



DEVELOPMENT OF ROBUST AI BASED X-RAY ANALYSIS WORKFLOW BASED ON REAL-TIME OUT OF DISTRIBUTION DETECTION

Lóránt SZABÓ,¹ Zoltán WELTSCH²

¹ AI Research Center (John von Neumann University), Kecskemét, Hungary, szabo.lorant@nje.hu

² Széchenyi István University, Győr, Hungary. weltsch.zoltan@sze.hu

Abstract

In medicine, the role of artificial intelligence (AI) is becoming increasingly common. An important example here is the application of AI in X-Ray analysis, as a known aspect of medical imaging and finding-detection. However, the effectiveness of AI image analysis may be challenging due to the out-of-distribution (OOD) records, i.e. data that significantly differ from the data set used to train the model. These OOD data may result from symptoms, that the model is not prepared for, or even from unpredictable tool behaviour, environmental changes or new errors that have not occurred during the data-gathering phase. This paper shows that with proper OOD analysis the AI-based tool may be prepared for handling “unknown” input data.

Keywords: AI, OOD, ID, t-SNE.

1. Introduction

1.1. AI in medical imaging

More and more hospitals are starting to use different kinds of artificial intelligence (AI) within various departments. However, one of the biggest remaining challenges of using AI in these domains is how the proper performance of AI can be assured. In order to set up and support a powerful collaboration between the MDs and AI, a workflow should be defined, where AI is used for annotating anomalies on Medical Images while MDs diagnose the findings. In most cases the problem is not that AI might fail to classify or detect an anomaly. But, the problem arises when it is not 100% sure that AI is prepared to process the input, or that the incoming image is outside of its knowledge. In this paper an experiment is shown in the Chest X-Ray domain among four classes, such as Normal, Lung Opacity, Pneumonia and Covid. As part of the investigation, COVID is considered as a new type of finding; this paper shows how those data, which are “unknown” to the actual AI-tool in use, can be handled.

1.2. Out of Distribution Data

In the case of robust controlled systems, the adaptability of models in terms of new data and unpredictable conditions is of especially high importance. In the assessment of the effectiveness of AI-based techniques the out-of-distribution (OOD) data [1], present great challenges. OOD data differ in some way from the training database (in-distribution or ID data) in terms of either representation, context, or other conditions. That is, if $\mathbf{X}$ and $\mathbf{Z}$ are datasets with given distributions, and a model has been optimized on $x_1, x_2, \dots, x_n$ in $\mathbf{X}$, z in $\mathbf{Z}$ data is OOD exactly if and only if

$$\mathbf{X} \neq \mathbf{Z}$$

OOD data can significantly affect the performance of deep learning models, often leading to low accuracy or unexpected behaviour. The problem originates from the assumption that the training data is representative of all possible scenarios, however, this assumption is often violated in real-life situations.

The methods presented in this paper focus on the analysis of X-Ray data, available from online data set [2, 3]. These, shown by images, serve as the primary input data. These are classified into four different classes, based on the diagnoses: Normal, Lung Opacity, Pneumonia and Covid.

2. Evaluation Methods

In this chapter the methodological details of the project are described. After a domain-specific fine-tuning of a deep neural network, the results of filtering OOD data during the prediction phase were investigated.

2.1. Data prepared of OOD study

Again, each X-Ray image is associated with a class labelled by Normal, Lung Opacity, Pneumonia and Covid. Based on the actual dataset of four classes, the following strategy was defined: instead of training a classifier neural network using data from all the four classes, a network was trained to explicitly predict the first three classes only, i.e. Normal, Lung Opacity and Pneumonia.

The fourth class, Covid, was removed from the training set, thus, it is used here as the OOD sample set. This approach allows us to study the performance of the network and simultaneously address the challenges posed by OOD data.

2.2. OOD detection algorithms

Managing OOD data is a critical requirement in AI-based applications; it is not only about improving model performance, but also about ensuring reliability in diverse and dynamic environments. OOD detection methods aim to identify data points that differ significantly from the distribution of the training dataset.

In OOD data detection, the first step is determining the corresponding confidence level for the given input. This resulting confidence value is compared to a predefined threshold, thus, based on such a comparison it is easily concluded whether the given input belongs to OOD or ID. When the calculated confidence level lies below the threshold value, the data is considered as OOD data, otherwise the input is considered as ID data.

To calculate the confidence levels, we should consider the following methods:

- **Softmax-based methods** generally use the output of the softmax layer of the neural network to calculate the confidence level. ID data have higher softmax output than OOD data. Thus, defining a confidence score as a function of the softmax scores looks logical. However, such al-

gorithms only work for classification neural networks with a single softmax output [4–6].

- **Density-based methods** calculate confidence scores using the output of a selected layer other than the softmax; one of the feature-recognition layers is used here. The idea is based on the assumption, that OOD samples result in feature-scores that fall far from the "average features" of ID samples. In practice, density-based algorithms compute the distributions of the weights of features learned by the network, considering them as probabilistic models as a result. Thus, samples falling into areas of low density are assumed to be OOD observations. The distributions of the features, i.e. the density function of each feature is represented by a histogram. Thus, the method implements a histogram-based outlier score (HBOS) method. The training dataset corresponds to the ID data here, which is used for constructing histograms representing the feature-score density functions [7].

The choice of the OOD detection method depends on the specific requirements of the application, including the nature of the data, the architecture of the model and the desired level of robustness in performance [8]. Thanks to the fact that the HBOS algorithm is way more generally applicable, compared to the softmax-based methods, this method was chosen for presenting further application examples.

2.3. Actualization - Creating the right Neural Network

Altogether 24221 pairs of x - y data for training and 3058 pairs per class were used for validation. Resnet18 [9] network architecture was used and the X-Rays, i.e. input data are in the format of colour images with the dimensions of 299×299×3. With the help of a validation set separated from the training set, the training options and performance of the network could be optimized.

3. Results and discussions

When the Resnet18 architecture had been successfully deployed, a classifier neural network was trained on the entire dataset with 4 classes, and another network was trained on 3 classes only – omitting the one with Covid. Fig. 1. shows selected examples from each of the classes, as well as corresponding Grad-CAM "heat maps" indicating the most relevant areas in terms of the classification process. It is clear from the confusion matrix in Fig. 2, that a well-trained network

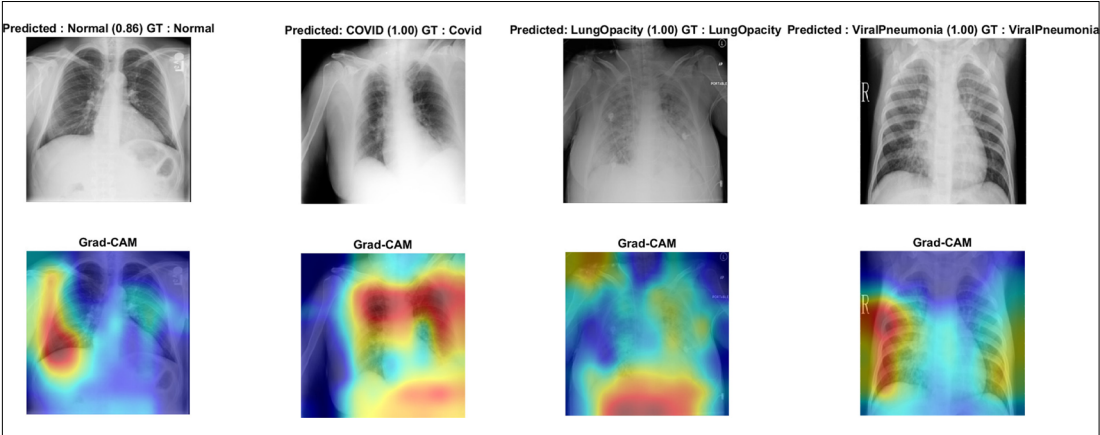


Fig. 1. X-Ray examples from classes Normal, Covid, Lung Opacity and Viral Pneumonia. Corresponding Grad-CAM pictures are also shown, indicating the most relevant areas in terms of classification.

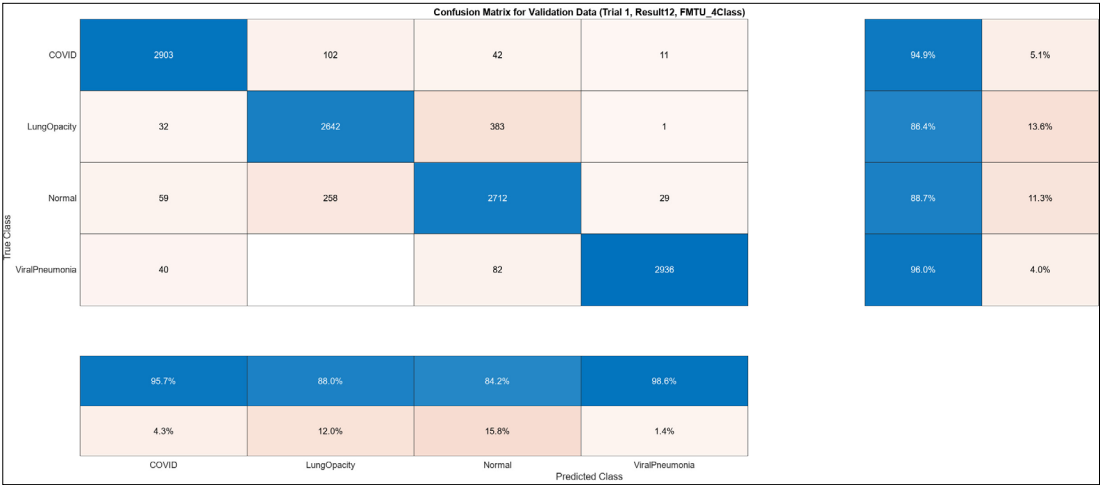


Fig. 2. Confusion matrix of the neural network trained with all the four classes.

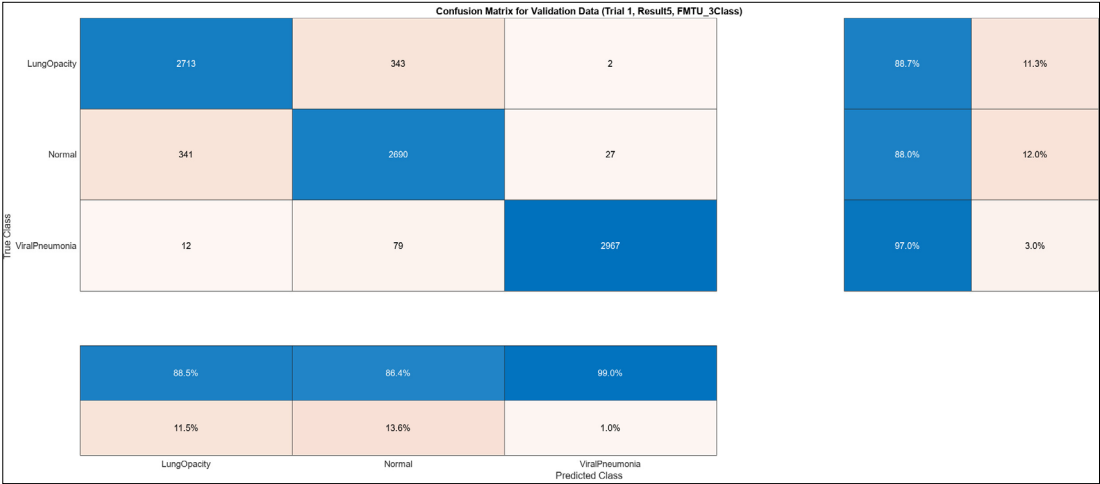


Fig. 3. KConfusion matrix of the neural network trained with three classes, omitting 'Covid'.

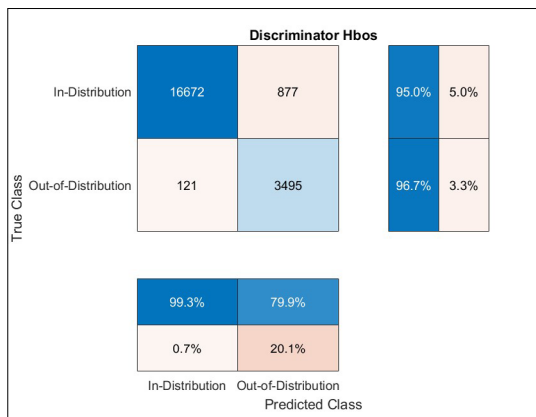


Fig. 4. Confusion matrix of the HBOS discriminator.

can effectively distinguish between all the four classes; similarly, the same applies, when only 3 of these classes were used for training – see Fig. 3

It is shown in Fig. 4 that the HBOS discriminator performs well with the network trained on 3 classes, considering the Covid class as OOD data. Even though, this might be not surprising in this actual situation, the fact that OOD data can be effectively separated from those on which the network was trained increases the usability of AI. In terms of diagnoses, it practically indicates that if a new disease should be considered, compared to those the AI-tool has been actually certified.

However, further research might be still needed on possible fine-tuning of a HBOS Discriminator. It is still a challenging task to automatically select the right layer within a neural network, such that the best possible performance by HBOS method is ensured. Authors of this paper will continue their research accordingly, to find the proper workflow for OOD sample detection.

References

- [1] Yang J., Zhou K., Li Y., Liu Z.: *Generalized out-of-Distribution Detection: A Survey*. (2021,10)
- [2] Chowdhury M. E. H., Rahman T., Khandakar A., Mazhar R., Kadir M. A., Mahbub Z. B., Islam K.R., Khan M. S., Iqbal A., Al-Emadi N., Reaz M. B. I., M. T. Islam: *Can AI Help in Screening Viral and COVID-19 Pneumonia?* IEEE Access, 8. (2020) 132665–132676.
- [3] Rahman T., Khandakar A., Qiblawey Y., Tahir A., Kiranyaz S., Kashem S.B.A., Islam M.T., Maadeed S.A., Zughaier S.M., Khan M.S., Chowdhury M.E., *Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection Using Chest X-ray Images*. arXiv preprint arXiv:2012.02238. 2020.
- [4] Shalev, Gal, Gabi Shalev, Joseph Keshet: *A Baseline for Detecting Out-of-Distribution Examples in Image Captioning*. In Proceedings of the 30th ACM International Conference on Multimedia, Lisboa Portugal. ACM, 2022. 4175–4184. <https://doi.org/10.1145/3503161.3548340>
- [5] Shiyu Liang, Yixuan Li, R. Srikant: *Enhancing the Reliability of Out-of-Distribution Image Detection in Neural Networks*. arXiv:1706.02690 [cs.LG], August 30, 2020. <http://arxiv.org/abs/1706.02690>.
- [6] Weitang Liu, Xiaoyun Wang, John D. Owens, Yixuan Li: *Energy-Based out-of-Distribution Detection* arXiv:2010.03759 [cs.LG], April 26, 2021. <http://arxiv.org/abs/2010.03759>
- [7] Markus Goldstein, Andreas Dengel: *Histogram-Based Outlier Score (hbos): A Fast Unsupervised Anomaly Detection Algorithm*. KI-2012: Poster and Demo Track 9 (2012).
- [8] Lee, Kimin, Kibok Lee, Honglak Lee, Jinwoo Shin: *A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks*. arXiv, October 27, 2018. <http://arxiv.org/abs/1807.03888>
- [9] He K., Zhang X., Ren S., Sun J.: *Deep Residual Learning for Image Recognition*. (2015. 12)