Informative and Intriguing Visual Features: UA.PT Bioinformatics in ImageCLEF Caption 2019

Ana Jorge Gonçalves*, Eduardo Pinho*, and Carlos Costa

DETI - Institute of Electronics and Informatics Engineering of Aveiro
University of Aveiro, Portugal
{ana.j.v.goncalves,eduardopinho,carlos.costa}@ua.pt

Abstract. Digital medical imaging has opened new advances in clinical decision support and treatment procedures since its inception. This leads to the creation of huge amounts of data that are often not fully exploited. The development and evaluation of representation learning techniques for automatic detection of concepts in medical images can make way for improved indexing, processing and retrieval capabilities in medical imaging archives.

This paper discloses several independent approaches for multi-label classification of biomedical concepts, in the context of the ImageCLEFmed Caption challenge of 2019. We emphasize the use of threshold tuning to optimize the quality of sample retrieval, as well as the differences between training a convolutional neural network end-to-end for supervised image classification, and training unsupervised learning models before linear classifiers. In the test results, the best mean F_1-score of 0.206 was obtained with the supervised approach, albeit with images of a larger resolution than for the dual-stage approaches.

Keywords: representation learning · deep learning · auto-encoders

1 Introduction

Medical imaging modalities are an essential and well established medium, and as the amount of medical images is dramatically growing, automatic and semi-automatic algorithms are quite pertinent for the extraction of information from biomedical image data [14]. Therefore, deep learning techniques are becoming increasingly useful and necessary for this aim, posing as a valuable key for the development of representation learning techniques, and ultimately for improving the quality of systems in healthcare.

[*] Both authors contributed equally, names are in alphabetical order.

The process of annotating images with useful information in this context is time-consuming and usually requires medical expertise. The development of powerful representations of images could enable the automatic detection of biomedical concepts in a medical imaging data set. The ImageCLEFmed initiative, inserted in ImageCLEF [4], has been focused on automatic concept detection, diagnosis, and question answering from medical images. In particular, the ImageCLEFmed Caption challenge of 2019 [11] has narrowed its scope into the task of *concept detection*, with the goal of recognizing biomedical concepts presented in medical images, using only the visual content.

This paper presents our solution proposal for the concept detection task, describing our methodology and evaluating its performance under the ImageCLEF 2019 challenge.

2 Methods

For this task, a data set with a total of 70,786 radiology images of several medical imaging modalities was provided from Radiology Objects in Context (ROCO) [12]. This global set was further split into training (56,629 images), validation (14,157 images) and test (10,000 images) sets by the organizers. Only the first two were accompanied with the list of concepts applicable to each image, whereas the testing set's ground truth was hidden from the participants.

The ImageCLEF Caption 2019 data set includes an overwhelming number of 5,216 unique concepts, not all of which can be reasonably considered due to the very small number of positive samples in the training and validation splits. In all of the methods described next, we have admitted only the 1,100 concepts with the highest number of samples with a positive occurrence of that concept (henceforth named *positive samples*). The label vectors were built based on a direct mapping from the UMLS concept unique identifier (CUI) to an index in the vector. The reverse mapping was kept for producing the textual list of concepts.

Also contributing to this decision, was the observed imbalance in the number of positives of each label, as discerned in Figure 1, making them difficult to train classifiers and evaluate them. By considering the 1,100 most frequent concepts, one could ensure, in the extreme case, a minimum number of 29 positive samples in the training set and 2 positive samples in the validation set. We admit that attempting to detect any less frequent concepts is unlikely to result in useful classifiers.

At ImageCLEF 2018, the highest mean F_1-score was obtained through unsupervised methods, namely by using the features of an adversarial auto-encoder, followed by logistic regression [13]. However, in relation to last year challenge, the number of images in the training set was reduced by 34 % and since all the data was annotated, we felt inclined to compare our past approach with the training of purely supervised methods.

Convolutional neural networks (CNNs) are considered one of the best approaches for image classification [15]. Unlike a 2-stage approach, where the ex-

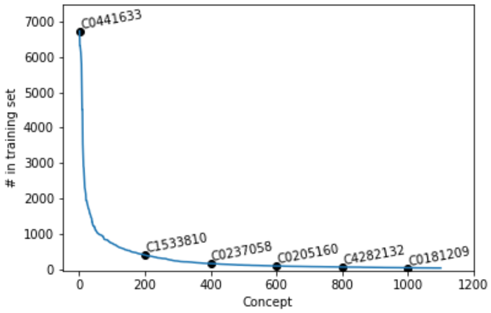

Fig. 1. Number of positives of each label, for the 1,100 more frequents ones. The concepts C0441633 (diagnostic scanning), C1533810(placed), C0237058 (no hydronephrosis), C0205160 (ruled out), C4282132 (malignancy) and C0181209 (hooks) are marked.

tracted feature descriptors are served as input to a trainable classifier, the images themselves are used in the learning process of the CNN. This could lead to a more focused guidance of the feature learning process across the multiple layers of the network. It was of our interest to compare this common supervised image classification method with the 2-stage pipeline involving unsupervised learning methods and simple classifiers.

Therefore, we have addressed the concept detection task with multiple independent approaches, which can be divided in two major groups:

- Through image representations, obtained by the implementation of several feature extraction methods:
  - Color and edge directivity descriptors, that were used as image descriptors.
  - An auto-encoder and an adversarial auto-encoder was trained and features were extracted from its bottleneck vector.
  - The ensemble of features obtained from the previous point was used for classification.
- An *end-to-end* approach, using two deep learning architectures:
  - A simple convolutional neural network model was assumed.
  - A residual neural network.

In every case, some form of optimum threshold tuning was employed, to overcome the classifier's focus on accuracy rather than F-measure. Further details are given in Sections 2.4 and 2.5.

Neural network training, feature extraction, and logistic regression were conducted using TensorFlow on one of the GPUs of an NVIDIA Tesla K80 graphics card in an Ubuntu server machine.

2.1 Color and Edge Directivity Descriptors

As traditional visual feature extraction algorithms are still very often considered in medical image recognition, these techniques contribute to a baseline, which we expect modern deep learning methods to surpass. For this purpose, we have extracted Color and Edge Directivity Descriptors (CEDDs) [2] from the images[1], after they were resized to a minimum size of 256 while keeping the aspect ratio. These low-level features accumulate color and texture information into a histogram of 144 bins per sample, and are known for their appealing accuracy in image retrieval tasks, when contrasted with their high compactness.

2.2 Adversarial Auto-encoder

For the unsupervised extraction of visual features from the medical images, an adversarial auto-encoder (AAE) [9] was trained on the given data set, with images resized to 64 pixels (64 × 64 × 3 inputs). While functioning as a typical auto-encoder, which seeks to minimize the information loss of passing samples through an information bottleneck (Equation 1), a discriminator D is also included. The purpose of D is to learn to distinguish latent codes produced by the encoder E from a prior code created by an arbitrary distribution $p(z)$, whereas E seeks to fool the code discriminator by approximating its output distribution to that of $p(z)$ (Equation 2). Based on the concept of Generative Adversarial Networks (GANs) [3], this min-max game of adversarial components provides variational inference to the basic auto-encoder structure while leading the encoder to match the prior distribution, thus regularizing the encoder. In this work, we have sampled $\epsilon \sim p(z)$ from a rectified unit-norm Gaussian distribution (as in, $\mathcal{N}(0, I)$ with all negative numbers replaced with zeros), which resulted in organically sparse latent codes.

$$\begin{aligned} x' &= G(E(x)) \\ \mathcal{L}_{rec}(x, x') &= \frac{1}{2N} \sum_{i}^{N} (x_i - x'_i)^2 \end{aligned} \tag{1}$$

$$V(E, D) = \min_{E} \max_{D} \mathbb{E}_{\mathrm{z} \sim p_{\mathrm{z}}}[\log D(\mathrm{z})] + \mathbb{E}_{\mathrm{x} \sim p(\mathrm{x})}[\log (1 - D(E(\mathrm{x})))] \tag{2}$$

Both encoder and decoder architecture are based on the ResNet19 [6], each component comprising four 2-layer residual blocks. At the end of the encoder, the final layer was subjected to a ReLU activation and a very light L_1 activation regularization (of factor 10^{-6}), thus contributing to the features' sparsity without deviating from the established prior. The code discriminator, on the other hand, is composed of three 1024-channel wide dense layers, with layer normalization [1], plus an output layer. Drop-out of rate 25% was also added before the output layer.

[1] Available on GitHub: `https://github.com/Enet4/ACEDD`

The AAE was trained for 30 epochs on the training set, with images resized to a minimum dimension of 72 pixels and then randomly cropped to a 64 x 64 square. All three RGB channels were kept with their values normalized to the [-1, 1] range with the formula $x/127.5 - 1$. Each iteration is composed of three optimization steps: the code discriminator step, the encoder regularization step, and the reconstruction step. The components were trained in this order with the Adam optimizer [5], a learning rate of 10^{-5}, beta parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and a mini-batch size of 32.

Moreover, in order to better understand the influence of the adversarial auto-encoder's regularization phase, a separate auto-encoder (AE) was trained with the same encoder and decoder characteristics as in the adversarial one, but without the adversarial loss. In this case, each iteration is only composed of the reconstruction step, mainly influenced by the loss $\mathcal{L}_{rec}$.

2.3 Early Fusion: Auto-encoder and Adversarial Auto-encoder

The combination of information from different techniques seems intuitively appealing for improving the performance of the learning process. More recently, there is an attempt of combining traditional image characteristics with high-level features [7].

Although it is known that the combination of low-level features, such as color and texture in CEDD, is important to the quality of visual features, it is not as clear whether the combination of high-level features obtained from two auto-encoders can benefit from an early fusion. Hence, we hereby took the features obtained from the AE and the AAE, and concatenated them to form a 1024-dimensional feature set, for its subsequent use in the logistic regression step alongside the other feature sets.

2.4 Logistic Regression

For each of the previously described set of features, logistic regression classifiers were trained for the chosen labels, for a set of predefined operating point thresholds: 0.075, 0.1, 0.125, and 0.15. The linear classifiers were trained with the Adam optimizer [5], with a mini-batch size of 128, until the best F_1-score among the various thresholds would reach a plateau. Our experiments suggests that training in this phase with a very small learning rate, often 10^{-5} in our experiments, for a large number of epochs (more than 500), helps the training process to find a more optimal solution.

Aware of the presence of concepts with a very low number of positive samples, one may wonder whether certain labels were not well trained or resulted in uninformative classifiers. To mitigate this, we calculated the area under the curve (AUC) of each binary classifier's receiver operating characteristic curve (ROC). Afterwards, we have tested whether ignoring the predictions of concepts where the AUC was lower that 0.5 would potentially improve the overall performance. Testing this hypothesis on the validation set, it is revealed that this would improve the mean F_1-score in most cases, albeit only slightly. For the features of the AAE, as an example, this tweak has only increased the score by 5×10^{-5}. With the features of the simple AE, the score was only improved by 7.3×10^{-4}. We held this mechanism away from the classifiers trained with the CEDD feature set.

Probabilistic classifiers minimizing binary cross-entropy inherently optimize for the accuracy of predictions. However, accuracy is overoptimistic when labels have a very low number of positives, as is the case in this task, making a poor metric for the classifiers' usefulness. As recognized by past work in the scope of ImageCLEF concept detection, adjusting the operating point thresholds to optimize the F_1-score provides significant improvements in the final metric values, in spite of the known implications of this practice [8]. In order to adjust the probabilistic threshold for optimizing the F_1-score, the provided validation set was split in five folds. For each one, we used a granular sequential search (with a granularity of 0.01) to identify the threshold resulting in the highest F_1-score and the calculated median of the five optimizing thresholds was used for the prediction over the testing set, using the trained classifiers.

In the event that a sample was predicted to have more than 100 concepts before submission, the list of concepts was trimmed by the less frequent concepts. In practice, this has only happened to the linear classifiers trained using CEDD.

2.5 End-to-end Convolutional Neural Network

A simple CNN was designed (Table 1) and trained for multi-label classification, thus once again treating concepts as labels. Conv2D stands for 2D convolution layer, GAP for global average pooling and FC for fully connected layer. Training samples were augmented using random square random crops, experimented with different sized squares. In one approach, denoted as CNN-A-256px, we used 256 pixel-wide and excluded the layer *Conv2D-5*. In CNN-B, the full CNN was used with 128 (CNN-B-128px) and 64 (CNN-B-64px) pixel-wide images. Validation and test samples were simply resized to fit these dimensions, according to the training process.

Table 1. The specification of the CNN used.

Layer Type	Kernel/Stride	Output Shape	Details
Conv2D-1	5 x 5 / 2	$64 \times 64 \times 64$	ReLU activation
Conv2D-2	3 x 3 / 2	$32 \times 32 \times 128$	ReLU activation
Conv2D-3	3 x 3 / 2	$16 \times 16 \times 256$	ReLU activation
Conv2D-4	3 x 3 / 2	$8 \times 8 \times 512$	ReLU activation
Conv2D-5	3 x 3 / 2	$4 \times 4 \times 512$	ReLU activation
GAP	-	512	-
Dropout	-	-	50 %
FC	-	1100	sigmoid activation

Moreover, to serve as a more intuitive means of comparison, the same architecture as the encoder in the AAE and AE, based on ResNet19 [6], was trained for end-to-end classification. It is composed by five ResNet blocks, with the architectures depicted in Tables 2 and 3. We employed the same process of data augmentation as in the previously described CNN, resulting in 64×64 images. In Table 3, BN means batch normalization and c_{in} and c_{out} represent the input and output channels for the ResNet block, respectively. The *Addition* layer depicts the addition of the previous stages: the first, with one convolution layer and the second, with two convolution layers.

Table 2. ResNet architecture.

Layer Type	Output Shape
ResBlock	$64 \times 64 \times 64$
ResBlock	$32 \times 32 \times 128$
ResBlock	$16 \times 16 \times 256$
ResBlock	$8 \times 8 \times 256$
ResBlock	$4 \times 4 \times 512$
GAP, ReLU	512
Dropout	512
FC	1100

Table 3. Residual block specification.

Layer Type	Kernel/Stride	Output Shape
Conv2D	3×3 / 2	$h/2 \times w/2 \times c_{out}$
BN, ReLU	-	$h \times w \times c_{in}$
Conv2D	3×3 / 1	$h \times w \times c_{out}$
BN, ReLU	-	$h \times w \times c_{out}$
Conv2D, BN	3×3 / 2	$h/2 \times w/2 \times c_{out}$
Addition	-	$h/2 \times w/2 \times c_{out}$

Both end-to-end deep learning models were trained with the AMSGrad optimizer [16], with a batch size of 32 and a learning rate of 10^{-4} with a decay of $2x10^{-5}$ over each update and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For each model, threshold fine tuning was performed by evaluating the F_1-score performance for multiple thresholds, using the validation data set. Thereafter, we determined the threshold which would yield the optimal mean F_1-score on the validation set.

3 Results and Discussion

Alongside with the metrics obtained from our submissions, we also present a brief qualitative analysis as part of our results.

3.1 Qualitative Feature Analysis

The visualizations of the features for five of our models were obtained by training dimensionality reduction algorithms, namely principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) [10]. The visualizations are presented in Figures 2 and 3, respectively, using a stratified portion of 5 % of the training set, where the extreme outliers were removed from the figures. The official and open implementation in Python of UMAP[2] was used

[2] Available on GitHub: `https://github.com/lmcinnes/umap`

and the algorithm was configured with 15 as the number of neighbors and 0.1 as the minimum distance.

For the CNN, the features were extracted at the GAP layer, whereas in the remaining the visualizations depict the features extracted before the classification process. The points associated with the concepts C0441633 (diagnostic scanning) , C0817096 (thoracics) and C0935598 (sagittal planes set) are labeled in red, green and blue, respectively, each painted in an additive fashion.

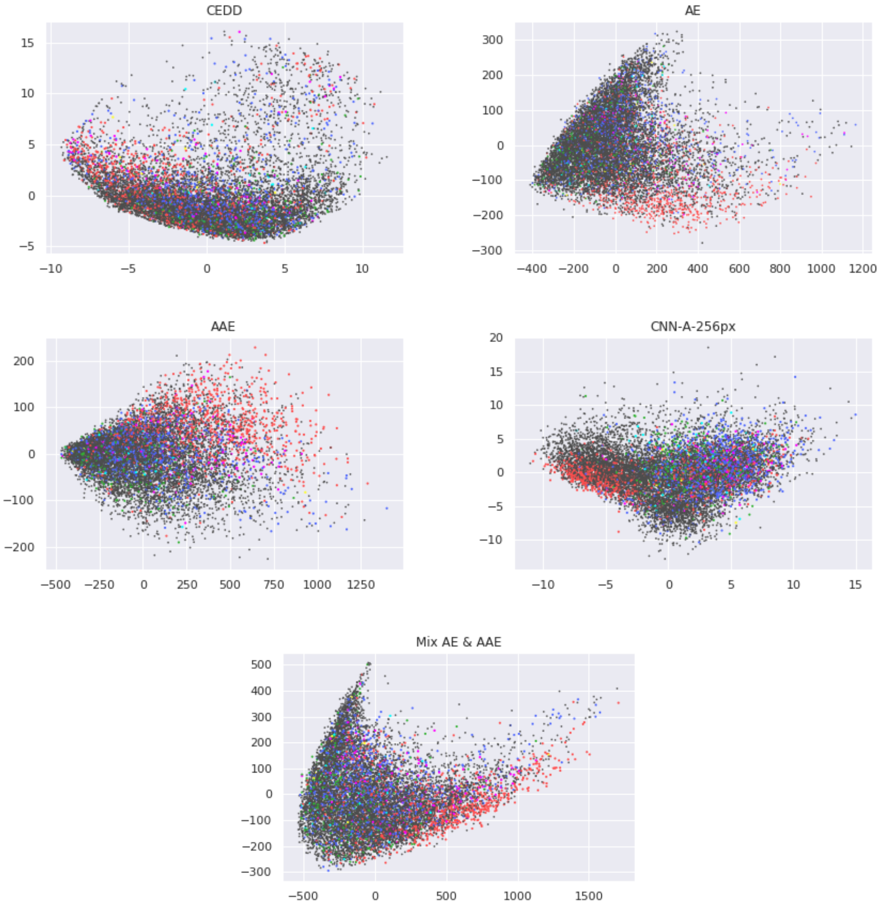

Fig. 2. The 2D projections of the features, obtained with PCA, for each model.

Comparing the two types of dimensionality reduction algorithms, we notice that the representations obtained with PCA have more outliers. In a good representation, the samples will be linearly separable based on their associated concepts. In both types of representations, we can identify regions in the man-

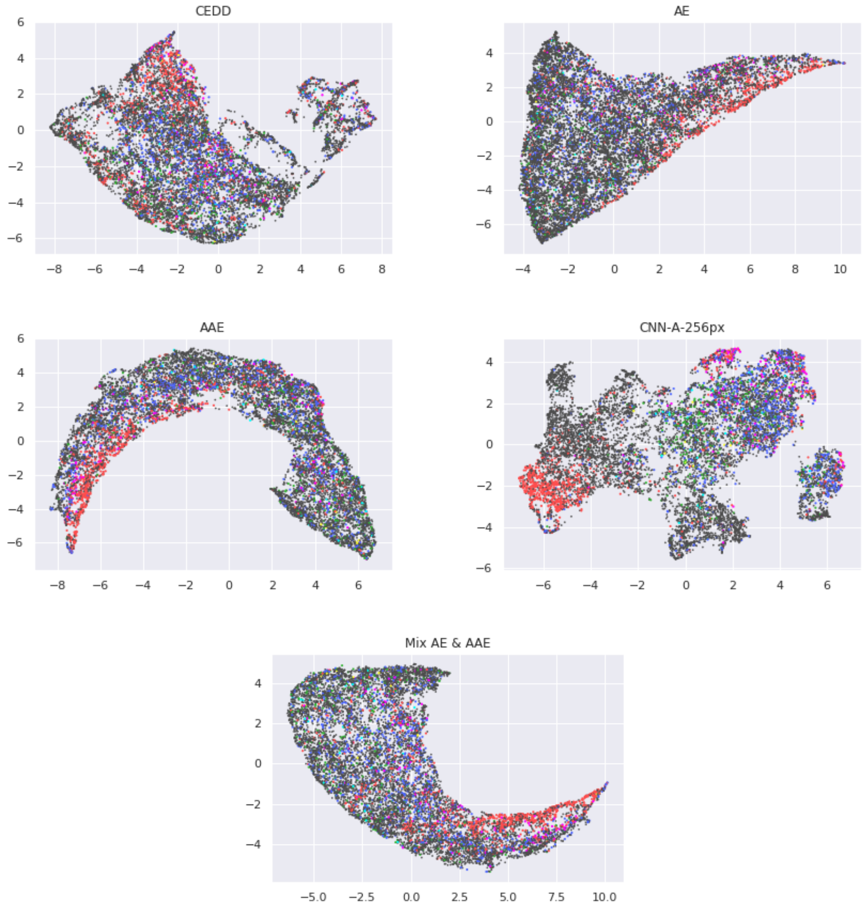

Fig. 3. The 2D projections of the features, obtained with UMAP, for each model.

ifold in which points of one of the chosen labels are mostly gathered, with this being more perceptible in the representations obtained with UMAP and for the CNN trained end-to-end. In fact, the clustering of representations with common labels is highly expected, even more so for the CNN, since the feature learning in this case was uniquely guided by the target labels. In general, these observations are a rough approximation of the effective performance of each method.

3.2 Quantitative Results

The metrics on the validation and test set for each submission are depicted in Table 4. Both *Val* F_1*-score* and *Test* F_1*-score* represent the F_1-scores averaged sample-wise. When applied to the validation set, all concepts of the ground truth were considered, even though the predictions were made only assuming the 1,100 most frequent concepts on the training set, and so always predicting ”negative” for the remaining concepts.

Table 4. The final results obtained in the concept detection task, ranked by the list of all valid submissions.

Rank	Run file name	Details	Val F_1-score	Test F_1-score
16	simplenet	CNN-A-256px	0.19103	0.20586
19	simplenet128x128	CNN-B-128px	0.17636	0.18934
20	mix-1100-o0-2019-05-06_1311	AE + AAE	0.16973	0.18254
21	aae-1100-o0-2019-05-02_1509	AAE	0.16064	0.17601
24	ae-1100-o0-2019-05-02_1453	AE	0.16021	0.17152
25	cedd-1100-o0-2019-05-03_0937-trim	CEDD	0.15725	0.16679
38	simplenet64x64	CNN-B-64px	0.11747	0.12799
39	resnet19-cnn	CNN-RN	0.11813	0.12695

The scores obtained from end-to-end CNN models (*CNN-A-256px*, *CNN-B-128px*, *CNN-B-64px*) was highly varied, which demonstrates the impact of the input shape, as well as neural network architecture, in the performance of the model. With an image resolution of 64×64, this approach did not perform better than any of the 2-stage procedures. On the other hand, higher resolutions have contributed to significantly better scores.

Concerning the unsupervised methods, the mean F_1-score obtained with CEDDs (*CEDD*) was lower than with the deep learning architectures, probably because they lack representation ability for high-level problems, an effect that was also observed in prior work [13]. Even with a smaller data set than the previous edition of the concept detection task, unsupervised methods have pushed the performance limits within the initially proposed input shape.

The early fusion of the features obtained from the two auto-encoders (*AE + AAE*) was also beneficial, resulting in a higher score than any of the two forms independently(*AE* and *AAE*), suggesting that this aggregation was not entirely redundant, thus providing another useful distribution.

It is also worth noting that, much unlike in our previous participations in the same task, the instance-wise mean F_1-scores on the testing set were higher than on the validation set. This effect is even more noteworthy, since these methods relied on the validation set for threshold optimization, and as such the classifiers were fit for both the training and validation sets. This consistent discrepancy is due to the fact that the test set did not include any new concepts that were not present in the overall data set provided at the beginning of the challenge, whereas the validation set contained some concepts which were not present in the training set.

4 Conclusion

In the context of the ImageCLEFmed Caption challenge, we did an assessment of feature learning techniques for concept detection from radiology images of several medical imaging modalities. The extraction of informative – and intriguing – visual features can yield great potential for multiple use cases in medical imaging systems, including automated image labelling and content-based retrieval.

We had confirmed the greater potential of deeper architectures for the construction of more powerful representations, in comparison with low-level feature extraction algorithms. With the data set size being significantly smaller in this edition of the challenge, this was seen as an opportunity to compare end-to-end classification models with the use of unsupervised learning methods. The outperforming CNN model had a larger image size as input, making this factor a counterbalance to obtain a better F_1-score than with the unsupervised models. In fact, at a late stage of these experiments, we have identified that the attempted resolution of 64×64 is insufficient to attain better results. In the end, a simple CNN with a higher resolution showed the best performance among these submissions. Time constraints have not enabled us to combine the two ideas together in our submissions.

The quality of the results obtained in this edition may also be attributed to the use of threshold tuning to optimize the F_1-score. Without an adjustment of the classifiers' operating point, these methods would have a focus towards the highest accuracy of a prediction, which is not as useful in the context of information retrieval. When the number of positive samples is low, the potential retrieval of less relevant entries is compensated by a significantly greater chance of receiving relevant images. Nevertheless, we understand that the focus of a single metric can distort the perception of quality among multiple methods in the challenge, such that a change of performance metric could result in different rankings [8]. Therefore, it may be insightful for future editions to also present other metrics alongside the main metric, such as the mean precision and recall on the testing set.

This year presented an increase in participants engagement in the challenge, which might echo the interest in solving timely situations in medical information retrieval and automated medical data analysis. We believe that further investment in the challenge, both from participants and organizers, will enable the implementation of these solutions in real-world scenarios.

Acknowledgments

This work was supported by the Integrated Programme of SR&TD "SOCA" (Ref. CENTRO-01-0145-FEDER-000010), co-funded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund.

References

1. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer Normalization (jul 2016). https://doi.org/10.1038/nature14236, `http://arxiv.org/abs/1607.06450`
2. Chatzichristofis, S., Boutalis, Y.: CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. pp. 312–322 (01 2008). https://doi.org/10.1007/978-3-540-79547-6_30
3. Goodfellow, I.J., Pouget-abadie, J., Mirza, M., Xu, B., Warde-farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Nets pp. 1–9 (2014)
4. Ionescu, B., Müller, H., Péteri, R., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Cid, Y.D., Liauchuk, V., Kovalev, V., Ben Abacha, A., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Pelka, O., Friedrich, C.M., Chamberlain, J., Clark, A., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasillopoulos, N., Karampidis, K.: Overview of ImageCLEF 2019: Challenges, datasets and evaluation. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Tenth International Conference of the CLEF Association (CLEF 2019), LNCS Lecture Notes in Computer Science, Springer, Lugano, Switzerland (September 09-12 2019)
5. Kingma, D.P., Ba, J.L.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2015), `https://arxiv.org/pdf/1412.6980.pdf`
6. Kurach, K., Lucic, M., Zhai, X., Michalski, M., Gelly, S.: The GAN landscape: Losses, architectures, regularization, and normalization. CoRR **abs/1807.04720** (2018), `http://arxiv.org/abs/1807.04720`
7. Lai, Z., Deng, H.: Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. Computational Intelligence and Neuroscience **2018**, 1–13 (09 2018). https://doi.org/10.1155/2018/2061516
8. Lipton, Z.C., Elkan, C., Narayanaswamy, B.: Thresholding Classifiers to Maximize F1 Score. Machine Learning and Knowledge Discovery in Databases **8725**, 225—-239 (feb 2014), `http://arxiv.org/abs/1402.1892`
9. Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B.: Adversarial autoencoders (nov 2015), `http://arxiv.org/abs/1511.05644`
10. McInnes, L., Healy, J., Melville, J.: UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv e-prints arXiv:1802.03426 (Feb 2018)
11. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Overview of the ImageCLEFmed 2019 concept prediction task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR- WS.org), ISSN 1613-0073, vol. 2380. Lugano, Switzerland (September 09-12 2019)
12. Pelka, O., Koitka, S., Rückert, J., Nensa, F., Friedrich, C.M.: Radiology objects in context (ROCO): A multimodal image dataset. In: Stoyanov, D., Taylor, Z., Balocco, S., Sznitman, R., Martel, A., Maier-Hein, L., Duong, L., Zahnd, G., Demirci, S., Albarqouni, S., Lee, S.L., Moriconi, S., Cheplygina, V., Mateus, D., Trucco, E., Granger, E., Jannin, P. (eds.) Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. pp. 180–189. Springer International Publishing, Cham (2018)
13. Pinho, E., Costa, C.: Feature learning with adversarial networks for concept detection in medical images: Ua.pt bioinformatics at imageclef 2018. In: Working Notes of CLEF (09 2018)

14. Pinho, E., Costa, C.: Unsupervised learning for concept detection in medical images: A comparative analysis. Applied Sciences **8**(8) (2018). https://doi.org/10.3390/app8081213
15. Rawat, W., Wang, Z.: Deep convolutional neural networks for image classification: A comprehensive review. Neural Computation **29**(9), 2352 – 2449 (2017). https://doi.org/10.1162/neco_a_00990
16. Reddi, S.J., Kale, S., Kumar, S.: On the Convergence of Adam and Beyond. In: International Conference on Learning Representations (2018), `https://openreview.net/forum?id=ryQu7f-RZ`