Breast cancer classification methods for augmented reality microscopes

Robin Heckenauer[1], Jonathan Weber[1], Cédric Wemmert[2],
Michel Hassenforder[1], Pierre-Alain Muller[1], and Germain Forestier[1]

[1] IRIMAS, Université de Haute-Alsace, France
name.surname@uha.fr
[2] ICube, Université de Strasbourg, France
wemmert@unistra.fr

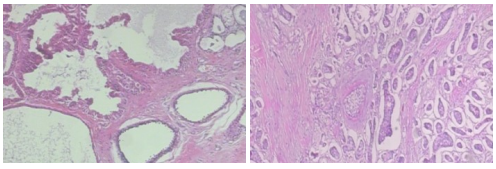
Abstract. An Augmented Reality Microscope (ARM) displays additional information about the tissues being analyzed by a pathologist. The image analysis methods used by these microscopes must be robust to changes in magnification level in order to follow the displacements in the glass slides. In this paper, we propose to take advantage of features present in some key magnification levels that improve the results of other magnification levels. We propose a real-time method robust to changes in magnification levels by training deep neural networks on certain key levels of Whole Slide Images (WSI) and testing them on the levels of a microscope. We show that our approach outperforms or equals naive methods on a breast cancer dataset for the Inception-ResNet-v2 deep learning architecture.

Keywords: Real-time breast cancer classification; Digital pathology; Augmented reality microscope; Deep convolutional networks.

1 Introduction

1.1 Digital pathology

The introduction of slide scanners in the late 1990s enabled the advent of digital pathology [1]. Indeed, scanners have made it possible to digitize glass slides by creating high resolution tissue images (Whole Slide Image - WSI). An example of such images is shown in Figure 1.



(a) **(b)**

Fig. 1. Patches of benign and malignant tumors. Examples of benign tumor (a) and malignant tumor (b) as seen in a WSI at 40X magnification level, in the case of breast cancer (H&E stain).

The digitization of the slides has made it possible to solve many problems related to microscopic study. First, unlike glass slides, it is possible to preserve the tissues indefinitely and without deterioration. Consequently, the archiving of medical images makes it possible to create large databases around particular pathologies, which facilitates research work concerning these diseases. Then it becomes possible to share WSIs for peer review purposes to obtain multiple remote diagnoses of the same glass slide. Alternatively, microscope virtualization and WSI sharing can be used for training purposes. Indeed, to train specialists in pathology, it is no longer necessary to have glass slides, nor even a microscope available. A virtual microscope with WSI is sufficient, thus allowing easy access to education. These technological advances come with new challenges related to the size of WSIs. Because of their high resolution, WSIs weigh several gigabytes. Therefore, appropriate IT systems are needed to store, process and share WSI.

1.2 Digital pathology & AI

Following the encouraging results shown by artificial intelligence (AI) in the field of computer vision [2], these methods were quickly used on medical images. In the field of digital pathology, AIs have the potential to provide

solutions in education, quality control, clinical diagnoses, image analysis [3–5]. However, faced with these new opportunities, the challenges are numerous: the lack of annotated data; high consumption of computing resources; the lack of transparency and explainability of AI predictions; ethical and legal issues.

Among the methods of AI, we find in particular deep learning methods. Deep neural networks were quickly applied to various pathologies, including breast cancer, which is the deadliest for women [6, 7].

Thus, studies have proposed methods for classifying tumor patches in the case of breast cancer [8–12]. These works have shown the superiority of deep learning approaches compared to machine learning approaches.

1.3 Virtual microscope & augmented reality

Driven by the evolution of digital pathology, virtual microscopes have appeared, thus making it possible to emulate the functions of an optical microscope on a computer [13, 14]. In addition to not needing a microscope, virtual microscopes allow tissues to be zoomed in and out without delay, unlike optical microscopes, which only offer a few levels of magnification and require focusing of the tissue observed at each level change.

Despite these advantages, in a context of redundant work such as glass slide diagnosis, using a virtual microscope is not suitable, because it leads to a loss of time during the digitization of the glass slide, then during the diagnosis phase [15]. An approach based on augmented reality tools integrated into microscopes (ARM), where additional information is displayed during the microscopic study seems better suited to help pathologists in their routine.

During a microscopic study carried out with an optical microscope, the pathologist analyzes a glass slide in search of structures of interest to establish a diagnosis. The glass slide is studied in particular at different levels of magnification. To change the level of magnification, optical microscopes have a limited number of objectives. We believe that the intermediate magnification levels unattainable with a light microscope are rich in features and can provide insight into tissue at the magnification levels accessible by a microscope.

Following previous work [16] on real-time object detection in WSIs, we propose to:

- Classify in real-time patches of benign and malignant tumors in the case of breast cancer using the Inception-ResNet-v2 [17] architecture.
- Use the intermediate magnification levels of a WSI to improve the performance of an ARM method that uses standard magnification levels of an optical microscope (20X, 40X, 100X).
- Compare three tumor prediction approaches at different magnification levels: Train networks at multiple levels; Train networks on each level; Train networks on an intermediate level.

For the sake of reproducibility, the experiments were performed on the public dataset BreakHis [18], and the source code is available here [3].

The paper is organized as follows. First, we introduce three methods to process images of multiple magnification levels in the case of an ARM application in section 2.1. Then, we introduce the public dataset BreakHis [18] and datasets created from BreakHis in section 2.2. We compare the methods presented in 2.1 in section 3, before providing an analysis on the experiments performed in section 4.

2 Materials and methods

2.1 Methods

In recent years, the rapid evolution of the computational capabilities of graphics processing units (GPUs) has largely contributed to the emergence of new approaches that improve the overall performance of deep learning methods, allowing them to be considered more seriously for real-time applications such as augmented reality microscopy [19, 20]. One of the challenges of using the augmented reality microscope (ARM) is the frequent changes in magnification level during diagnosis sessions. Indeed, changing the magnification alters the structures, shapes and borders of the observed tissues. To overcome this problem, we consider two naive approaches to develop a robust real-time method that performs well at different magnifications.

The first approach consists of training a single network on several magnifications, so that it generalizes its understanding of structures at several levels. This method requires only one model to be trained, however, the performance is largely related to the ability of the architecture to generalize over multiple magnifications. The second approach consists in training one network per level. In this case, several models are required, depending on the range of magnification levels studied. We then propose a third approach where we train a network on a level near to the level used for the test phase.

[3] https://github.com/RobinHCK/Breast-cancer-classification-methods-for-augmented-reality-microscopes

Among the architectures that have shown the best results on the task of classifying benign and malignant tumor patches in the case of breast cancer are: Inception-v3 [21], ResNet [22], GoogleNet [23], VGG [24]. Inception-v3 seems to show slightly better results than the other methods although the difference may be minimal depending on the dataset used. The Inception-ResNet-v2 architecture, which is an evolution of the Inception-v3 architecture, proposes to add residual connections to the Inception modules in order to speed up training and outperform Inception architectures without residual connections [17].

We compare the three approaches by studying the results of different Inception-ResNet-v2 models, over a wide range of magnification levels to be able to select the best networks over a given range.

2.2 Datasets

BreakHis The public dataset BreakHis [18] was used because it contains many images at different magnification levels. It consists of 7,909 patches stained in H&E (Hematoxylin and Eosin), size 700x460 pixels, at 40X, 100X, 200X and 400X magnification levels. This dataset is destined for the breast cancer classification task. 2,480 benign and 5,429 malignant tumor patches were collected from 82 patients.

Creation of intermediate magnification levels To cover a wider range of magnification levels, we interpolate patches of 20 intermediate magnification levels from the original dataset.

To do this, each original patch is cropped according to Equation 1 where *crop size* is the size of the patch to be kept when cropping, *patch size* is the size of the patch that will be used by the deep neural network, and *nearest level* is the nearest lower magnification level that exists in the patches of the original dataset.

$$crop\ size = \frac{patch\ size\ *\ nearest\ level}{desired\ level} \tag{1}$$

Next, we choose bicubic interpolation [25], which is a good compromise between interpolation quality and processing time [26], to size the patches to the dimension *patch size*. Thus, we obtain a dataset containing patches of size 350x230 pixels[4] for each of the 24 magnification levels in our range (from 20X to 400X). We create a last balanced dataset named allX from the 40X, 100X, 200X and 400X patches.

Finally, we apply data augmentation methods to the datasets to improve the diversity of the data and counteract the lack of data that is a recurrent problem when using medical data. Patches are randomly rotated, horizontally flipped, and vertically flipped.

3 Experimentations

In section 2.1, we presented three approaches that are robust to changes in magnification levels. To study and compare these approaches, we train Inception-ResNet-v2 networks on datasets built from BreakHis over a wide range of magnification levels.

Before starting training, we transfer weights from a pre-trained model on ImageNet [27] (ImageNet is the largest image database, commonly used for classification tasks, containing 14,197,122 images organized into 21,841 classes) to our model. We freeze the model weights, before re-training the last layers of the network on our data. This method is a standard procedure as reported in [28] for refining the training of models with little data. This saves training time and reduces overfitting.

We train the networks with the same parameters and hyperparameters to facilitate comparison of results: learning rate = 0.01; momentum = 0.9; epochs = 30; batch size = 64; optimizer = SGD; cost function = softmax; regularization = dropout. In addition, we use EarlyStopping to stop training when no improvement occurs on the validation loss in the last 15 epochs to avoid overfitting. To take full advantage of the data we have and to tackle overfitting, we perform a 5-fold (K=5) cross-validation for each magnification level. Patches are distributed in each fold respecting the organization given by BreakHis. Each fold is used to test once the results obtained on a network trained on all other folds. Thus, each patch of the dataset is classified once.

We perform all the computational processing on GPUs provided by the computing center of the University of Strasbourg. In total, at a rate of 10 minutes of training per network, it took about 1 day of cumulative computation to train the 125 networks (cross-validation K=5x25 datasets) that this experiment counts.

Finally, we test each model on all magnification levels. For this step, we choose to use the accuracy defined in Equation 2 as the evaluation metric since our classes are balanced.

[4] The size of the patches in the BreakHis dataset is 700x460 pixels. In order to obtain patches with a lower magnification level, the original patches must be cropped and resized. Thus we lose some of the context of the original patch

$$accuracy = \frac{well\ classified\ patches}{total\ number\ of\ patches} \tag{2}$$

We build a heatmap illustrated in Figure 2 from the accuracies obtained by the networks.

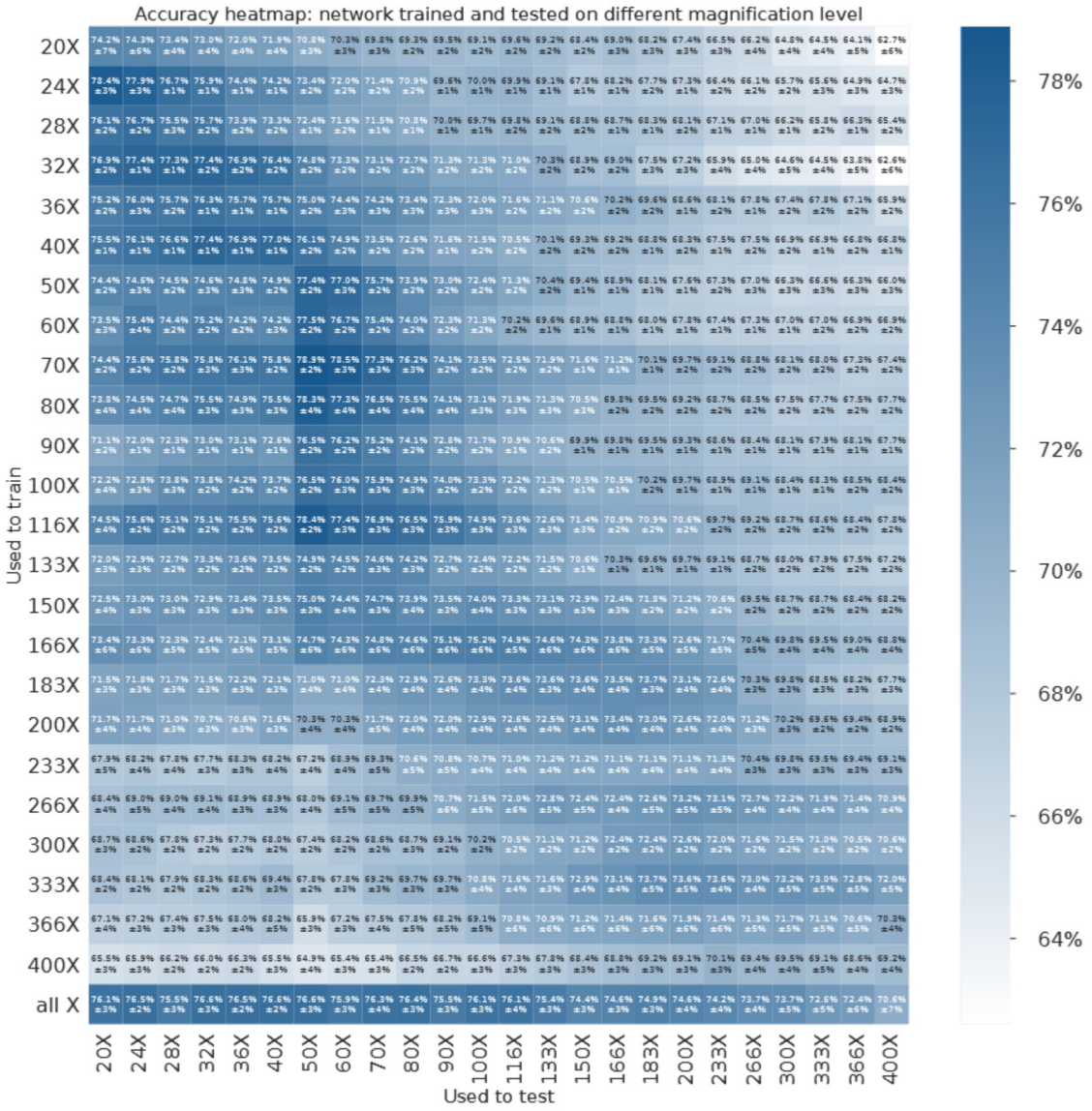


Fig. 2. Inception-ResNet-v2 Accuracy Heatmap. Shows the accuracies (%) obtained by individually tested trained networks on all datasets in a range of magnification levels from 20X to 400X. The higher the accuracy of a network on a dataset, the darker the box, and vice versa.

In parallel, we measure the number of patches processed per second during the testing step on a subsample of 10 networks. Then, assuming that this calculation is linear, we convert this number of patches per second into the number of 720p and 1080p definition frames tested per second (FPS) to obtain the Table 1. These measurements were carried out using a processor (CPU) [5] then two GPU [6].

[5] i5-11600KF 3.90GHz 3.91 GHz
[6] NVIDIA GeForce GTX 1080 & RTX 3080

Table 1. Number of FPS processed by Inception-ResNet-v2. The average number of frames per second (FPS) tested by a network depending on the type of resource used (GPU or CPU) and on the image definition (720p or 1080p).

Image definition	720p	1080p
CPU	1.3	0.6
GPU 1080	3.5	1.7
GPU 3080	7.3	3.2

4 Results and discussion

4.1 Real-time

Displaying augmented reality information on a microscope requires a real-time method capable of tracking a pathologist's movements. We define the notion of real-time based on [29]. Thus, for the tumor patch classification task, we speak of real time to mean that our method is able to predict the field of view observed under a microscope several times per second as shown in Table 1. Indeed, the Inception-ResNet-v2 architecture processes an average of 7.3 FPS at 720p resolution with a GPU, which can classify images quite quickly for a smooth result.

With 55,873,736 parameters and a depth of 572, Inception-ResNet-v2 is one of the computationally intensive networks. Despite its size, it is possible to perform real-time tasks using a single, latest-generation GPU. Thanks to recent advances in computational capacity, it is now possible to run deep learning networks in real time by embedding a commercial GPU on a microscope.

4.2 Heatmap

First, it is clear that networks trained on one magnification level have the best accuracies on nearby magnification levels, as shown in Figure 2. It is obvious that a network will generally be better on a magnification level near to the one used in the training step. However, by looking at the central diagonal, we can see that the best accuracy at one level of magnification is not always found by the trained network on that level of magnification. In other words, a slight zooming in or out during training improves the network's results. It appears that near context contains rich features that improve the results of a network. The results also show that at a given magnification level, networks trained at that level and nearby levels perform best. Moreover, these networks do not necessarily misjudge on the same cases. This means that each level of magnification does not provide the same features.

The second observation concerns the gradient on both sides of the diagonal of the darkest boxes. The further the magnification level of a tested dataset is from the magnification level of the dataset used for training, the lower the accuracy. This seems to confirm the intuition that the near context contains the most useful features for a network.

Third, we note that the higher the magnification level, the worse the accuracies of networks trained on a single magnification level. At higher magnification levels, the spatial context for understanding the structure of a tumor is lost. Therefore, observing tissue at high magnification is not relevant for tumor classification in breast cancer. We note that magnification levels between 20X and 80X are more rich in features for our networks in our case. In practice, this coincides with the behavior of the pathologist, who spends more time in the lower magnification levels.

Fourth, note that the best networks give accuracies close to 80%. This performance is lower than the state of the art in tumor patch classification in breast cancer on BreakHis [30]. This is due to the way we build our datasets from BreakHis. Indeed, in order to study a large range of magnification levels, we build patches of new levels, so part of the context has been removed, which impacts the performance of the networks. We also notice that in the worst case, the lowest accuracy obtained is 62.6%. This means that in the case of breast cancer, the features to distinguish benign and malignant tumors are present at all magnification levels of the study range.

4.3 Comparison of approaches

From the heatmap in Figure 2, we can compare our approach with two naive methods as reported in Table 2.

First, we find that the first approach, where we train the networks on the magnification level used for the test, manages to obtain the best accuracies on some levels in the 20X to 70X range.

Second, the networks trained on all magnification levels achieve best accuracies over the 100X to 300X range. This second approach succeeds in generalizing the learned features over multiple levels, thus obtaining better accuracies over higher magnification levels. However, this method fails to outperform specialized networks on lower magnification levels (20X to 100X).

Table 2. Comparison of approaches. Comparison of the accuracies (%) of three tumor prediction approaches on different magnification levels: Training networks on several levels; Training networks on each level; Training networks on an intermediate level.

Levels of magnification tested	network trained & tested on one level	network trained on 40X, 100X, 200X, 400X (allX)	network trained on an intermediate level (Our method)	The best network
20X	0.742	0.761	**0.784**	24X
24X	**0.779**	0.765	**0.779**	24X
28X	0.755	0.755	**0.773**	32X
32X	**0.774**	0.766	**0.774**	32-40X
36X	0.757	0.765	**0.769**	32-40X
40X	**0.770**	0.766	**0.770**	40X
50X	0.774	0.766	**0.789**	70X
60X	0.767	0.759	**0.785**	70X
70X	**0.773**	0.763	**0.773**	70X
80X	0.755	0.764	**0.765**	116X
90X	0.728	0.755	**0.759**	116X
100X	0.733	**0.761**	0.752	allX
116X	0.736	**0.761**	0.749	allX
133X	0.715	**0.754**	0.746	allX
150X	0.729	**0.744**	0.743	allX
166X	0.738	**0.746**	0.738	allX
183X	0.737	**0.749**	0.737	allX
200X	0.726	**0.746**	0.736	allX
233X	0.713	**0.742**	0.736	allX
266X	0.727	**0.737**	0.730	allX
300X	0.715	**0.737**	0.732	allX
333X	**0.730**	0.726	**0.730**	333X
366X	0.706	0.724	**0.728**	333X
400X	0.692	0.706	**0.720**	333X
Average	0.740	0.751	**0.754**	

Third, we find that the proposed approach, where we select the best networks trained on any magnification level, achieves the best accuracies for the 20X to 90X range. A magnification level contains features that sometimes yield better results on nearby magnification levels.

On average, the proposed approach (75.4%) outperforms or equals the naive approaches (74% & 75.1%) over the range of magnification levels studied with the Inception-ResNet-v2 architecture on the BreakHis dataset. Our approach shows better results on low magnification levels (20X to 90X). From 100X, the allX approach becomes better. Beyond a certain magnification level, the tumor features present in the image become less relevant. However, the features at low magnification levels significantly help the understanding of the features at high magnification levels.

Looking at the best networks by magnification levels shown in Table 2, we observe that the 24X, 32X, 40X, 70X, 116X, 333X, and allX networks are able to cover the entire range of magnification levels studied, while achieving the best results.

In the context of an ARM method, we want to display additional information about the observed tissue at the few magnification levels offered by the optical microscope. Let's take for example a microscope with 20X, 40X and 100X magnification levels. In order to obtain the best results in our case, we must use: the 24X model on the 20X level; the 40X model on the 40X level; the allX model on the 100X level.

It seems interesting to use networks trained on a single level of magnification for the lowest levels of magnification up to 90X. Then, from 100X, it is advisable to use gratings trained on several levels of magnification.

Finally, it should be kept in mind that these results were obtained with the Inception-ResNet-v2 architecture on a breast cancer dataset. Depending on the method and the data used, the conclusions may change.

5 Conclusion

In recent years, the rapid evolution of the computing capabilities of graphics cards has largely contributed to the appearance of new approaches improving the overall performance of deep learning methods. These advances allow

us to more seriously consider deep learning methods for embedded real-time microscope tools. Such applications can be used to assist pathologists in their daily routine.

We have proposed an approach that takes advantage of the intermediate magnification levels of WSIs to improve the performance of an ARM method that uses standard magnification levels of an optical microscope.

Future work should focus on the use of ensemble learning methods to take full advantage of the features of each level of magnification. Finally, it would be interesting to see real-time approaches being implemented on different tasks in various medical and other applications. This will raise awareness of the challenges related to the use of real-time deep learning methods in real conditions.

6 Acknowledgments

The authors would like to acknowledge the High Performance Computing center of the University of Strasbourg for supporting this work by providing scientific support and access to computing resources. Part of the computing resources were funded by the Equipex Equip@Meso project (Programme Investissements d'Avenir) and the CPER Alsacalcul/Big Data.

References

1. Al-Janabi, S., Huisman, A., Van Diest, P.J.: Digital pathology: current status and future perspectives. Histopathology **61**(1) (2012) 1–9
2. Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E.: Deep learning for computer vision: A brief review. Computational intelligence and neuroscience **2018** (2018)
3. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: Challenges and opportunities. Medical image analysis **33** (2016) 170–175
4. Niazi, M.K.K., Parwani, A.V., Gurcan, M.N.: Digital pathology and artificial intelligence. The lancet oncology **20**(5) (2019) e253–e261
5. Tizhoosh, H.R., Pantanowitz, L.: Artificial intelligence and digital pathology: challenges and opportunities. Journal of pathology informatics **9** (2018)
6. Chugh, G., Kumar, S., Singh, N.: Survey on machine learning and deep learning applications in breast cancer diagnosis. Cognitive Computation **13**(6) (2021) 1451–1470
7. Debelee, T.G., Schwenker, F., Ibenthal, A., Yohannes, D.: Survey of deep learning in breast cancer image analysis. Evolving Systems **11**(1) (2020) 143–163
8. Chen, H., Li, C., Li, X., Rahaman, M.M., Hu, W., Li, Y., Liu, W., Sun, C., Sun, H., Huang, X., et al.: Il-mcam: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. Computers in Biology and Medicine **143** (2022) 105265
9. Golatkar, A., Anand, D., Sethi, A.: Classification of breast cancer histology using deep learning. In: International conference image analysis and recognition, Springer (2018) 837–844
10. Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., Li, S.: Breast cancer multi-classification from histopathological images with structured deep learning model. Scientific reports **7**(1) (2017) 1–10
11. Khan, S., Islam, N., Jan, Z., Din, I.U., Rodrigues, J.J.C.: A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition Letters **125** (2019) 1–6
12. Wang, D., Khosla, A., Gargeya, R., Irshad, H., Beck, A.H.: Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718 (2016)
13. Afework, A., Beynon, M.D., Bustamante, F., Cho, S., Demarzo, A., Ferreira, R., Miller, R., Silberman, M., Saltz, J., Sussman, A., et al.: Digital dynamic telepathology–the virtual microscope. In: Proceedings of the AMIA Symposium, American Medical Informatics Association (1998) 912
14. Ferreira, R., Moon, B., Humphries, J., Sussman, A., Saltz, J., Miller, R., Demarzo, A.: The virtual microscope. In: Proceedings of the AMIA Annual Fall Symposium, American Medical Informatics Association (1997) 449
15. Treanor, D., Jordan-Owers, N., Hodrien, J., Wood, J., Quirke, P., Ruddle, R.A.: Virtual reality powerwall versus conventional microscope for viewing pathology slides: an experimental comparison. Histopathology **55**(3) (2009) 294–300
16. Heckenauer, R., Weber, J., Wemmert, C., Feuerhake, F., Hassenforder, M., Muller, P.A., Forestier, G.: Real-time detection of glomeruli in renal pathology. In: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), IEEE (2020) 350–355
17. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence. (2017)
18. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. Ieee transactions on biomedical engineering **63**(7) (2015) 1455–1462
19. Chen, P.H.C., Gadepalli, K., MacDonald, R., Liu, Y., Nagpal, K., Kohlberger, T., Dean, J., Corrado, G.S., Hipp, J.D., Stumpe, M.C.: Microscope 2.0: An augmented reality microscope with real-time artificial intelligence integration. arXiv preprint arXiv:1812.00825 (2018)
20. Razavian, N.: Augmented reality microscopes for cancer histopathology. Nature Medicine **25**(9) (2019) 1334–1336

21. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 2818–2826
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770–778
23. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 1–9
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
25. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing **29**(6) (1981) 1153–1160
26. Parsania, P.S., Virparia, P.V.: A comparative analysis of image interpolation algorithms. International Journal of Advanced Research in Computer and Communication Engineering **5**(1) (2016) 29–34
27. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, Ieee (2009) 248–255
28. Kora, P., Ooi, C.P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W.Y., Meenakshi, K., Swaraja, K., Plawiak, P., Acharya, U.R.: Transfer learning techniques for medical image analysis: A review. Biocybernetics and Biomedical Engineering (2021)
29. Dodhiawala, R.T., Sridharan, N., Raulefs, P., Pickering, C.: Real-time ai systems: A definition and an architecture. In: IJCAI, Citeseer (1989) 256–264
30. Tufail, A.B., Ma, Y.K., Kaabar, M.K., Martínez, F., Junejo, A., Ullah, I., Khan, R.: Deep learning in cancer diagnosis and prognosis prediction: a minireview on challenges, recent trends, and future directions. Computational and Mathematical Methods in Medicine **2021** (2021)