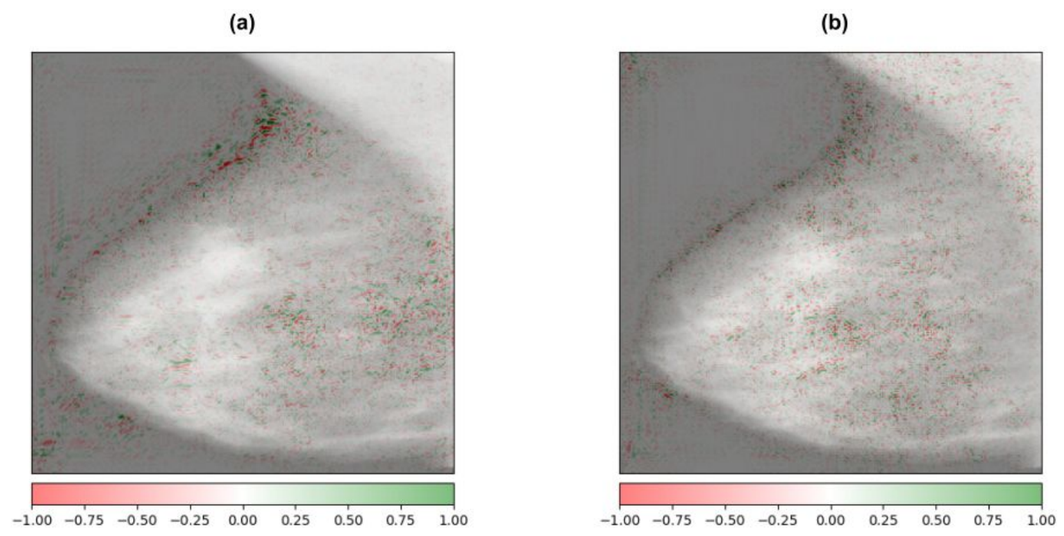

Figure 1 - Examples of saliency maps obtained for an image with label “0” of the CBIS-DDSM data set, using the DenseNet-121 backbone model: (a) - Baseline, (b) - Baseline and MLDAM.

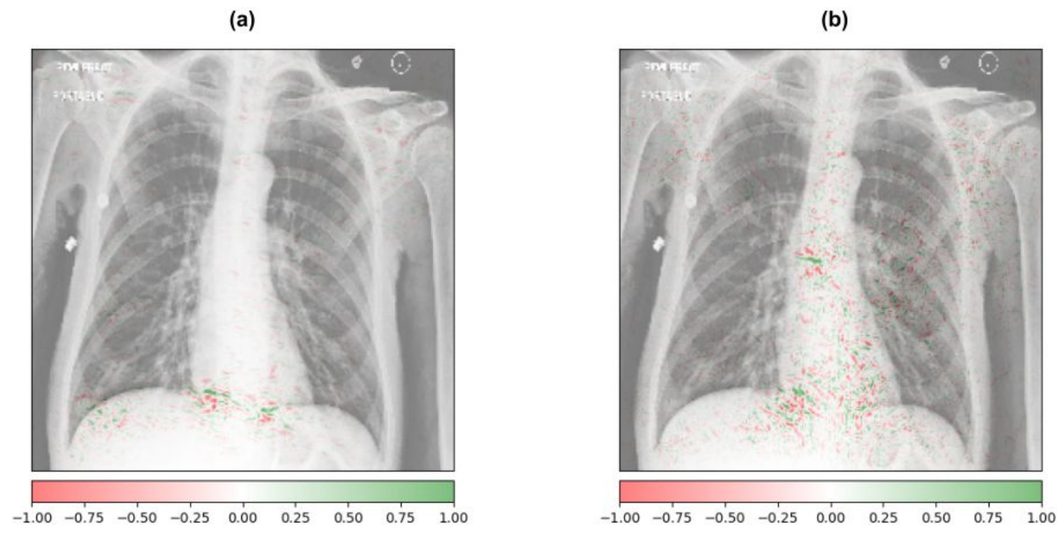

Figure 2 - Examples of saliency maps obtained for an image with label “0” of the MIMIC-CXR data set, using the DenseNet-121 backbone model: (a) - Baseline, (b) - Baseline and MLDAM.

Table 1 - Accuracy results obtained for the test set of the MIMIC-CXR data set: (a) - Baseline, (b) - Baseline with Data Augmentation, (c) - Baseline and MLDAM, (d) - Baseline and MLDAM with Data Augmentation.

Model	Weights	(a)	(b)	(c)	(d)
DenseNet-121	Training	0.8451	0.8312	0.8349	0.8386
	Validation	0.8629	0.8498	0.8535	0.8666
ResNet-50	Training	0.8340	0.8424	0.8470	0.8386
	Validation	0.8535	0.8694	0.8563	0.8414
VGG-16	Training	0.8507	0.8330	0.8293	0.8461
	Validation	0.8629	0.8731	0.8535	0.8647

Conclusions & Future Work

Our experiments did not present conclusive results on the impact of attention mechanisms in two healthcare use-cases, using three different state-of-the-art backbones.

Further work should be devoted to:

1) Development of new experiences with different data processing and augmentation strategies;
2) Design of different attention mechanisms that capture features from different scales or levels;
3) Generation of saliency maps with other methods to see if the results are dependent on the post-model interpretability method;
4) Experimentation of different state-of-the-art backbones to see if performances differ;
5) Evaluation of different data sets and different tasks to assess if results are data or task-dependent.

Acknowledgements

This work was partially funded by the Project TAMI - Transparent Artificial Medical Intelligence (NORTE-01-0247-FEDER-045905) financed by ERDF - European Regional Fund through the North Portugal Regional Operational Program - NORTE 2020 and by the Portuguese Foundation for Science and Technology - FCT under the CMU - Portugal International Partnership and the PhD grant ``2020.06434.BD''.

References

[1] Finale Doshi-Velez and Been Kim. Towards A Rigorous Science of Interpretable Machine Learning.

[2] Cynthia Rudin. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.

[3] Bahdanau et al. Neural machine translation by jointly learning to align and translate.

[4] Mishra et al. Multi-level dual-attention based cnn for macular optical coherence tomography classification.

[5] Cunjian Chen and Arun Ross. An explainable attention-guided iris presentation attack detector.

[6] Selvaraju et al. Grad-cam:Visual explanations from deep networks via gradient-based localization.

[7] He et al. Deep residual learning for image recognition.

[8] Huang et al. Densely connected convolutional networks.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition.

[10] Simonyan et al. Deep inside convolutional networks: Visualising image classification models and saliency maps.