HAL open science

Curriculum Incremental Deep Learning on BreakHis DataSet

Mona Sabrine Mayouf, Florence Dupin de Saint-Cyr

To cite this version:

Mona Sabrine Mayouf, Florence Dupin de Saint-Cyr. Curriculum Incremental Deep Learning on BreakHis DataSet. 8th International Conference on Computer Technology Applications (ICCTA 2022), May 2022, Vienna, Austria. pp.35–41, 10.1145/3543712.3543747 . hal-03649718v2

HAL Id: hal-03649718

https://hal.science/hal-03649718v2

Submitted on 3 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Curriculum Incremental Deep Learning on BreakHis DataSet

Mona Sabrine Mayouf
mouna-sabrine.mayouf@irit.fr
IRIT, Université de Toulouse
Toulouse, France

Florence Dupin De Saint-Cyr
florence.bannay@irit.fr
IRIT, Université de Toulouse
Toulouse, France

ABSTRACT

This paper examines methodological aspects of the training procedure of neural networks for medical image classification. The research question concerns the conjecture that: *feeding a network with datasets of increasing magnification leverages high-level knowledge and helps the network to better classify.* This study confirms this hypothesis by an experiment carried out on a dataset of breast cancer histopathological images. Results are presented that underline the importance of the order in which data is introduced to the neural network during the training phase. Extensive experiments done on the BreakHis dataset demonstrate that curriculum incremental learning reaches 98.76% accuracy for binary classification while the best state of the art approach only reaches 96.78.%. Concerning multi-class classification, curriculum incremental learning reaches 95.93% while the state of the art approaches only reaches 95.49%. Moreover both the computational time and the stabilization time of the learning process of the incremental curriculum learning approach are reduced (respectively by 6% and by more than 20%) wrt a non curriculum learning approach.

KEYWORDS

breast cancer, deep learning, histopathological images, learning process.

ACM Reference Format:
Mona Sabrine Mayouf and Florence Dupin De Saint-Cyr. 2022. Curriculum Incremental Deep Learning on BreakHis DataSet . In *2022 8th International Conference on Computer Technology Applications (ICCTA 2022), May 12–14, 2022, Vienna, Austria.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3543712.3543747

1 INTRODUCTION

According to the World Health Organization (WHO), breast cancer is the most common and dangerous cancer types among women with the highest mortality rate [6]. The process of tumour diagnosis begins with the extraction of a biopsy from the suspected area of the breast, then this organ fragment, called a histological specimen, is cut into thin slides, these slides are histochemically colored and are observed under the microscope producing histological images. The diagnosis made by the pathologist consists of the interpretation of the histological image to decide on the characteristics of the tumour. Computer aided diagnosis has been a main research topic over the last years with the aim to improve the diagnosis accuracy and to support the pathologist in his decisive responsibility. In more than 40 years of research, many studies were conducted on automatic breast cancer image classification. Neural networks have revolutionized medical imaging field by their high performance outperforming the traditional image processing approaches. BreakHis dataset is one of the biggest and most used dataset of histopathological images for breast cancer detection. Considering the performances of CNN approaches on this dataset (see Table 1), we observe that:

The lower the magnification, the higher the accuracy.

From this observation, we propose to study the importance of incremental magnifications of images in the learning phase of a CNN. Indeed, some pathologists have testified that the global information present in the lowest magnification is crucial in order to identify the malignancy, compared to the information present in higher magnifications. Moreover, the protocol used by doctors is focusing first on global views in order to rule out images that are already easy to classify [2, 19]. This leads us to study the behavior of an automatic classification approach based on a guided learning from the lowest magnification to the highest one. This kind of learning is named *curriculum incremental learning* after Bengio et al. [5]. The initial idea first introduced by Elman [12] was to perform a guided learning of a language by starting to train a RNN with simple sentences first and progressively introduce more and more complex ones. In the work of Bengio et al., *curriculum learning* was used to conduct several experiments in machine learning: supervised classification with a SVM of randomly generated points in the presence of more and more noise, supervised shape classification by first learning regular shapes (circle, square and equilateral triangle) then not necessarily regular ones (elipsis, rectangle, triangle), next word prediction after a sequence by first learning sentences concerning only a restricted vocabulary of 5000 words and stepping on towards sentences concerned with more and more vocabulary by adding 5000 words more each time. All the experiments confirm that curriculum learning performs better than non curriculum (see Section 2.2 for more details). Some CNN approaches have also used curriculum learning namely [14, 15] (who defined curriculum learning on more and more "difficult" items, meaning difficult to classify by a given CNN) however, as far as we know, there were no study on the impact of magnification, i.e., extrinsic notion of difficulty, in a curriculum training.

More precisely, the scientific question addressed in this paper is *whether curriculum incremental learning where the incrementation is concerning magnification of images enhances the recognition rate.* For this purpose, we describe a set of experiments that take into account the ordering of the training phase. The training starts with the VGG19 neural network and then learns sets of images of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICCTA 2022, May 12–14, 2022, Vienna, Austria
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9622-6/22/05…$15.00
https://doi.org/10.1145/3543712.3543747

different magnification with their labels. This sets of images are issued from the BreakHis dataset, the CNN is fed with a sequence of patches of images ordered by magnification either in ascending, descending, or random order.

The main *originality* of this approach is that, to the best of our knowledge, no studies have been done that exploit the inherent hierarchical ordering of image magnifications, and the way these are fed to a CNN. Indeed this experiment is done only on one dataset (BreakHis) but the result is however interesting since it confirms the results obtained with curriculum learning on other datasets and with other incremental criteria.

2 STATE OF THE ART

This state of the art starts by introducing the BreakHis dataset and the works dealing with it, then we introduce the studies using the curricuum learning.

2.1 Deep Learning approaches on BreakHis

BreakHis [35] stands for "Breast Cancer Histopathological Images", it is a public dataset of histopathological images of breast cancer. The current version of BreakHis is composed of 7909 histopathological biopsy images collected from 82 patients by P&D Laboratory in Brazil. BreakHis contains several images observed by different microscopic magnifications: 40X, 100X, 200X and 400X (Figure 1 illustrates the four magnifications used in BreakHis dataset [35]).

Table 1 illustrates BreakHis composition. The images are divided into benign and malignant tumors that are themselves divided into eight different breast tumor subtypes (benign tumors have four subtypes: Adenosis (A), Fibro Adenoma (F), Tubular Adenoma (TA) and Phyllodes Tumor (PT); malignant ones have four other subtypes: Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC) and Papillary Carcinoma (PC)) since tumor subtypes are also very important for the treatment phase and the prognostic estimation in the clinical procedure. Binary classification aims at classifying the BreakHis images into benign and malignant, while multi-class classification aims at classifying the images into one of the eight tumor subtypes.

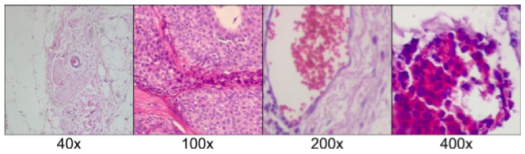
In this article, we focus only on CNN-based researches that use the BreakHis dataset; accuracy results of the different approaches are summarized in Table 1. The different approaches presented in Table 1 are all based on transfer learning from famous "teacher" networks (AlexNet, DenseNet, Inception, ResNet, VGG, etc.). The accuracy rates obtained by these works (around 90%), demonstrate the efficiency of CNNs in breast cancer histopathology classification. For almost all approaches, we observe that *the lower the magnification the higher the accuracy.*

2.2 Curriculum learning approaches

According to Bengio [5], experiments showed that humans and animals learn better when the examples are presented in a suitable order. The idea of using this strategy in machine learning can be tracked back to [12] who trained a network to process complex sentences. [18] has also used this kind of guided learning to train a recurrent neural network to predict the next word.Similar ideas were also explored in robotics [31], by gradually raising the task difficulty. Bengio et al. [5] have introduced the term of "Curriculum learning" to define this process of learning gradually from simple examples towards more and more difficult ones. Several experiments are presented in this article, the first one is about training a support vector machine (SVM) on a dataset made of randomly generated points in the presence of more and more noise, in another experiment, a Perceptron is trained to classify generated pairs (x, y) where y is function of x with more or less additional pairs (x', y') where x' is irrelevant to y' (the more irrelevant pairs the more difficult the learning task). They analyzed three training strategies: curriculum (the examples are ordered by increasing difficulty), anti-curriculum (the examples are ordered by decreasing difficulty) and lastly the examples were ordered randomly. The results show that the curriculum strategy generates less test errors. A last experiment concerns shapes classification is using two datasets "BasicShapes" and "GeomShapes" that contains respectively regular shapes images (circles, squares and equilateral triangles) and not necessarily regular shapes images (ellipses, rectangles, triangles). The first dataset being less variable than the second one is considered "easier to learn". The curriculum consists in a 2-step schedule: first perform the gradient descent with the BasicShapes then perform it with the GeomShapes training sets. The best generalization is obtained when the model spends the half of its training epochs on the easiest examples then moves to the hardest examples of "GeomShapes". The last experiment was about training a recurrent neural network to predict the most suitable word that will occur after a sequence of words. The results show that learning gradually with sentences that are restricted to a limited vocabulary of 5000 words and progressively integrating sentences using more and more vocabulary (increasing the vocabulary by steps of 5000 words) gives better results than feeding the network with all the sentences of the training set directly from the beginning.

Inspired by theses results, several works have adopted the curriculum strategy for their network training. In [15], curriculum learning is defined by increasing progressively the difficulty of the samples in each training batch. They propose two definitions of the *difficulty* associated to a sample: 1) the number of neurons activated at the penultimate layer 2) the confidence degree of the classification (the lower the confidence the most difficult the sample). In [14], a neural network is also used to estimate the difficulty of a given task wrt the neural network "learning progress" and this indicator is then used to select the next samples to learn. In our approach, we define the difficulty of a sample by extra-information, namely the magnification (meaning that a more detailed image is considered as more difficult to learn), moreover the curriculum is done by training with more and more difficult sets of patches instead of increasing the difficulty inside each patch.

The expression "Incremental" learning is also used in machine learning fields like decision trees [32, 37], decision rules [25], SVM [10], for characterizing the learning task done with data that arrives continuously over time. Note also that the term was used in the context of neural networks research, to allow the neural network to accommodate with new arriving data and also new additional classes. For instance, [8] and [36] are using incremental learning in the context of fuzzy systems (systems where the frontiers of the classes are not clear-cut, since some items may be associated with different classes as time goes by). It has also been used by [7] which adapts the network with new heterogeneous remote



Classes	Sub-classes	number of samples per Magnif. factors				Total
		40X	100X	200X	400X	
Benign	A	114	113	111	106	444
	F	253	260	264	237	1014
	TA	109	121	108	115	453
	PT	149	150	140	130	569
	total	625	644	623	588	2480
Mali-gnant	DC	864	903	896	788	3451
	LC	156	170	163	137	626
	MC	205	222	196	169	792
	PC	145	142	135	138	560
	total	1370	1437	1390	1232	5429
Total		1995	2081	2013	1820	7909

Ref.	Based-architecture	40X	100X	200X	400X	All
Binary classification						
[4]	AlexNet	81.87	83.39	82.56	80.69	NA
[34]	VGG-VD	87.00	86.2	85.20	82.90	NA
[23]	3-Conv CNN	90.4	86.3	83.1	81.3	NA
[11]	AlexNet	90.96	90.46	90.37	89.75	NA
[41]	VGG16	91.28	91.45	88.57	84.58	NA
[3]	5-Conv CNN	96.82	96.96	96.36	95.97	NA
[39]	InceptionV3	96.84	96.76	96.49	94.71	NA
[38]	BiCNN	97.89	97.64	97.56	97.97	NA
[39]	InceptionResNetv2	97.90	96.88	96.98	96.98	NA
[1]	IRRCNN	97.95	97.57	97.32	97.36	NA
[20]	DenseNet21-AnoGAN	99.13	96.39	86.38	85.20	NA
[24]	VGG16	NA	NA	NA	NA	92.60
[28]	VGG16	NA	NA	NA	NA	94.40
[27]	ResNetV1	NA	NA	NA	NA	98.70
Multi-class classification						
[39]	InceptionV3	90.28	85.35	83.99	82.08	NA
[39]	InceptionResNetv2	92.07	88.06	87.62	84.50	NA
[16]	CSDCNN	92.8	93.9	93.7	92.9	NA
[27]	ResNetV1	NA	NA	NA	NA	99.2

Figure 1: Top left: Images from BreakHis DataSet with different magnifications. Bottom left: BreakHis dataset composition. Right: State of the art results for binary classification ("NA" meaning "Not available" in the cited articles)

sensing data and with the possibility to have evolving classes. [9] are addressing the problem of incremental learning that occurs when new data concerns new classes that were not present before, which could bring bad performances to the network. For all these works, the "incremental" aspect aims at enabling the network to integrate new data coming from the changing environment and to discover new classes, while in our approach the incremental aspect consists in giving the data to the network in a guided way where the data is divided into different sets that are introduced sequentially according to this division. The division of the data is a guide for the learning network because it is done thanks to high level external knowledge, namely the magnification level.

3 INCREMENTAL CURRICULUM LEARNING

In order to demonstrate our hypothesis, we used the proposed curriculum training strategy to enhance the classification of breast cancer histopathological images in order to identify the tumor type and subtype.

3.1 CNN Model and computational parameters

Before introducing our architecture, let us recall that *transfer learning* has been recognized as one of the most fruitful techniques in deep learning [13], it consists in re-training for a new task an existing network that had been successfully implemented for another task, it requires also to fine-tune the original network. Nevertheless, it should be noted that the choice of the most suitable model in addition to the fine-tuning of the model's parameters is very challenging.

In this work, after several attempts, we have chosen the VGG19 model [33]. VGG19 is a pre-trained convolutional network model, with 47 layers. There are 19 layers with learnable weights: 16 convolutional layers and 3 fully connected layers, with a total of 144M parameters, see Figure 2. VGG19 was trained for many iterations on millions of images of the ImageNet dataset [17]. This pre-trained network can classify images into 1000 daily objects categories, such as keyboard, mouse, pencil, and many animals. VGG19 was placed second in classification and first in localization in ILSVRC'2014 competition [30]. The reason of our choice is also justified by the great success of this network on various other images classification tasks ([33] is cited more than 70 000 times).

In order to enhance the network training, the following regularization techniques are performed: *L2* is used with α set to 0.01. Note that, due to the heterogeneity of the histopathological images and especially when we added chromatic transformations, the data distribution became sparser and we get divergence gradient problems especially with the L2 regularization. To counteract this problem, we decided to standardize the data subtracting the mean and dividing by the standard deviation. The CNN's behaviors were improved thanks to this standardization. *Early stopping* and *dropout* techniques are also used (dropout is used on the two last layers).

The model is implemented with the Keras library, using Tensorflow GPU and deep learning libraries. The model is trained with 4000 epochs with a batch size of 128. Adam optimizer is used with an initial learning rate set to 0.0001. The implementation is done on a computation node of the OSIRIM (Observatory of Systems Information Retrieval and Indexing of Multimedia contents) Platform [29][1]. The sources are available in [21].

3.2 Experimental protocol

We split BreakHis dataset into three parts: 70% for training, 20% for validation and 10% for the final tests (all the sets are also distributed into four sets according to the magnifications). Data preparation is made by performing four kinds of operations: vertical and horizontal flips, color inversion, HSV color transformation. These operations are described in [22], the data preparation enabled us to obtain a dataset which is balanced wrt to the eight subtypes. We first re-implement the best approaches of the recent state of the art, namely InceptionV3 and ResNetV1 (these best results approaches were published by Xie et al. in 2019 [39]) with the hyper-parameters described above. We call these methods *classic*. We also implemented a classic version based on the VGG19 network (described in Section 3.1) called *VGG19* in Table 1. We then

[1]The compute node has the following characteristics: DELL T630 server, 2 Xeon 2640 V4 processors (20 threads), 2 x 400 Go RAID1 SSD disk, 1 x 10 Gb/s network, 4 GPUs Nvidia GTX 1080 TI (3584 cuda cores each, 11 Go RAM).

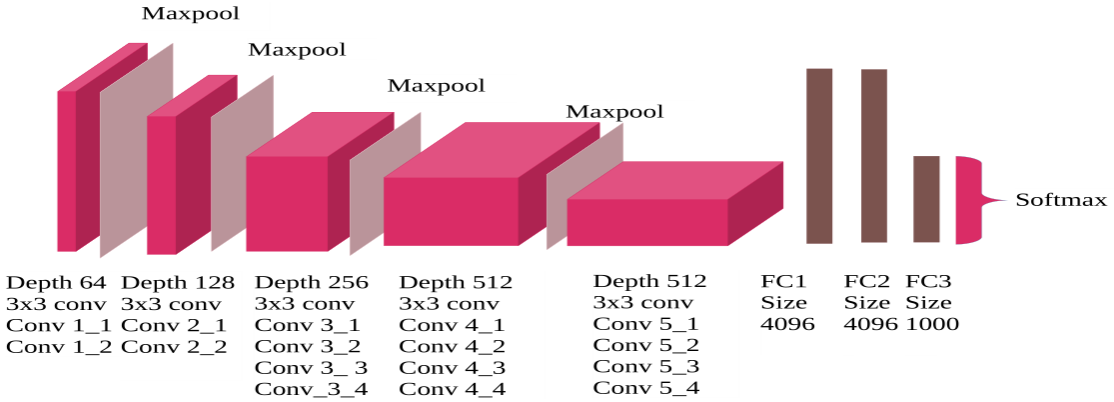


Figure 2: VGG19 architecture [40], `Conv`i_j are convolutional layers with kernels of size 3×3, `FC1` and `FC2` are fully connected layers.

introduced six approaches for studying the magnification impact on the learning process:

- The INCR (for incremental) approach aims at training the model on the grounds of the weights obtained successively for the 40X, 100X, 200X and 400X magnifications as illustrated on Figure 3. More precisely INCR trains the network on 40X images then uses the latest weights as starting point to train on the 100X images and so on for other images) starting from VGG19 weights.
- The Ctrl_INCR (for control incremental) approach consists in training VGG19 independently with the same set of images as the incremental approaches but not incrementally (first train and test on 40X images then train on a *mixed* set composed of 40X and 100X images and test on 100X, then train on 40X+100X+200X and test on 200X then train on all images and test on the 400X), as shown in Figure 3.
- The DECR (for decremental) approach aims at training and testing successively on 400X, 200X, 100X and 40X magnifications (it is the converse of INCR).
- The Ctrl_DECR approach consists in training VGG19 similarly to the Ctrl_INCR approach but with the respective appropriate sets of images (400X then 200X+400X then 100X+200X+400X then all).
- The DIS (for disordered) approach aims at training successively on 200X, 40X, 400X, 100X.
- The Ctrl_DIS approach consists in training VGG19 similarly to the two others control approaches with the successive sets (200X, 40X + 200X, 40X + 200X + 400X, AllX).

4 RESULTS AND DISCUSSION

All the results of the experiments are shown in Table 1. The main conclusions that follow from these results are that 1) the hypothesis is verified and 2) the best approach is INCR.

Accuracy results. First, the experiments performed on the three classic approaches (VGG19, InceptionV3 and ResNetV1), where we trained then tested each magnification separately for both binary and multi-class classification, confirm that the lower the magnification the higher the accuracy.

Second, in most cases VGG19 has a better accuracy than InceptionV3 and ResNetV1 on BreakHis images. Moreover the 6 new approaches implemented are more accurate than VGG19 (thanks to the benefit of transfer learning from a more suitable field). Lastly, we observe that among the new approaches, the INCR approach, that sequentially feeds images with a larger zoom factor to the network, gives more accurate results with a general test accuracy of 98.76% (enhancing the VGG19 classic approach, and improving the state of the art), it shows also a better behavior than the Ctrl_INCR approach (that trains on the same amount of images but in a random way) and than the DECR and DIS ones. This is summarized by the following rank ordering that holds for Binary and Multi-classification:

$$\text{ResNetV1} <_{All} \text{IncepV3} <_{All} \text{VGG19} =_{All} \text{Ctrl_}_{*} <_{All} \text{DECR} <_{All} \text{DIS} <_{All} \text{INCR}$$

where $<_{All}$ means more accurate in the test of images of any magnification (last column of Table 1).

In order to support these results, we used the Layer Wise Relevant Propagation (LRP) technique [26] which selects the pixels (forming a *saliency-map*) that are the most relevant for guiding the classification task. The LRP approach exploits the CNN structure and weights in order to discover this map. We use LRP on several wrongly classified images from the 400X test with CTRL_INCR but correctly classified with INCR, e.g. the image `SOB_M_DC-14-5694400-014` shown on Figure 4. This figure shows the initial image on the left, on the center its LRP saliency map obtained with the control protocol and on the right, the one obtained with the incremental protocol. We can see that the incremental protocol activates more regions of interests for the classification

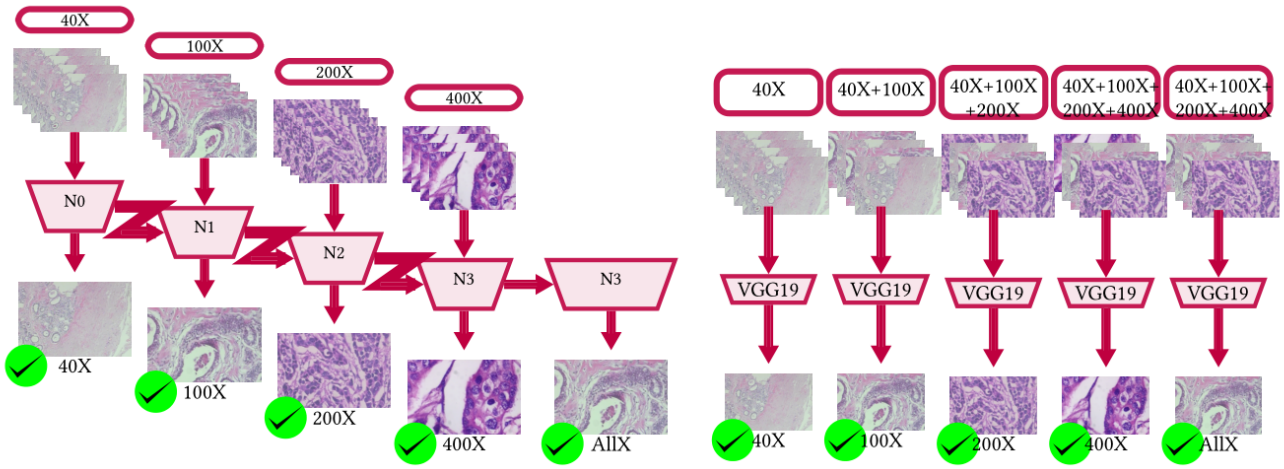


Figure 3: The INCR approach (on the left) and the CTRL_INCR approach (on the right)

Table 1: Computational training time in (minutes:seconds) (right) and Accuracy Results (left). The approaches are in lines, the datasets are in columns, the training is done with 4000 epochs for each magnification set and 16000 epochs for the whole training set.

	40X	100X	200X	400X	All
classicapproaches: Binary classification					
VGG19	95.90	95.35	**95.30**	**94.98**	**97.05**
InceptionV3	95.68	95.52	94.78	94.22	96.95
ResNetV1	**96.20**	**95.68**	94.25	93.08	96.78
Incremental approaches: Binary classification					
CTRL_incr	95.90	95.96	95.52	96.89	97.05
INCR	95.90	96.23	**97.01**	**98.86**	**98.76**
CTRL_decr	96.82	96.97	95.41	94.98	97.05
DECR	**96.91**	96.06	95.43	94.98	97.38
CTRL_dis	96.17	**97.25**	95.30	96.98	97.05
DIS	96.33	97.06	95.30	97.03	97.11
classicapproaches: Multi classification					
VGG19	94.06	93.63	93.80	93.25	**95.49**
InceptionV3	**95.12**	**94.33**	**93.57**	93.15	95.37
ResNetV1	93.87	93.25	92.89	**93.52**	94.76
Incremental approaches: Multi classification					
CTRL_incr	94.06	94.17	94.71	95.62	95.49
INCR	94.06	95.35	**95.68**	**96.26**	**95.93**
CTRL_decr	95.68	94.91	94.22	93.25	95.49
DECR	**96.03**	94.54	94.40	93.25	95.67
CTRL_dis	95.18	96.07	93.80	95.32	95.49
DIS	95.23	**96.13**	93.80	95.78	95.71

	40X	100X	200X	400X	All
classicapproaches: Binary classification					
VGG19	07:23	08:05	07:56	07:02	30:18
InceptionV3	08:49	8:52	09:06	08:12	34:12
ResNetV1	09:11	09:40	09:33	08:57	38:03
Incremental approaches: Binary classification					
CTRL_incr	07:23	15:20	22:07	30:18	30:18
INCR	07:23	14:22	20:25	28:09	**28:09**
CTRL_decr	07:02	15:38	23:58	30:18	30:18
DECR	07:02	15:27	24:21	31:09	31:09
CTRL_dis	07:56	15:34	23:13	30:18	30:18
DIS	07:56	15:49	22:58	32:05	32:05
classicapproaches: Multi classification					
VGG19	7.29	08:11	08:02	07:05	31:02
InceptionV3	08:55	09:13	9:15	8:24	35:28
ResNetV1	9:19	09:52	09:47	09:20	40:05
Incremental approaches: Multi classification					
CTRL_incr	7.29	15:33	23:02	31:02	31:02
INCR	7.29	15:12	22:05	29:57	**29:57**
CTRL_decr	07:15	15:46	24:02	31:02	31:02
DECR	07:15	15:39	24:36	31:37	31:37
CTRL_dis	08:02	15:34	23:13	31:02	31:02
DIS	08:02	16:23	23:02	32:18	32:18

task than the image obtained by the control protocol. This experiment reinforces the idea that the network is learning better with the curriculum strategy.

Let us recall that a fair comparison with state of the art approaches is impossible, since as mentioned above, most of the state of the art proposals did not precise neither their data preparation nor the parameters of their network, nor the size and nature of the test set. Besides, none of the studies cited in Table 1 made experiments covering both the classification on the datasets of the four different magnifications separately and the classification on the whole dataset. This is why we re-implemented the two architectures that gave the best results in [27], namely InceptionV3 and ResNetV1. The accuracy rate obtained in table 1 demonstrates that our choice of VGG19 network and our incremental curriculum strategy give better results than the two other architectures. Note that Figure 1 shows that [27] obtained an accuracy rate of 99.2% in the multi-class classification of an image of any magnification. However, in this work, an additional external dataset has been added to supplement the lack of data.

Computational time: The right table of Table 1 represents the computational time taken by the training process of the different protocols. Note that the training does not stop directly after stabilization since every protocol is systematically trained on a fixed number of epochs (4000 epochs for each magnification training set and 16000 epochs for the whole training set) for sake of fair comparison. The average time taken for the training phases with the whole dataset

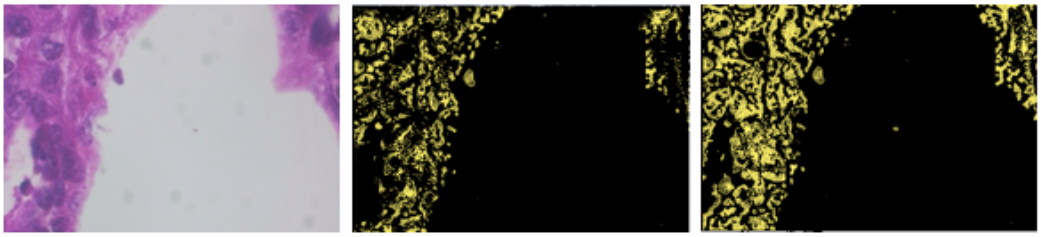


Figure 4: An Image from BreakHis (left), its associated saliency maps obtained with CTRL_INCR (center) and INCR (right)

(column All) is around 30 minutes and 59 seconds for the binary classification and around 31 minutes and 54 seconds for the multi-class classification. We remark that for the classic approaches of the literature where each dataset is learned independently, VGG19 is the fastest network for both binary and multi-class classification for every magnification, namely the training time for the whole set is 30mn18s for the binary classification and 31mn02s for multi-class classification. This can be explained by the fact that InceptionV3 and ResNetV1 are deeper networks (with respectively 48 and 159 layers compared to 19 layers for VGG19) with complex intermingled blocks (contrarily to VGG19 which is sequentially organized).

Concerning the experiments done on curriculum learning, we can see that in the binary classification, the incremental approach gains *124 seconds* (which represents 6% of the training time) wrt the control approach and *65 seconds* in the multi-class classification. Moreover the incremental approach reduces the training time compared with the decremental approach and the disordered approach.

Number of epochs for training stabilization: To confirm the claim of this article that INCR is the best approach, we compare also the number of epochs for training stabilization. The stabilization epoch is the epoch from which the weights of the network are evolving no more. This epoch is computed thanks to the "Early stopping technique". Due to the fact that the training was done for a fixed number of epochs (hence not stopped directly after stabilization), the stabilization epoch is useful for comparing the networks behaviors. We give here only two remarkable numbers: the stabilization epoch for INCR with the whole dataset is 12313 while the one of DIS approach is 14659 given that these curriculum approaches already require 12000 epochs to train on the whole dataset (see Figure 3, 4000 epochs are required to obtain the network N0 then 4000 for obtaining N1, and so on, hence 12000 epochs are required to obtain the final network N3). We note that only 313 supplementary epochs are necessary to reach stabilization for the INCR approach, while the additional epochs required by DIS are 1659. So the stabilization time of the INCR approach is 5 times faster.

5 CONCLUSION

One main goal of this article was to experiment the use of curriculum learning based on magnification knowledge in the cancer type and sub-type classification. Eighteen experimental protocols have been implemented to highlight these results and explore deeply curriculum learning. The experiment has not only demonstrated the benefits of a guided feeding consisting in *training the network from the lowest magnification images towards the highest magnification* but also showed that our approach has outperformed the state of the art breast cancer images classification results by achieving an accuracy of 98.76% for tumor type detection and 95.93% for tumor subtype classification.

This technique of guided CNN feeding which we proved very useful for histopathological images can be extended to other fields in order to optimize the learning process. Along this study, we have noticed shortcomings in the previous research work: a lack of transparency concerning data preparation, the absence of a common set of benchmarks to standardize the tests for a fair and plausible comparison between the different works. There is also a limitation concerning BreakHis dataset itself since the same area is not available in the different magnifications of the images. Indeed, contrarily to what could be expected from the encoding of the names of the files in BreakHis, there is no link between two images with same "slide-id" (also called "patient id") and same sequential number. For instance the images called `SOB_B_A-14-22549AB-40-007.png` and `SOB_B_A-14-22549AB-100-007.png` are named as if they were magnification X40 and X100 taken from the slide 22549 area 007, but in fact 007 is just a random number, hence it does not imply that the two images cover the same area. This observation gave birth to doubts about the correctness of the existing labeling of BreakHis since all the images of the same slide-id have the same cancer type label. This leads us to believe that the label of some images could have been wrongly inferred (instead of obtained from an expert). This doubt about BreakHis dataset labeling is increased when one considers that some works are using the label of an entire image for labeling its subarea (while a tissue image may contain normal, malignant and benign regions at the same time). As a perspective to this work, we would explore the curriculum learning on other datasets and exploit high level knowledge during the training phase. Finally we wish to emphasize the need for transparency in the preparation of the data and in the definition of the test set.

REFERENCES

[1] Md Zahangir Alom, Chris Yakopcic, Mst Shamima Nasrin, Tarek M Taha, and Vijayan K Asari. 2019. Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *Journal of digital imaging* 32, 4 (2019), 605–617. http://dx.doi.org/10.1007/s10278-019-00182-7.

[2] Jose F Arocha, Dongwen Wang, and Vimla L Patel. 2005. Identifying reasoning strategies in medical decision making: a methodological guide. *J. of biomedical informatics* 38, 2 (2005), 154–171. https://doi.org/10.1016/j.jbi.2005.02.001.

[3] Dalal Bardou, Kun Zhang, and Sayed Mohammad Ahmad. 2018. Classification of breast cancer based on histology images using convolutional neural networks. *IEEE Access* 6 (2018), 24680–24693. https://doi.org/10.1109/ACCESS.2018.2831280.

[4] Neslihan Bayramoglu, Juho Kannala, and Janne Heikkilä. 2016. Deep learning for magnification independent breast cancer histopathology image classification. In *2016 23rd International conference on pattern recognition (ICPR)*. IEEE, 2440–2445. https://doi.org/10.1109/ICPR.2016.7900002.

[5] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*. 41–48. https://doi.org/10.1145/1553374.1553380.

[6] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 68, 6 (2018), 394–424. https://doi.org/10.3322/caac.21492.

[7] Lorenzo Bruzzone and D Fernandez Prieto. 1999. An incremental-learning neural network for the classification of remote-sensing images. *Pattern Recognition Letters* 20, 11-13 (1999), 1241–1248. https://doi.org/10.1016/S0167-8655(99)00091-4.

[8] Gail A Carpenter, Stephen Grossberg, Natalya Markuzon, John H Reynolds, David B Rosen, et al. 1992. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on neural networks* 3, 5 (1992), 698–713. https://doi.org/10.1109/72.159059.

[9] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. 2018. End-to-end incremental learning. In *Proceedings of the European conference on computer vision (ECCV)*. 233–248. https://arxiv.org/abs/1807.09536.

[10] Christopher P Diehl and Gert Cauwenberghs. 2003. SVM incremental learning, adaptation and optimization. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, Vol. 4. IEEE, 2685–2690. https://doi.org/10.1109/IJCNN.2003.1223991.

[11] Baolin Du, Qi Qi, Han Zheng, Yue Huang, and Xinghao Ding. 2018. Breast cancer histopathological image classification via deep active learning and confidence boosting. In *Int. Conf. on Artificial Neural Networks*. Springer, 109–116. https://doi.org/10.1007/978-3-030-01421-6_11.

[12] Jeffrey L Elman. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48, 1 (1993), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4.

[13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press. https://mitpress.mit.edu/books/deep-learning.

[14] Alex Graves, Marc G Bellemare, Jacob Menick, Remi Munos, and Koray Kavukcuoglu. 2017. Automated curriculum learning for neural networks. In *international conference on machine learning*. PMLR, 1311–1320. http://proceedings.mlr.press/v70/graves17a.

[15] Guy Hacohen and Daphna Weinshall. 2019. On the power of curriculum learning in training deep networks. In *International Conference on Machine Learning*. PMLR, 2535–2544. http://proceedings.mlr.press/v97/hacohen19a.html.

[16] Zhongyi Han, Benzheng Wei, Yuanjie Zheng, Yilong Yin, Kejian Li, and Shuo Li. 2017. Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific reports* 7, 1 (2017), 1–10. https://www.nature.com/articles/s41598-017-04075-z.

[17] imagenet [n.d.]. Imagenet. http://www.image-net.org.

[18] Kai A Krueger and Peter Dayan. 2009. Flexible shaping: How learning in small steps helps. *Cognition* 110, 3 (2009), 380–394. https://doi.org/10.1016/j.cognition.2008.11.014.

[19] Roy E. Lee, Long Phi Le, and John Gilbertson. 2013. *Pathology Informatics*. Springer New York, New York, NY, 499–513. https://link.springer.com/content/pdf/10.1007%2F978-1-4614-4800-6_18.pdf.

[20] Rui Man, Ping Yang, and Bowen Xu. 2020. Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screened by Generative Adversarial Networks. *IEEE Access* 8 (2020), 155362–155377. https://doi.org/10.1109/ACCESS.2020.3019327.

[21] Mona Mayouf. 2020. Incremental Learning. https://github.com/monamayouf/incremental_learning.

[22] Mouna Sabrine MAYOUF and Florence Dupin De Saint Cyr. 2021. *On Data-Preparation Efficiency Application on Breast Cancer Classification*. Technical Report. IRIT, Toulouse University. https://hal.archives-ouvertes.fr/hal-03246076.

[23] Rajesh Mehra et al. 2018. Automatic magnification independent classification of breast cancer tissue in histological images using deep convolutional neural network. In *Int. Conf. on Advanced Informatics for Computing Research*. Springer, 772–781. https://doi.org/10.1007/978-981-13-3140-4_69.

[24] Rajesh Mehra et al. 2018. Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express* 4, 4 (2018), 247–254. https://doi.org/10.1016/j.icte.2018.10.007.

[25] Ryszard S Michalski, Igor Mozetic, Jiarong Hong, and Nada Lavrac. 1986. The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. *Proc. AAAI 1986* (1986), 1–041. https://dl.acm.org/doi/10.5555/2887770.2887940.

[26] Grégoire Montavon, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. 2019. Layer-wise relevance propagation: an overview. *Explainable AI: interpreting, explaining and visualizing deep learning* (2019), 193–209. https://doi.org/10.1155/2013/832509.

[27] Nima Habibzadeh Motlagh, Mahboobeh Jannesary, HamidReza Aboulkheyr, Pegah Khosravi, Olivier Elemento, Mehdi Totonchi, and Iman Hajirasouliha. 2018. Breast cancer histopathological image classification: A deep learning approach. *bioRxiv* (2018), 242818. https://doi.org/10.1101/242818.

[28] Abdullah Al Nahid and Yinan Kong. 2018. Histopathological breast-image classification using local and frequency domains by convolutional neural network. *Information* 9, 1 (2018), 19. https://doi.org/10.3390/info9010019.

[29] osirim [n.d.]. OSIRIM (Observatory of Systems Information Retrieval and Indexing of Multimedia contents) platform description. https://osirim.irit.fr/site/.

[30] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *Int. J of computer vision* 115, 3 (2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y.

[31] Terence D Sanger. 1994. Neural network learning control of robot manipulators using gradually increasing task difficulty. *IEEE transactions on Robotics and Automation* 10, 3 (1994), 323–333. https://doi.org/10.1109/70.294207.

[32] Jeffrey C Schlimmer and Douglas Fisher. 1986. A case study of incremental concept induction. In *AAAI*, Vol. 86. 496–501. https://www.aaai.org/Library/AAAI/1986/aaai86-083.php.

[33] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*. https://www.robots.ox.ac.uk/~vgg/publications/2015/Simonyan15/simonyan15.pdf.

[34] Yang Song, Ju Jia Zou, Hang Chang, and Weidong Cai. 2017. Adapting fisher vectors for histopathology image classification. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 600–603. https://doi.org/10.1109/ISBI.2017.7950592.

[35] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte. 2016. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Transactions on Biomedical Engineering* 63, 7 (2016), 1455–1462. https://doi.org/10.1109/TBME.2015.2496264.

[36] Marko Tscherepanow, Marco Kortkamp, and Marc Kammer. 2011. A hierarchical ART network for the stable incremental learning of topological structures and associations from noisy data. *Neural Networks* 24, 8 (2011), 906–916. https://doi.org/10.1016/j.neunet.2011.05.009.

[37] Paul E Utgoff. 1989. Incremental induction of decision trees. *Machine learning* 4, 2 (1989), 161–186. https://doi.org/10.1023/A:1022699900025.

[38] Benzheng Wei, Zhongyi Han, Xueying He, and Yilong Yin. 2017. Deep learning model based breast cancer histopathological image classification. In *2017 IEEE 2nd international conference on cloud computing and big data analysis (ICCCBDA)*. IEEE, 348–353. 10.1109/ICCCBDA.2017.7951937.

[39] Juanying Xie, Ran Liu, Joseph Luttrell IV, and Chaoyang Zhang. 2019. Deep learning based analysis of histopathological images of breast cancer. *Frontiers in genetics* 10 (2019), 80. https://doi.org/10.3389/fgene.2019.00080.

[40] Yufeng Zheng, Clifford Yang, and Alex Merkulov. 2018. Breast cancer screening using convolutional neural network and follow-up digital mammography. In *Computational Imaging III*, Vol. 10669. International Society for Optics and Photonics, 1066905. https://doi.org/10.1117/12.2304564.

[41] Weiming Zhi, Henry Wing Fung Yueng, Zhenghao Chen, Seid Miad Zandavi, Zhicheng Lu, and Yuk Ying Chung. 2017. Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images. In *Int. C. on Neural Information Proc.* Springer, 669–676. https://doi.org/10.1007/978-3-319-70093-9_71.