

Automated classification of multiphoton microscopy images of ovarian tissue using deep learning

Mikko J. Huttunen
Abdurahman Hassan
Curtis W. McCloskey
Sijyl Fasih
Jeremy Upham
Barbara C. Vanderhyden
Robert W. Boyd
Sangeeta Murugkar

Automated classification of multiphoton microscopy images of ovarian tissue using deep learning

Mikko J. Huttunen,^{a,b,*} Abdurahman Hassan,^a Curtis W. McCloskey,^{c,d} Sijyl Fasih,^a Jeremy Upham,^a Barbara C. Vanderhyden,^{c,d} Robert W. Boyd,^{a,e} and Sangeeta Murugkar^f

^aUniversity of Ottawa, Department of Physics, Ottawa, Ontario, Canada

^bTampere University of Technology, Laboratory of Photonics, Tampere, Finland

^cUniversity of Ottawa, Department of Cellular and Molecular Medicine, Ottawa, Ontario, Canada

^dOttawa Hospital Research Institute, Centre for Cancer Therapeutics, Ottawa, Ontario, Canada

^eUniversity of Rochester, Institute of Optics, Department of Physics and Astronomy, Rochester, New York, United States

^fCarleton University, Department of Physics, Ottawa, Ontario, Canada

Abstract. Histopathological image analysis of stained tissue slides is routinely used in tumor detection and classification. However, diagnosis requires a highly trained pathologist and can thus be time-consuming, labor-intensive, and potentially risk bias. Here, we demonstrate a potential complementary approach for diagnosis. We show that multiphoton microscopy images from unstained, reproductive tissues can be robustly classified using deep learning techniques. We fine-train four pretrained convolutional neural networks using over 200 murine tissue images based on combined second-harmonic generation and two-photon excitation fluorescence contrast, to classify the tissues either as healthy or associated with high-grade serous carcinoma with over 95% sensitivity and 97% specificity. Our approach shows promise for applications involving automated disease diagnosis. It could also be readily applied to other tissues, diseases, and related classification problems.

© 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JBO.23.6.066002](https://doi.org/10.1117/1.JBO.23.6.066002)]

Keywords: medical and biological imaging; ovarian cancer; optical pathology; tissue characterization; nonlinear microscopy; convolutional neural networks.

Paper 180174R received Mar. 22, 2018; accepted for publication May 31, 2018; published online Jun. 13, 2018.

1 Introduction

Ovarian cancer is the most lethal gynecological malignancy with an estimated 22,280 new cases and 14,240 deaths in 2016 in the United States alone.¹ High-grade serous carcinoma (HGSC) is the most common type of epithelial ovarian cancer accounting for 70% of the cases and associated with a low 5-year survival rate of only 40%.² Due to the lack of effective screening and diagnostic imaging techniques, the disease is normally detected at a late stage after wide-spread dissemination. Furthermore, the existing techniques do not permit the detection of microscopic residual disease at the time of surgery. There is thus an urgent need for developing a high-resolution imaging technique that permits the rapid and automated detection of early and recurrent ovarian cancer from tissue biopsies with high accuracy.

Multiphoton microscopy is a high-resolution optical imaging technique that is becoming an indispensable tool in cancer research and diagnosis.³⁻⁹ In this imaging paradigm, the nonlinear optical signals are generated only at the focal point of the excitation beam, providing intrinsic three-dimensional (3-D) optical sectioning and permitting nondestructive, label-free imaging. In particular, second-harmonic generation (SHG) imaging provides intrinsic contrast to visualize the organization of collagen fibers and elastin, which are major constituents of the extracellular matrix (ECM), the distribution of which can be a key identifier for several diseases.^{8,9} Another example is two-photon excitation fluorescence (TPEF) imaging of intrinsic

tissue fluorescence, which enables the identification of changes in cellular morphology and organization. SHG and TPEF imaging have been utilized to demonstrate that remodeling of the ECM is associated with cancer progression.⁷⁻¹² Wen et al.¹³ implemented two-dimensional texture analysis of SHG images from unstained ovarian tissue to quantify the remodeling of the ECM. Recently, the approach was generalized to 3-D texture analysis and to classify SHG images from six different ovarian tissue types.¹⁴ These studies demonstrate the potential of machine learning-based evaluation of SHG images for improved diagnostic accuracy of ovarian cancer detection.

In machine learning, the computer programs learn to perform data analysis tasks, such as image classification, that are hard to perform algorithmically due to the complexity of the data set. Image classification is often achieved using supervised learning, where the task is learned by using labeled training images. In general, the labeled images are used to learn a more optimal representation of the image data, which facilitates clustering of the images into clearly separated sets and thus enables their classification. Several supervised learning approaches exist for classification tasks; support vector machines (SVMs) and logistic regression are among the most commonly used due to their relative simplicity and performance.¹⁵ However, these classification approaches require extensive image processing and handcrafted feature extraction procedures. In contrast, deep learning is a rapidly growing area of machine learning, in which data are analyzed using multilayered artificial neural

*Address all correspondence to: Mikko J. Huttunen, E-mail: mikkojhuttunen@gmail.com

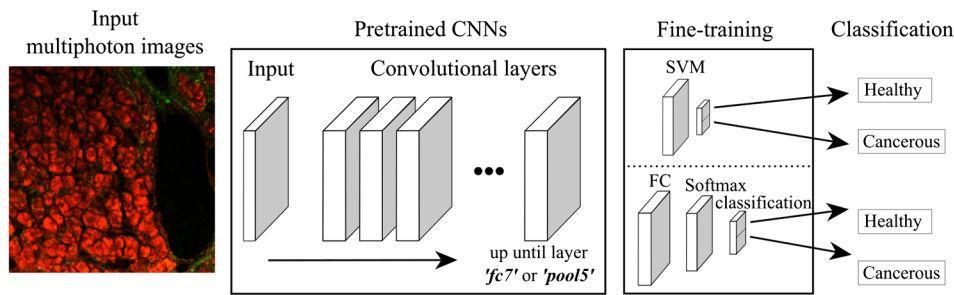


Fig. 1 Schematic of the two transfer learning approaches used in this study for classifying the input multiphoton images either as healthy or cancerous (HGSC). In both cases, the input images are fed to the pretrained CNNs, which transform the data into a more optimal representation enabling robust classification. In the first approach, the output of the pretrained CNN is fed to a trained SVM classifier. In the second approach, the final FC layers of the pretrained CNNs are replaced by new FC layers more suitable for binary classification.

networks that avoid extensive human intervention.¹⁶ In particular, convolutional neural networks (CNNs) have been applied also for classifying images of stained tissue biopsy slides.^{16–22} In these studies, the CNNs have been trained using large amounts of data consisting of millions of images.^{23,24} But so far, the use of CNNs in the classification of multiphoton images has been remarkably limited,²⁵ mainly because of the small size of the typically available data set. However, with the development of deep learning techniques for high-accuracy classification that requires fewer training images, its application to multiphoton image data sets has become more viable, which could lead to rapid and reliable automated diagnostic tools.

In this paper, we demonstrate the use of deep neural networks for robust and real-time classification of multiphoton microscopy images of unstained tissues. We acquire SHG and TPEF images of ovarian and upper reproductive tract tissue from healthy mice and tumor tissue from orthotopic syngeneic HGSC murine models. We construct binary image classifiers (healthy versus HGSC) by fine-tuning pretrained CNNs using a relatively small acquired data set consisting of ~200 multiphoton images. We study the performance of four pretrained CNNs (AlexNet, VGG-16, VGG-19, and GoogLeNet), and examine the role of data augmentation on the results. We demonstrate classification of the acquired images with over 95% sensitivity and 97% specificity. In particular, we show that best classification performance is achieved when the combined TPEF and SHG data are used compared to using only the SHG or TPEF data. The trained classifiers are also shown to outperform more traditional classifiers based on SVMs. Because the demonstrated approach is minimally invasive, operates in real-time, and requires very little sample preparation, it has potential for clinical applications and computer-aided diagnosis.

2 Image Classification Using Pretrained Convolutional Neural Networks

Deep learning and CNNs have recently proved useful for various computer vision tasks.^{16–22} Although several CNNs with different architectures and configurations exist, their overall working principles are similar. The input image is passed through the CNN consisting of different layers, such as convolutional, pooling, activation, and fully connected (FC), where each layer performs specific types of data operations. The layers are made of artificial neurons, which calculate a weighted sum of

the inputs and transform it, often with a bias, to an output using a transfer function. During the training process of the CNN, the weights and biases of the artificial neurons are optimized leading to the desired performance of the network, such as distinguishing between healthy and diseased tissue samples.

In convolutional layers, the input data are convolved using various filters into a more useful representation, which can be used, for example, in feature detection/extraction. The number of sequential convolutional layers, i.e., the depth of the CNN, varies from a few layers to hundreds of layers where the deeper CNNs are computationally more expensive but often outperform shallower ones.^{16,18} Pooling layers downsample the input to reduce its dimensionality. Activation layers, such as rectified linear units, provide nonlinearity to the signal processing allowing faster and more effective training of the network.¹⁶ At the end of the CNN, FC layers are used to compute the output, in our case the binary class scores (healthy versus HGSC) for each input image. Alternatively, the FC layers can be replaced by other classifiers, for example, based on logistic regression or SVMs, which are optimized for the task of classification.²⁶

After the CNN is designed, it needs to be trained for the particular task. For the case of supervised learning this is done by forming a cost function for the network and using it to compare the calculated output of the network with the desired output. The network is then trained by iteratively optimizing its weights and biases to minimize the cost function. This process utilizes gradient descent method and a procedure known as backpropagation.²⁷ First and foremost, a large data set is needed to successfully train a network from scratch and to overcome problems related to overfitting. For example, the well-known AlexNet was trained using ~1.2 million images divided into 1000 categories.^{16,23}

For our task of binary classification of multiphoton images from ovarian and surrounding reproductive tract tissues, no extensive data sets yet existed, and neither was it feasible to generate a vast amount of data. Therefore, instead of training a CNN from scratch, we used four pretrained CNNs (AlexNet, VGG-16, VGG-19, and GoogLeNet). These CNNs were chosen as they are openly available and due to their success in the ImageNet Large Scale Visual Recognition Challenges.^{23,24} AlexNet was the first successful CNN winning the 2012 challenge outperforming thus the more conventional approaches. The more sophisticated VGG-16 and VGG-19 networks were the winners of the following year and were again superseded

by the GoogLeNet in the 2014 competition. Since we had no prior knowledge on how well each of these CNNs could perform on our classification task, we fine-trained all of them.

We replaced their last few FC layers, originally responsible for the 1000-way classification of ImageNet data,^{23,24} with a binary classifier enabling fine-training of the modified CNN using a considerably smaller data set consisting of ~ 200 images. Since it was not *a priori* clear what kind of classifier would result in the best classification performance, we used two different approaches. In the first, we replaced the final FC layers by a linear SVM, since SVMs are often used for binary image classification. In the second approach, we replaced the final FC layers by new layers (sequential FC, Softmax, and classification layers) more suitable for binary classification. Figure 1 shows a layout illustrating the two chosen approaches. Since in these approaches we were fine-training the modified CNNs using smaller amounts of data, overfitting could cause problems, but such problems were mitigated by data augmentation and dropout as shown in earlier reports focusing on medical image analysis.^{21,26,28–30}

3 Experiments and Results

Animal experiments were performed in accordance with the Canadian Council on Animal Care's Guidelines for the Care and Use of Animals under a protocol approved by the University of Ottawa's Animal Care Committee. Samples were acquired from five healthy FVB/n mice and five syngeneic mice with HGSC-like ovarian cancer generated by injection of spontaneously transformed ovarian surface epithelial (STOSE) cells under the ovarian bursa.² Five $6\text{-}\mu\text{m}$ -thick sections were prepared both from the upper reproductive tract of healthy mice ($n = 5$) and from STOSE ovarian tumors ($n = 5$). Four sections from each sample were left unstained and imaged using a multiphoton microscope. One section per sample was stained with picosirius red and was used for overall inspection of the tissues.

All samples were imaged by measuring backscattered TPEF and SHG signals. In order to ensure that the trained classifiers could correctly classify images where parts of surrounding non-ovarian tissues are present, tissues from the upper part of the reproductive tract were also imaged. A Ti:sapphire femtosecond laser (Mai Tai HP, Spectra Physics) with 80-MHz repetition rate

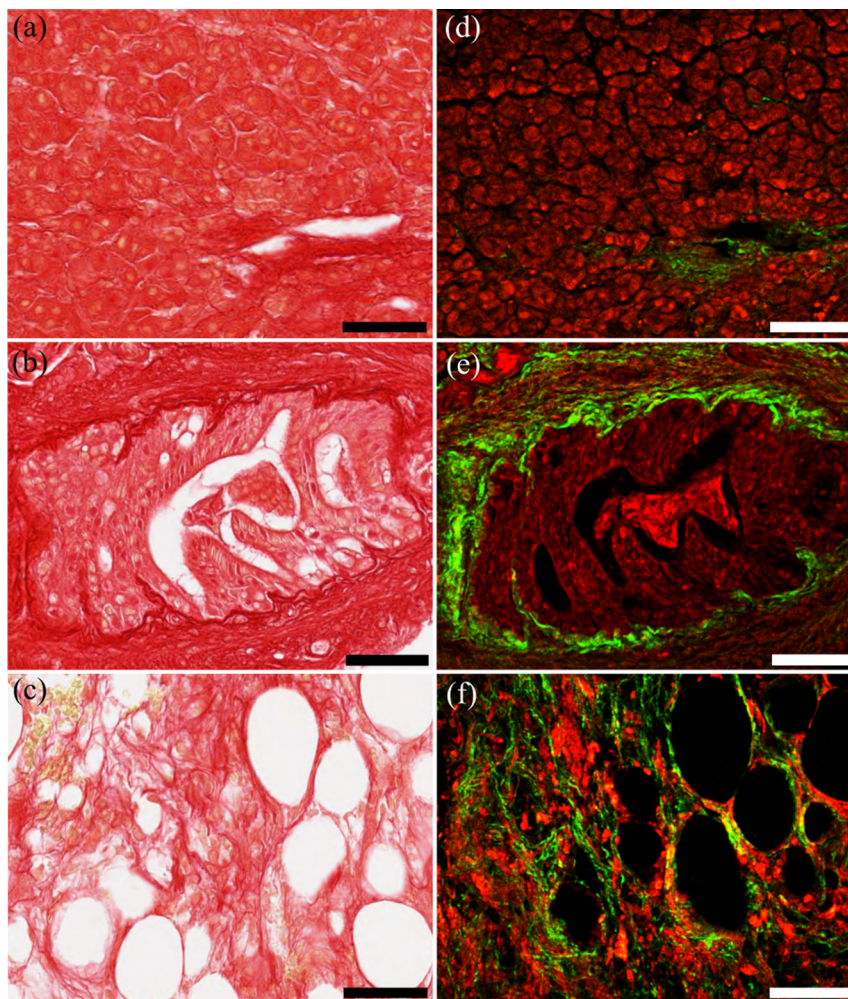


Fig. 2 (Left) Representative bright-field images from stained murine model (a) healthy ovarian tissue, (b) healthy reproductive tract tissue, and (c) HGSC tissue. Collagen appears dark red in the stained tissue images. (Right) (d)–(f) Corresponding multiphoton images from adjacent unstained sections, respectively. Relative to healthy ovary (a) and (d), remodeling of ECM is visible in cases of HGSC (c) and (f) as an increase in the amount of collagen and consequent SHG signal (green). In addition, the overall tissue morphology becomes less organized which is visible in the intrinsic TPEF signal (red). Scale bars are $50\ \mu\text{m}$.

and ~ 150 -fs pulses at the incident wavelength of 840 nm was used for excitation in conjunction with a laser-scanning microscope (Fluoview FVMPE-RS, Olympus). All measurements were taken with a $40\times$ (NA = 0.8) water-immersion objective (LUMPlanFL, Olympus). The average incident power at the sample plane was 5 to 10 mW, which was adjusted using a polarizer and a rotating half-wave plate along the beam line. A quarter-wave plate and a Soleil-Babinet compensator were used to ensure that the incident polarization at the sample plane was circular. Circular polarization was used to make sure that anisotropic structures, in our case mainly the collagen fibers, were evenly excited and imaged. The backscattered nonlinear signals were separated from the fundamental beam using a dichroic mirror (DM690, Olympus). The TPEF signal was separated from the SHG signal using another dichroic mirror (FF452-Di01, Semrock) and the SHG signal was further filtered using a band-pass filter (FF01-420/10, Semrock).

Both SHG and TPEF images consisting of 800×800 pixels were simultaneously acquired with a field-of-view of $\sim 250 \times 250 \mu\text{m}^2$. A pixel dwell time of $8 \mu\text{s}$ was used and each image pixel was averaged 16 times to improve the signal-to-noise ratio, resulting in an imaging speed of 82 s per image. The raw data were transformed into RGB images, where the red (green) channel corresponded to TPEF (SHG) signal and the blue channel was set to zero. Representative multiphoton images from healthy and cancerous reproductive tissues along side with the corresponding bright-field images from adjacent stained sections are shown in Fig. 2. Remodeling of the ECM is visible as an increase in the amount of collagen and thus SHG signal in the cancerous tissue, while changes in the overall tissue morphology are seen in the TPEF signal [compare Figs. 2(c) and 2(f)].

As the data set of ~ 200 images was relatively small for our purposes, we first augmented the data using patch extraction. The original RGB images were divided into N evenly spaced patches (see Fig. 3) consisting of 227×227 (224×224) pixels, to match the input size requirements of the pretrained CNN

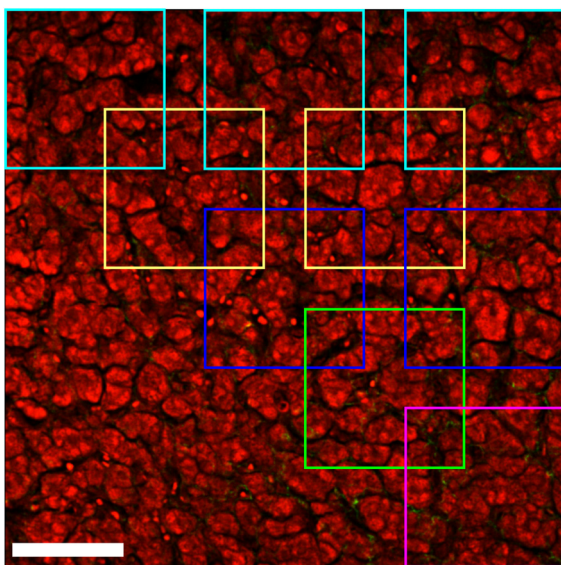


Fig. 3 Schematic illustrating the overlap between the extracted patches (colored squares) for the case of $N = 25$. For clarity, only every second patch in each row on the upper triangle of the image is shown. Scale bar is $50 \mu\text{m}$.

AlexNet (VGG-16, VGG-19, and GoogLeNet). This choice also maintained the same field-of-view in the patches, as varying field-of-view might affect the results. To minimize the amount of overlapping data, we only considered cases $N = 1, 4, 9, 16,$ and 25 . The performed patch extraction for one example image for the case of $N = 25$ is shown in Fig. 3. Due to the reduced field-of-view some of the image patches were found to be very dark, containing only minimal image features. As such patches could compromise the training, patches with mean pixel values below 3% of the maximum pixel count value were excluded from the analysis. Data sets processed in this way were further augmented using horizontal and vertical reflections together resulting in further threefold increase in the data set size. Therefore, the overall data augmentation scheme, consisting of patch extraction along with horizontal and vertical reflections, led up to a 75-fold increase in the training set size.

The whole data set was randomly divided into training and validation sets using a ratio of 60/40, respectively. The classifiers were then trained using the training data set and validated using the validation set by calculating the classification sensitivity (true-positive rate), specificity (true-negative rate), and accuracy (number of correct classifications divided by the total number of cases). The classification performance of the two studied approaches as discussed in Sec. 2 (using SVMs with learned features from pretrained CNNs versus fine-trained modified CNNs) was quantified in this way. Since the training and validation sets were randomly chosen, the calculated accuracies varied slightly for each training event. Therefore, training events were repeated 25 times and the mean sensitivities, specificities, and accuracies (along with their standard deviations) are reported for better representation of the results. The results for all the studied classifiers are shown in Fig. 4.

As a second step, we estimated how well the approach generalizes to independent data sets by performing leave-two-mice-out cross-validation, where the classifiers are trained using image data taken from eight mice and validated using the two remaining independent ones. This better represents a realistic scenario in which the classifier is first trained on known samples, and then used to diagnose a sample being observed for the first time. Because the approach of fine-training CNNs resulted in better classification performance compared to using SVMs with learned features from the pretrained CNNs, only the approach based on fine-training CNNs was used for this validation test. During this test, the CNNs were independently trained on sets from eight samples before being validated on the remaining two samples, which they were seeing for the first time. The training process was repeated for all the 25 possible data set permutations and the results for the calculated sensitivities, specificities, and accuracies with their standard deviations are shown in Fig. 5.

4 Discussion

In general, three trends are visible in our results. First, it is clear that the patch extraction improves the results since increasing N systematically improves the classification performance (see the colored markers in Fig. 4). Second, more conventional classifiers based on SVMs [see Fig. 4(a)] are clearly outperformed by the classifiers based on fine-trained CNNs [see Fig. 4(b)]. When fine-trained CNNs are used, the classification sensitivity, specificity, and accuracy all increase on average by $\sim 3\%$, which is a marked improvement. Third, classification performance (sensitivity, specificity, and accuracy) increases by

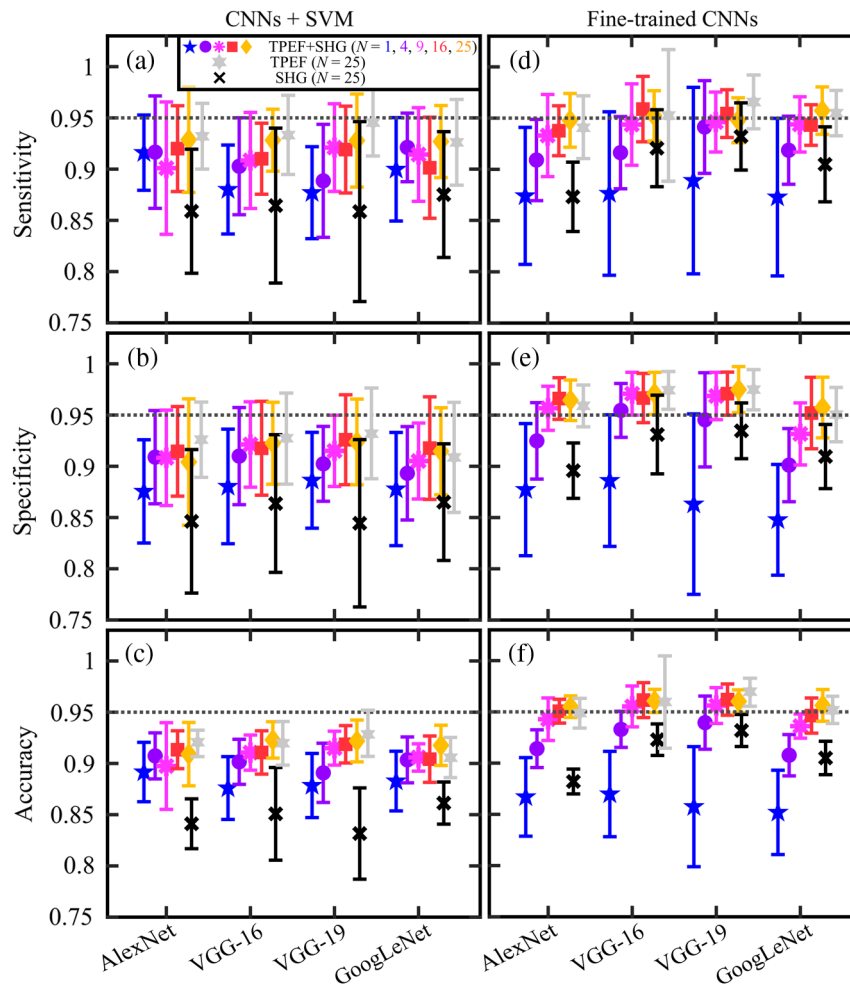


Fig. 4 (Left) Calculated (a)–(c) sensitivity, specificity, and accuracy for the classifiers using SVMs with learned features from the pretrained CNNs, respectively. (Right) Calculated (d)–(f) sensitivity, specificity, and accuracy for the classifiers formed by fine-training the CNNs. In general, increasing number of image patches N improves the results (see colored markers). Each data point is the mean result of 25 separately trained classifiers with the error bars corresponding to the respective standard deviation. Classification performance using only the SHG (TPEF) data are shown with black crosses (gray stars), on average resulting in $\sim 5\%$ ($\sim 0.3\%$) decrease in the classification performance compared to classifiers trained using both the TPEF and the SHG data.

$\sim 5\%$ when the classifiers are trained using both the TPEF and the SHG data (see the colored markers in Fig. 4), compared to training using only the SHG data (see the black crosses in Fig. 4). However, when the classifiers were trained by using only the TPEF data, the classification performance decreased only marginally ($\sim 0.3\%$) compared to training with both TPEF and SHG data. This is a somewhat surprising result, because one intuitively expects a clear increase in classification performance when more data are used. Further investigation would be necessary to determine whether this performance difference is typical. Therefore, combined TPEF and SHG microscopy seem beneficial over solely SHG (or TPEF) microscopy. This is somewhat expected since the data set is twice as big, and since the TPEF + SHG images can support additional features not visible in bare SHG or TPEF images.

The highest mean sensitivity ($95.2 \pm 2.5\%$), specificity ($97.1 \pm 2.1\%$), and accuracy ($96.1 \pm 1.1\%$) were found by fine-training the VGG-16 network using $N = 25$ image patches while using the training/validation scheme (see Fig. 4). But we note that all the studied CNNs performed almost equally well,

implying that the choice of which pretrained CNN to use is not crucial. We believe that this is mostly because the studied CNNs were originally designed and trained to classify images into 1000 of different classes, which is a considerably more challenging computer vision task than the binary classification performed in this work. Therefore, it seems plausible that all of the studied CNNs exhibited adequately complex network structures to allow their successful training for the simpler task of binary classification. However, the size of the training data set was found important and should be maximized, for example, using data augmentation, as done in this work.

Then we discuss the leave-two-mice-out cross-validation results (see Fig. 5). In general, the calculated sensitivities, specificities, and accuracies were slightly lower ($\sim 3\%$ to 4%) than what we achieved using the randomized training/validation scheme (see Fig. 4). However, the best performing classifier (fine-trained modified VGG-19) still resulted in very high classification sensitivity ($94.1 \pm 4.4\%$), specificity ($93 \pm 7.5\%$), and accuracy ($93 \pm 4.5\%$) for the case of $N = 25$ (marked as yellow diamonds). Therefore, the results suggest that the studied

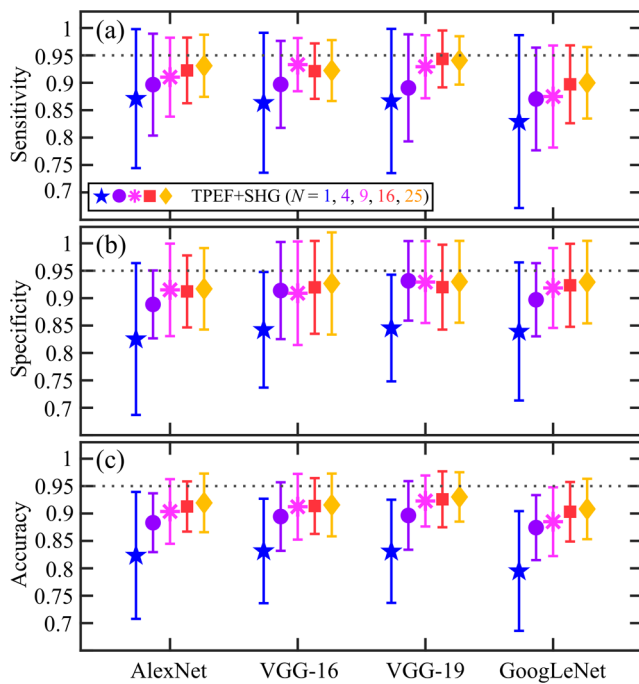


Fig. 5 Calculated (a) sensitivity, (b) specificity, and (c) accuracy for the four fine-trained CNN classifiers using leave-two-mice-out cross-validation with the error bars corresponding to the respective standard deviation. Both TPEF and SHG data were used in the training and analysis. The fine-trained VGG-19 network showed the best classification sensitivity ($94.1 \pm 4.4\%$), specificity ($93 \pm 7.5\%$), and accuracy ($93 \pm 4.5\%$) for the case of $N = 25$ (marked as yellow diamonds), respectively.

approach could provide automated and reliable ovarian tissue classification based on label-free multiphoton microscopy images.

Label-free images based on contrast from intrinsic multiphoton SHG and TPEF processes were used to demonstrate the deep learning technique in this study. Among the many advantages of the demonstrated approach are that it scales very favorably with the increasing amount of data. This is not necessarily the case for more conventional approaches based on user-defined filters and data analysis.^{14,26} The amount of training data could be increased further using a multimodal approach based on other label-free nonlinear modalities, such as third-harmonic generation,^{31,32} coherent anti-Stokes Raman scattering,²⁵ or polarized SHG.^{33–36} In addition, considerably larger data sets could be generated, for example, by switching to 3-D volumetric imaging. Recent work suggests that such a switch could improve the classification accuracy.¹⁴

The method demonstrated in this study is quite general and could be readily extended to other tasks, such as multiclass classification of tissues between known cancer types or stage classification of malignant tumors.^{14,37} We also believe that this approach is not restricted only to cancerous tissues, but could be straightforwardly extended to study and classify other diseases/disorders known to correlate with ECM remodeling, such as many fibrotic diseases.^{10,11,34,36}

Finally, we discuss the speed of the approach. The complexity of the used CNN and the amount of data define the training time along with the used training parameters. Training was performed using stochastic gradient method with a batch size of 50, initial learning rate of 0.0001 for up to four epochs.¹⁶

Fine-tuning the simplest CNN (AlexNet) using 25 image patches took around 300 s, whereas the same training took ~ 1 h for the computationally most demanding CNN (VGG-19). A graphics processing unit (NVIDIA GeForce GTX 1080 Ti) was used to speed-up the training. We note that the training times were considerably shorter when the learned features of pretrained CNNs were used to train an SVM classifier. But we emphasize that irrespective of the training time, which in general could be long, the actual classification process using the learned classifiers is quite fast (8 to 50 ms/image). Therefore, the computationally demanding training process does not compromise potential applications, since real-time image classification is perfectly feasible.

5 Conclusion

We have performed combined SHG and TPEF microscopy on normal and cancerous murine ovarian and surrounding reproductive tissues. We demonstrated that already with a relatively small data set consisting of ~ 200 images, pretrained CNNs can be fine-trained into binary image classifiers to correctly classify the images with over sensitivity 95% and 97% specificity. We compared four pretrained networks (AlexNet, VGG-16, VGG-19, and GoogLeNet) and investigated how data augmentation improves the classification performance. We also showed that training the classifiers using both the TPEF and SHG data is beneficial compared to using only the SHG data.

Histopathological image analysis of stained tissue slides is routinely used in tumor detection and classification. Diagnosis requires a highly trained pathologist and can thus be time-consuming, labor-intensive, and potentially risks bias. The trained classifiers demonstrated in this paper perform in real-time and could thus be potentially useful for clinical applications, such as for computer-aided diagnosis. The technique demonstrated here will also be valuable for investigating the etiology of ovarian cancer. Since the approach is very general, it could be easily extended to other nonlinear optical imaging modalities and to various biomedical applications.

Disclosures

The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose.

Acknowledgments

The Canada Excellence Research Chairs and Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIn-418389-2012-RWB), NSERC-Discovery Grant (SM), Finnish Cultural Foundation (00160028-MJH), Academy of Finland (310428-MJH), and Vanier Canada graduate scholarship (CM).

References

- R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: Cancer J. Clin.* **66**, 7–30 (2016).
- C. W. McCloskey et al., "A new spontaneously transformed syngeneic model of high-grade serous ovarian cancer with a tumor-initiating cell population," *Front. Oncol.* **4**, 53 (2014).
- W. Denk, J. H. Strickler, and W. W. Webb, "Two-photon laser scanning fluorescence microscopy," *Science* **248**, 73–76 (1990).
- J. Prat, "New insights into ovarian cancer pathology," *Ann. Oncol.* **23**, x111–x117 (2012).
- J. M. Watson et al., "In vivo time-serial multi-modality optical imaging in a mouse model of ovarian tumorigenesis," *Cancer Biol. Ther.* **15**(1), 42–60 (2014).

6. P. J. Campagnola, "Second harmonic generation imaging microscopy: applications to diseases diagnostics," *Anal. Chem.* **83**, 3224–3231 (2011).
7. R. M. Williams et al., "Strategies for high-resolution imaging of epithelial ovarian cancer by laparoscopic nonlinear microscopy," *Transl. Oncol.* **3**, 181–194 (2010).
8. O. Nadiarnykh et al., "Alterations of the extracellular matrix in ovarian cancer studied by second harmonic generation imaging microscopy," *BMC Cancer* **10**, 94 (2010).
9. N. D. Kirkpatrick, M. A. Brewer, and U. Utzinger, "Endogenous optical biomarkers of ovarian cancer evaluated with multiphoton microscopy," *Cancer Epidemiol. Biomarkers Prev.* **16**, 2048–2057 (2007).
10. T. R. Cox and J. T. Erler, "Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer," *Dis. Model. Mech.* **4**, 165–178 (2011).
11. C. Bonnans, J. Chou, and Z. Werb, "Remodelling the extracellular matrix in development and disease," *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014).
12. P. P. Provenzano et al., "Collagen density promotes mammary tumor initiation and progression," *BMC Med.* **6**, 11 (2008).
13. B. L. Wen et al., "Texture analysis applied to second harmonic generation image data for ovarian cancer classification," *J. Biomed. Opt.* **19**, 096007 (2014).
14. B. Wen et al., "3D texture analysis for classification of second harmonic generation images of human ovarian cancer," *Sci. Rep.* **6**, 35734 (2016).
15. O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Networks* **10**, 1055–1064 (1999).
16. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. of the 25th Int. Conf. on Neural Information Processing Systems*, pp. 1097–1105 (2012).
17. B. van Ginneken, S. Kerkstra, and J. Meakin, "Grand Challenges in Biomedical Image Analysis," 2012, <https://grand-challenge.org/> (21 March 2018).
18. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556> (21 March 2018).
19. K. He et al., "Deep residual learning for image recognition," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE (2016).
20. C. Szegedy et al., "Going deeper with convolutions," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–9, IEEE (2015).
21. D. Wang et al., "Deep learning for identifying metastatic breast cancer," 2016, <https://arxiv.org/abs/1606.05718> (21 March 2018).
22. J. Donahue et al., "Decaf: a deep convolutional activation feature for generic visual recognition," 2013, <https://arxiv.org/abs/1310.1531> (21 March 2018).
23. J. Deng et al., "Imagenet: a large-scale hierarchical image database," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 248–255, IEEE (2009).
24. O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.* **115**, 211–252 (2015).
25. S. Weng et al., "Combining deep learning and coherent anti-Stokes Raman scattering imaging for automated differential diagnosis of lung cancer," *J. Biomed. Opt.* **22**, 1–10 (2017).
26. L. B. Mostaço-Guidolin et al., "Collagen morphology and texture analysis: from statistics to classification," *Sci. Rep.* **3**, 2190 (2013).
27. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323**, 533–536 (1986).
28. N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
29. N. Tajbakhsh et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?" *IEEE Trans. Med. Imaging* **35**, 1299–1312 (2016).
30. Y. Bar et al., "Deep learning with non-medical training used for chest pathology identification," *Proc. SPIE* **9414**, 94140V (2015).
31. D. Débarre et al., "Imaging lipid bodies in cells and tissues using third-harmonic generation microscopy," *Nat. Methods* **3**, 47–53 (2006).
32. B. Weigelin, G. J. Bakker, and P. Friedl, "Third harmonic generation microscopy of cells and tissue organization," *J. Cell Sci.* **129**, 245–255 (2016).
33. A. Golaraei et al., "Characterization of collagen in non-small cell lung carcinoma with second harmonic polarization microscopy," *Biomed. Opt. Express* **5**, 3562–3567 (2014).
34. M. Strupler et al., "Second harmonic imaging and scoring of collagen in fibrotic tissues," *Opt. Express* **15**, 4054–4065 (2007).
35. H. Lee et al., "Chiral imaging of collagen by second-harmonic generation circular dichroism," *Biomed. Opt. Express* **4**, 909–916 (2013).
36. D. Rouède et al., "Determination of extracellular matrix collagen fibril architectures and pathological remodeling by polarization dependent second harmonic microscopy," *Sci. Rep.* **7**, 12197 (2017).
37. J. D. Brierley, M. K. Gospodarowicz, and C. Wittekind, *TNM Classification of Malignant Tumours*, 8th ed., John Wiley & Sons, Oxford, England (2017).

Biographies for the authors are not available.