

Fuzzy Clustering of Histopathological Images Using Deep Learning Embeddings

Salvatore Calderaro¹, Giosué Lo Bosco¹, Riccardo Rizzo² and Filippo Vella²

¹*Dipartimento di Matematica e Informatica, Università degli studi di Palermo, Via Archirafi 34, 90123 Palermo, Italy*

²*Institute for High-Performance Computing and Networking (ICAR), National Research Council of Italy (CNR), Italy, 90146, Palermo*

Abstract

Metric learning is a machine learning approach that aims to learn a new distance metric by increasing (reducing) the similarity of examples belonging to the same (different) classes. The output of these approaches are embeddings, where the input data are mapped to improve a crisp or fuzzy classification process. The deep metric learning approaches regard metric learning, implemented by using deep neural networks. Such models have the advantage to discover very representative nonlinear embeddings. In this work, we propose a triplet network deep metric learning approach, based on ResNet50, to find a representative embedding for the unsupervised fuzzy classification of benign and malignant histopathological images of breast cancer tissues. Experiments computed on the BreakHis benchmark dataset, using Fuzzy C-Means Clustering, show the benefit of using very low dimensional embeddings found by the deep metric learning approach.

Keywords

Histopathological Images Classification, Deep Learning, Metric Learning,

1. Introduction

Digital pathology has emerged with the use of Whole Slide Images (WSI), a kind of digitization of patient tissue sample usually with a high resolution [1]. WSI can have a resolution of 10.000×10.000 pixels and present high morphological variance and various artefacts. A diagnosis based on WSI can be obtained with a very time-consuming process by a trained pathologist. Systems based on Machine Learning (ML) algorithms can save time by helping pathologists focus on the interesting part of histopathological images.

Today the standard computer hardware is not powerful enough to process the WSI input using a standard approach, i.e. it is not possible to input the whole image to the ML algorithm. It is also impossible to down-sample the whole image because it will lose critical features, so that the WSI processing is usually carried on by dividing the image into tiles or picking random patches.

The label of the WSI will be the combination of the labels obtained by processing the tiles or patches extracted [2]. Due to the morphological variance of the WSI, the labelling process can

Wilf 2021: The 13th International Workshop on Fuzzy Logic and Applications, December 20–22, 2021, Vietri sul Mare, Italy

✉ salvatore.calderaro01@community.unipa.it (S. Calderaro); giosue.lobosco@unipa.it (G. Lo Bosco); riccardo.rizzo@cnr.it (R. Rizzo); filippo.vella@icar.cnr.it (F. Vella)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

be difficult. Many patches can present only healthy tissue even if the global label is different. Moreover, patch-level processing should be carried out at mixed magnification levels to maintain an idea of the context of image details.

Patch analysis also has an intuitive drawback: it cannot detect characteristics bigger than the patch itself as shape and magnitude of the tumour, which adds another difficulty to the automatic processing of the WSI images.

Even considering these drawbacks, patch analysis remains one of the most used techniques to approach WSI processing, and the development of patch processing and classification systems is an active research field. Some datasets are considered standard and used to compare the performances of different systems. One of these is the BreakHis dataset (BH) [3] that is constituted by 7909 histopathological images of 8 types of breast cancer taken from 82 patients; a more in the deep description of the BH dataset is in Section 2.

The present work aims to study the application of embedding techniques on the fuzzy clustering of histopathological images coming from the BH dataset. Using fuzzy clustering as a methodology providing a continuous value membership rather than a crisp one could be more informative and useful for physicians.

The method used for the embedding creation is metric learning [4, 5] that aims to learn a new embedding by reducing a chosen distance measures between examples of the same class and increasing the same distance between objects of different classes. In this embedding, similar data will be closer, dissimilar data will be far away, and this should improve the performances of clustering and classification algorithms. In recent years new deep neural networks [5] have been used to learn a metric that allows to measure the similarity/dissimilarity between objects and to create a mapping in a new representation space in which the above conditions are respected. Some applications of deep metric learning, and in general of metric learning, are [4, 6]: image retrieval, visual tracking, person re-identification, medical problems, face verification and recognition, signature verification, zero-shot learning etc. The rest of the paper is organized as follows: Section 2 describes the deep metric learning used method, and the BH dataset, Section 3 summarizes the experiments conducted to test the embedding effectiveness for the case of fuzzy C-means, and Section 4 contains the concluding remarks.

2. Materials and Methods

In this work, we want to explore the effectiveness of embedded medical images in a fuzzy clustering task. The embedding method used is based on a deep neural network called triplet network, implemented in the framework Pytorch Metric Learning [7]. We scan the embedding dimensions from 2 to 256 to understand the most effective dimension in fuzzy clustering. The following subsections deeply describe the BH dataset, the architecture of the triplet network and its training procedure.

2.1. BreakHis dataset

The BreakHis dataset comprises 7909 images at a resolution of 700×460 pixels obtained from tissue samples of 82 breast cancer patients. These samples are organized into two classes,

Table 1

The BreakHis dataset images.

Benign 2480	Adenosis (A)	444
	Fibroadenoma (F)	1014
	Tubular Adenoma (TA)	453
	Phyllodes Tumour (PT)	569
Malignant 5429	Ductal Carcinoma (DA)	3451
	Lobular Carcinoma (LC)	626
	Mucinous Carcinoma (MC)	792
	Papillary Carcinoma (PC)	560

benign and malignant, each of them separated into four sub-classes, according to the structure in Table 1. The images were acquired at different magnification levels ($40\times$, $100\times$, $200\times$, $400\times$), but we neglect this detail in the following discussion. In this paper, we are only interested in the dichotomy of benign vs malignant, and Table 1 shows that the number of images in the malignant class doubles the number of benign images. We implement an augmentation procedure (flipping and rotating images) to balance the dataset increasing the images from classes adenosis, tubular and phyllodes tumour. We excluded the fibroadenoma class from the augmentation because it is the most populated class. Using this strategy, we obtain 5412 images from the benign class, so the dataset used for the experiments was made by 5429 malignant images and 5412 benign images. Moreover, in this work, we do not consider any pre-processing technique on images, e.g. stain-normalization or whitening. We also decided not to consider the (little) noise on data labelling since the same images for the patient ID:13412 are present in two malignant sub-classes (Ductal Carcinoma and Lobular Carcinoma).

Other considerations should be made on the training and test splitting of the dataset after augmentation. As written before, we do not consider the magnification factor and cluster all the images as benign and malignant. When splitting these images between the training and test datasets, we use the same split for the embedding neural network and the classification or clustering algorithm so that, during the test, the algorithm will calculate the embedding of a new image (a test image) and the classifier or the clustering algorithm will process an embedded data never seen before. Moreover, the training/test splitting can be done in two different modalities: at the image level or patient level. At the image level, the training/test dataset split neglects the patient information, meaning that some images of the same patient can be in the training dataset and some in the test dataset; the accuracy calculated using this splitting is called Image-Level Accuracy (ILA). At the patient level, the patient information became central, and the training/test split contains 58 patients as part of the training set and 24 patients as part of the test set; this means that all the images of a patient can be in the training set or in the test set. In this case, the accuracy for classification is called Patient-Level Accuracy (PLA) and is calculated as reported in [8].

2.2. Metric Learning

In deep metric learning, the embedding is implemented by using one or more neural networks, the two most used methods use siamese networks [9], and triplet networks [10]. Both architectures are composed of neural networks with sharing weights: siamese uses two neural networks while triplet networks three. In this work, we adopt a triplet network.

Triplets networks use 3 inputs: an anchor image x_a , an image of the same class of the anchor x_p (the positive example), and an image of a different class x_n (the negative example).

The goal is that the distance between anchor and negative sample representation $d(r_a, r_n)$ is greater than a margin m respect to the anchor and positive sample representations $d(r_a, r_p)$. There are several choices for the metric loss [5, 6]: contrastive loss, pair loss, triplet loss, lifted structure loss, binomial deviance loss etc. Another key factor in this technique is the choice of the metric to measure the similarity/dissimilarity between samples. The common choices are cosine similarity and euclidean distance.

In this work the chosen loss is triplet margin, defined as:

$$L_{triplet} = \max \{0, m + d(r_a, r_p) - d(r_a, r_n)\} \quad (1)$$

During the training phase, a mining procedure selects, from the input minibatch, the more effective triplets to use in the network training [5, 6]; the most common strategies are easy negative mining, hard negative mining and semi-hard negative mining, which is the one used.

With semi-hard negative mining, the negative sample is farther away from the anchor with respect to the positive one, but this distance is not greater than the margin m . In this case, $L_{triplet} < m$ as in the equation below:

$$d(r_a, r_p) < d(r_a, r_n) < d(r_a, r_p) + m. \quad (2)$$

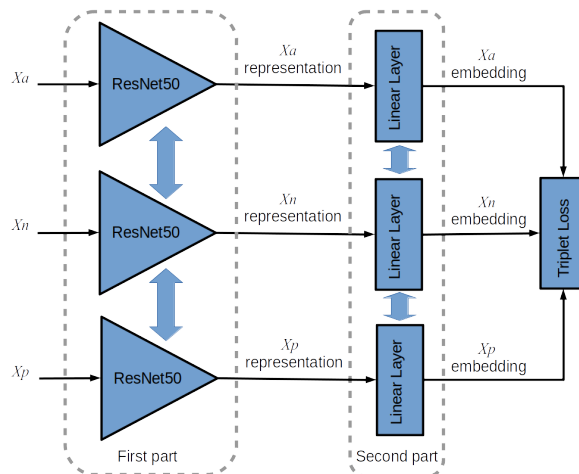


Figure 1: A representation of the network; the vertical double arrow indicates the weights sharing.

2.3. Implementation Details

The network used for the embedding has two parts: the first part implements a feature extraction and transforms the input images in a feature vector of dimension 2048; this is a ResNet50 network [11] pre-trained using the ImageNet dataset; a single linear layer constitutes the second part.

Only the second part of the network is trained by using the $L_{triplet}$ loss in eq. 1; the distance function used is the cosine similarity, and the optimization algorithm is the RMSprop with a learning rate of 1×10^{-5} and a weight decay factor of 1×10^{-4} .

A representation of the network is in Fig. 1.

3. Experiments and Results

For the given training set, the following elaborations are performed on every batch:

- the deep neural network, ResNet50 process the input data;
- the mining process identifies all the triplets found in the batch, according to the element labels. From this set, the *semi-hard* triplets are extracted as in eq. 2;
- the found triplets are used to calculate the loss, through the cosine similarity;
- the *reducer* calculates the average among *loss* value higher than a threshold, empirically set to zero.

The training process was repeated for the embedding sizes 2, 3, 4, 8, 16, 32, 64, 128, 256.

Figure 2 reports a representation of these embeddings. The figures for embedding sizes 2 and 3, (Figure 2a, 2b) are shown as 2d and 3d plots. A locus of points with radial symmetry contains all the mapped samples. This radial symmetry is probably due to the cosine similarity chosen as the metric to estimate the sample closeness.

The embedding with sizes greater than 3 (Figures 2c-2i) have been mapped in a 2-d space through the UMAP Uniform Manifold Approximation and Projection for Dimension Reduction library [12].

3.1. BreakHis Images Fuzzy Clustering

Differently from crisp clustering, fuzzy clustering allows associating a sample to centroids with a belonging degree. This property is useful to cope with the uncertainty of the hard attribution of values to artificially created clusters and have a smooth behaviour for samples far from centroids and potentially assimilable to multiple clusters.

In the experiment, as a fuzzy clustering solution, we have used fuzzy C-means [13]. The used implementation adopts the Euclidean distance as a clustering metric. To evaluate the goodness of the fuzzy clustering, some measures have been adopted, such as Partition Coefficient (PC)[14], Partition Entropy (PE)[15], Modified Partition Coefficient (MPC) and Modified Partition Entropy(MPE)[16].

The Partition Coefficient (PC) measures the average fuzzy degree of belonging of the samples to the clusters using the partition matrix: the larger is the value, the better is the partition [14]. The Partition Entropy (PE) measures the average entropy of the fuzzy belonging of a sample to

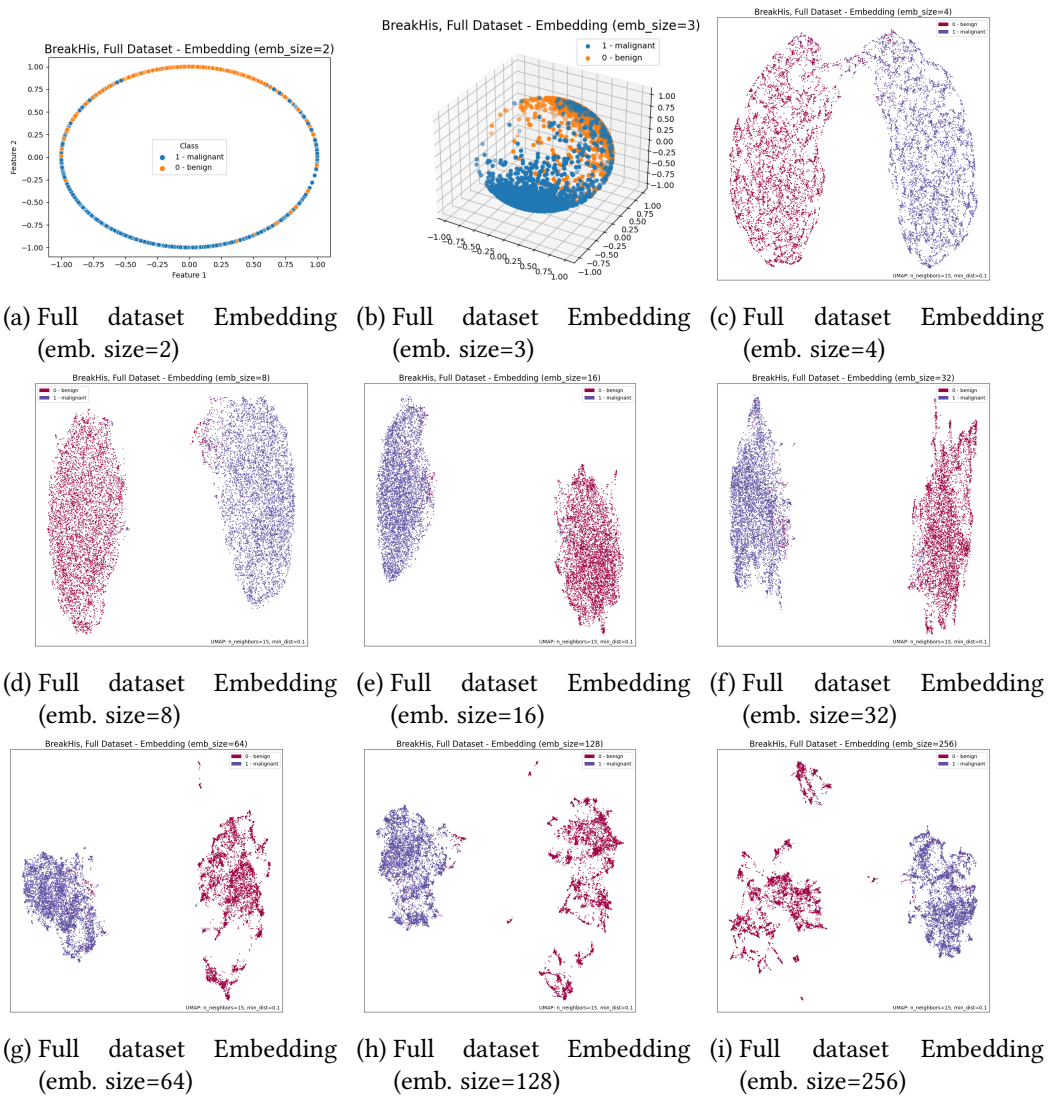


Figure 2: Bidimensional Plot of dataset embeddings. Figures (a) and (b) show the original coordinate space in 2d and 3d. The colours differentiate the object into two classes, benign and malignant. Figures (c)-(i) show the data with embedding sizes ranging from 4 to 256 visualized in 2d by the UMAP approach.

a cluster. The lower is the value, the better is the partition [15]. MPE and MPC are modifications of PE and PC that take into account the number of centroids and the number of samples. MPE should be minimized while MPC should be maximized [16].

Figure 3 shows the value of four metrics of the fuzzy clustering process when input data are processed with triplet loss. The PE and MPE values, which result coincident for all the experiments, show that the values are lower for a reduced number of centroids, and the entropy values increase proportionally to the dimension of the embeddings. These metrics advise using few clusters and a reduced embedding size. The values of PC and MPC are coherent and tend to

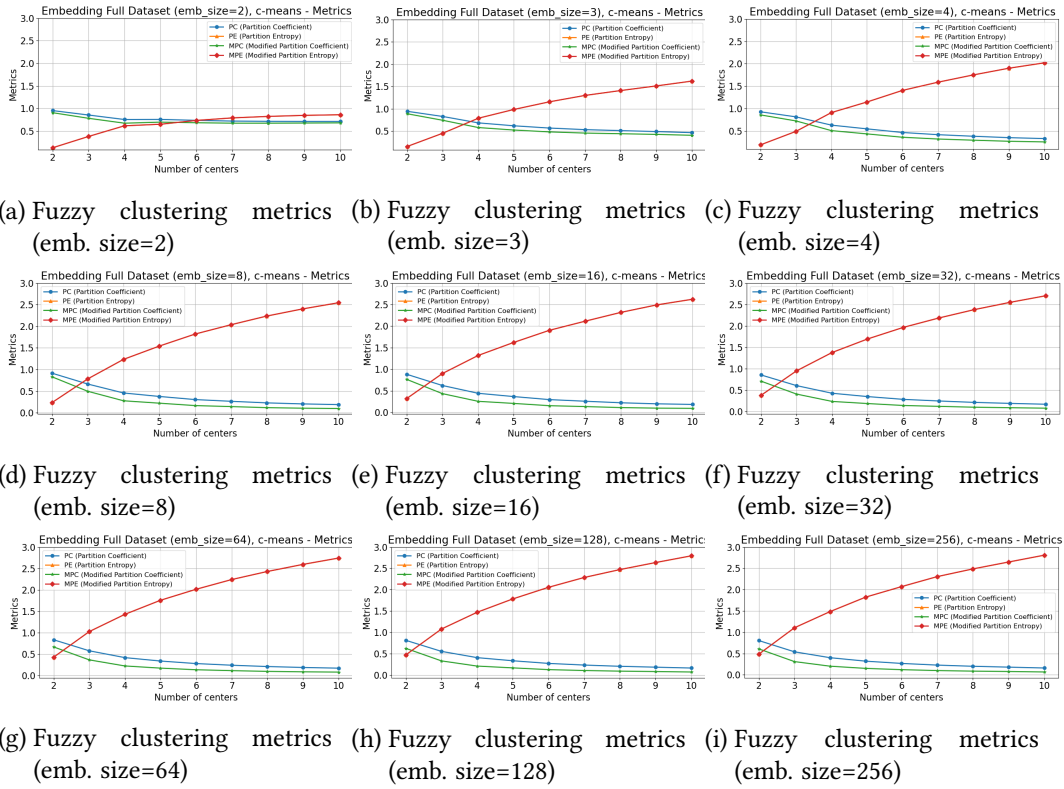


Figure 3: Metrics plot for fuzzy clustering with a dimension of embeddings ranging from 2 to 256. PE is plotted with orange triangles, MPE is plotted with red diamonds, PC is plotted with blue dots, MPC is plotted with green stars. In all the plots, MPE is superimposed to PE

decrease with a larger number of clusters. When the embedding size is larger, the values are even lower. Also, for these metrics, the best results are obtained with lower dimensions and few clusters.

To evaluate the clusters obtained with different embedding dimensions we considered the sets of images whose membership is in the range $[0.45, 0.55]$, i.e. the ones with an high level of uncertainty. The number of these images is reported in Table 2 for different embeddings (rows) and classes (columns). It is possible to notice that the number of uncertain images grows with embedding size. Moreover, there are 8 images which belongs to all the uncertain sets related to an embedding. These images are reported in Figure 4, and are composed by an adenosis image, two fibroadenomas, one tubular image and four ductal carcinoma images. It is worth noting that ductal carcinoma is the most common tumor in fact almost an half of the images in the original dataset are of this kind (3451 out of 7909), while fibroadenoma is often misclassified as malignant [3].

Table 2

Distribution of uncertain images (memberships in the range $[0.45, 0.55]$) for different embedding dimension (rows). In columns the different classes of images are represented by the following labels: Adenosis (A), Fibroadenoma (F), Tubular Adenoma (TA), Phyllodes Tumour (PT), Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), Papillary Carcinoma (PC).

Fuzzy c-means uncertain images									
Emb.size	A	F	TA	PT	DC	LC	MC	PC	Total
2	1	8	2	1	14	2	5	6	39
3	12	6	10	5	9	5	4	7	58
4	7	4	9	5	32	4	5	6	72
8	23	13	4	3	25	4	7	15	94
16	12	12	17	8	9	3	10	11	82
32	24	14	8	14	33	3	16	12	124
64	11	23	7	17	44	5	11	10	128
128	21	27	28	21	26	7	8	9	147
256	35	27	35	16	30	6	9	9	167

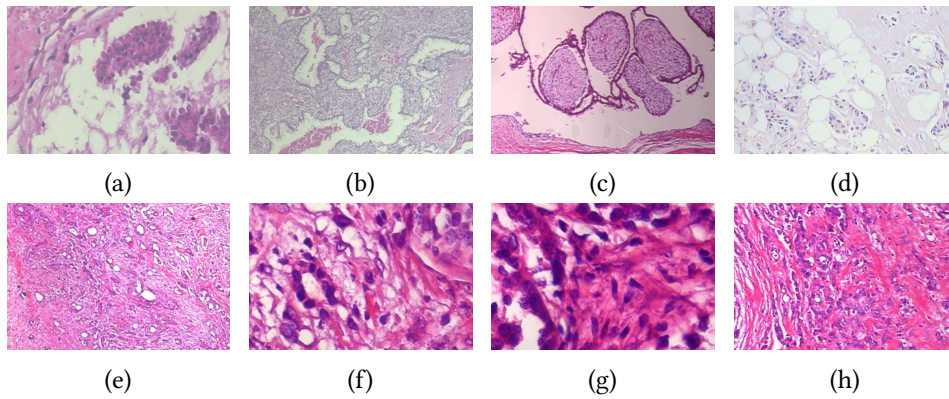


Figure 4: Some misclassified images: image (a) is an example of Adenosis, images (b) and (c) are from class Fibroadenoma, image (d) is from class Tubular Adenoma, and images from (e) to (h) are from Ductal Carcinoma class.

4. Conclusions

The adoption of fuzzy clustering in the medical field could be very useful for providing effective decision support systems. The outcomes of a fuzzy clustering provide continuous value membership that could be more informative and useful for a final decision, particularly in the case of histopathological image classification. In this work, we have proposed a triplet network deep learning model to derive effective embeddings for fuzzy clustering. The images projected in these embeddings lead to a better clusterization even at a very low dimension. This is a very important result due to the relevance of the data dimension in every supervised and unsupervised classification (e.g. curse of dimensionality and computational issues). In the near future, we plan to study how different loss functions used for the metric learning can affect the final classification. Moreover, we want also to investigate the sensitivity of the embeddings to

the used clustering distance, with the final goal of providing a decision support system tool for the physicians in the specific problem of histopathological image classification.

References

- [1] N. Dimitriou, O. Arandjelović, P. D. Caie, Deep Learning for Whole Slide Image Analysis: An Overview, *Frontiers in Medicine* 6 (2019) 264. doi:10.3389/fmed.2019.00264.
- [2] G. Campanella, M. G. Hanna, L. Geneslaw, A. Mirafior, W. K. S. et al., Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, *Nature Medicine* 25 (2019) 1301–1309.
- [3] F. A. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte, A Dataset for Breast Cancer Histopathological Image Classification, *IEEE Transactions on Biomedical Engineering* 63 (2016).
- [4] K. Musgrave, S. Belongie, S.-N. Lim, A metric learning reality check, in: *European Conference on Computer Vision*, Springer, 2020, pp. 681–699.
- [5] M. Kaya, H. Bilge, Deep metric learning: A survey, *Symmetry* 11 (2019).
- [6] X. Wang, X. Han, W. Huang, D. Dong, M. R. Scott, Multi-similarity loss with general pair weighting for deep metric learning, 2020. arXiv:1904.06627.
- [7] K. Musgrave, S. Belongie, S.-N. Lim, Pytorch metric learning, arXiv preprint arXiv:2008.09164 (2020).
- [8] Y. Benhammou, B. Achchab, F. Herrera, S. Tabik, BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights, *Neurocomputing* 375 (2020) 9–24.
- [9] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah, Signature verification using a "siamese" time delay neural network, in: *Proc. of the Int Conf on Neural Information Processing Systems, NIPS'93*, Morgan Kaufmann Publishers Inc., 1993, p. 737–744.
- [10] E. Hoffer, N. Ailon, Deep metric learning using triplet network, in: A. Feragen, M. Pelillo, M. Loog (Eds.), *Similarity-Based Pattern Recognition*, Springer International Publishing, Cham, 2015, pp. 84–92.
- [11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.
- [12] L. McInnes, J. Healy, Umap: Uniform manifold approximation and projection for dimension reduction, *ArXiv abs/1802.03426* (2018).
- [13] J. C. Bezdek, R. Ehrlich, W. Full, Fcm: The fuzzy c-means clustering algorithm, *Computers Geosciences* 10 (1984) 191–203. doi:https://doi.org/10.1016/0098-3004(84)90020-7.
- [14] J. C. Bezdek, Numerical taxonomy with fuzzy sets, *Journal of mathematical biology* 1 (1974) 57–71.
- [15] J. C. Bezdek, Cluster validity with fuzzy sets (1973).
- [16] R. N. Dave, Validating fuzzy partitions obtained through c-shells clustering, *Pattern recognition letters* 17 (1996) 613–623.