

FULL ARTICLE

Hierarchical deep convolutional neural networks combine spectral and spatial information for highly accurate Raman-microscopy-based cytopathology

Sascha D. Krauß  | Raphael Roy | Hesham K. Yosef  | Tatjana Lechtenen  |Samir F. El-Mashtoly  | Klaus Gerwert  | Axel Mosig*

Department of Biophysics, Ruhr-University Bochum, Bochum, Germany

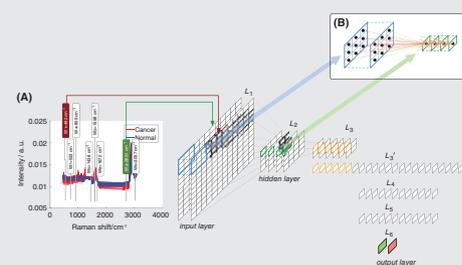
*Correspondence

Axel Mosig, Department of Biophysics, Ruhr-University Bochum, 44780 Bochum, Germany.
Email: axel.mosig@bph.rub.de

Funding information

European Regional Development Fund, European Union and North-Rhine Westphalia, Germany; Ministry of Innovation, Science and Research (MIWF) of North-Rhine Westphalia, Germany; Protein Research Unit Ruhr within Europe (PURE)

Hierarchical variants of so-called deep convolutional neural networks (DCNNs) have facilitated breakthrough results for numerous pattern recognition tasks in recent years. We assess the potential of these novel whole-image classifiers for Raman-microscopy-based cytopathology. Conceptually, DCNNs facilitate a flexible combination of spectral and spatial information for classifying cellular images as healthy or cancer-affected cells. As we demonstrate, this conceptual advantage translates into practice, where DCNNs exceed the accuracy of both conventional classifiers based on pixel spectra as well as classifiers based on morphological features extracted from Raman microscopic images. Remarkably, accuracies exceeding those of all previously proposed classifiers are obtained while using only a small fraction of the spectral information provided by the dataset. Overall, our results indicate a high potential for DCNNs in medical applications of not just Raman, but also infrared microscopy.



KEYWORDS

neural networks, Raman spectroscopy, supervised machine learning, urinary bladder neoplasms

1 | INTRODUCTION

Raman microscopy has been utilized extensively for diagnostic purposes in recent years. With applications ranging from cytopathology of different types of cancer [1, 2] to resolving tumors within their tissue context [3–5], these studies have relied on analyzing image spectra pixel-by-pixel, typically using multidimensional chemometric approaches to classify or decompose individual Raman spectra. Some recent approaches have deviated from this approach of considering one pixel spectrum at a time by

analyzing the spatial context contained in Raman microscopic images in a quantitative manner [6]. The authors of the present contribution demonstrated recently that the morphological differences between normal and cancerous urothelial cells can be captured using image features that have been commonly used for location studies of fluorescence microscopic images [7].

In recent years, hierarchical variants of *deep convolutional neural networks* (DCNNs) facilitated breakthrough results in numerous pattern recognition tasks. The rise of DCNNs has been triggered by the seminal work of

Krizhevsky et al. [8], which has no less than revolutionized the fields of machine learning and pattern recognition [9], including in particular breakthroughs in medical imaging and in diagnostics [10]. Obviously, this development promises impact on the analysis of Raman microscopic images and its applications, which constitutes the subject of our present contribution.

We demonstrate the use of DCNNs for Raman microscopic image analysis in the context of bladder cancer cytopathology, where cell material is obtained noninvasively from the sediment of urine samples. Urothelial cells from the sediment can be classified into normal and cancer cells. As has been suggested previously, such noninvasive approaches promise to reduce the need for urethrocystoscopy, which is the currently most accurate approach, yet suffering from its inherent invasiveness and high costs. The case has been made that noninvasive tests are a concern particularly relevant for bladder cancer due to the high rate of recidivating and the resulting need for frequent testing recurrence within short intervals. *Spectral cytopathology* has been introduced as a promising approach [11] in the context of other cancer entities. Subsequent work clearly demonstrated the potential of spectral cytopathology in the context of urine cytopathology [12].

1.1 | Raman-microscopy-based cell identification

It is well established that the collected Raman spectra of a specific microscopic pixel are highly representative for the biochemical status of the sample at the respective location. Numerous studies have demonstrated that image spectra can distinguish different cell types [7, 13–15], identify subcellular compartments [16, 17], or even resolve the distribution and metabolization of drugs within cells [18]. In these studies, numerous chemometric approaches have been utilized, ranging from different clustering approaches to various flavors of supervised classifiers [17, 19–21] as well as matrix factorization approaches [22]. In some cases, image spectra are classified on a pixel-by-pixel manner [17], while other studies rely on the analysis of either average spectra [23–29] or individual spectra that integratively represent a complete cell or region due to experimental setup [13, 19]. By Krauß et al. [7], it has been demonstrated that majority votes over classification outcome of individual pixel spectra within an image facilitate more accurate classification compared to approaches based on average spectra of Raman microscopic images.

The gold standard in such cytopathology studies usually is the visual identification of cancerous urothelial cells by a human expert in hematoxylin and eosin (HE)-stained images. In other words, cancer cells can be identified by morphological and textural features contained in the underlying imaging data, rather than by features of Raman microscopic image spectra which immediately reflect biochemical alterations at the respective pixel location. It is commonly

accepted across multiple forms of cancer that cells undergo morphological alterations in particular within their nuclei [30]. In the case of bladder cancer, NMP22 as a nuclear-matrix protein is one of the few approved urine biomarkers for bladder cancer [31]. NMP22 is a major constituent of the nuclear matrix, providing evidence that morphological alterations in the nuclei of cancer cells can be linked to molecular alterations. This is in line with the findings from ref. [7] that morpho-textural information obtained from only 3 wavenumbers in Raman microscopic images is sufficient to recognize cancer cells from urine sediment with very high accuracy.

The recognition of cancer cells in ref. [7] is based on a well-established approach to extract morphological and textural features from the intensity-image obtained for each of the 3 wavenumbers utilized for this previous study. Remarkably, this approach yields accurate classifiers while ignoring most of the spectral information contained in the Raman microscopic image spectra. The current state-of-the-art in Raman-microscopy-based cytopathology naturally raises the quest for classifiers that can take into account *both* morpho-textural information *and* the spectral information contained in Raman spectra. In this contribution, we assess the potential of deep convolutional neural networks to take this role.

2 | FROM CONVENTIONAL TO DCNNs

Artificial neural networks have been introduced in the early days of artificial intelligence research, and have been one of the predominant approaches in machine learning since. In their conventional form, neural networks have also been largely popular for classifying Raman and also infrared spectra [32–34], in particular when dealing with pixel spectra from corresponding microscopic images [35]. In these studies, conventional neural network classifiers utilize the biochemical status represented by an individual Raman spectrum, but do not take into account morphology.

In the following paragraphs, we provide a brief and very basic introduction to artificial neural networks. For a more thorough treatment of the topic, we refer to text books such as the excellent introduction by Goodfellow and Bengio [36]. Practitioners may find a suitable starting point in the book by Michael Nielsen [37], which introduces neural networks in a hands-on fashion accompanied by a highly descriptive and accessible presentation of the theoretical foundations.

2.1 | Basic terminology

We consider neural networks for binary classification tasks, where a n -dimensional input vector $x = (x_1, \dots, x_n)$ is classified into 1 out of 2 classes. In other words, the neural network computes a function $c(x)$, where $c(x) = 1$ or $c(x) = 2$ indicates that x is assigned to class 1 or 2, respectively. In conventional spectral classifiers, x may represent a Raman

spectrum where x_1, \dots, x_n indicate the intensities at different wavenumbers. Classes $c(x) = 1$ or $c(x) = 2$ may indicate healthy or disease, respectively, as the status of the sample.

In the most simple case, a neural network comprises 2 layers, as shown in Figure 1. A: Each input component x_i is represented by 1 vertex in the input layer L_1 , while the 2 output classes are represented by vertices y_1 and y_2 in output layer L_2 . Each output vertex y_j sums up the inputs x_i scaled by a weight of $w_{i,j}$, so that the *activation* at output neuron y_1 is now computed through

$$a(y_j) = w_{1,j}x_1 + w_{2,j}x_2 + w_{3,j}x_3 + b_j,$$

where b_j is a *bias* value required for technical reasons. The classification outcome can be obtained from identifying which of the 2 activations $a(y_1)$ or $a(y_2)$ is maximal.

The weights $w_{i,j}$ and b_j of a neural network are usually obtained from training data, which is a set data points with a specified class label. Training the neural network essentially translates to adjusting the parameters so that the classification error on the given training data will be minimized. This is famously accomplished by the *backpropagation algorithm* [38].

2.2 | Multilayered neural networks

Figure 1B illustrates a typical multilayered neural network consisting of several layers of vertices. As in the 2-layered case, each vertex integrates the incoming edges by multiplying the signal incoming along an edge with a weight associated with the edge. The integrated incoming weighted signals are passed through a *transfer function* on to the next vertex in the network. The first layer L_1 again constitutes the input layer, and the last layer L_4 is the output layer that determines the classification result. *Hidden layers* such as L_2

may involve a varying number of vertices. In conventional neural networks, vertices between consecutive layers are typically fully connected.

2.3 | Hierarchical DCNNs

Compared to conventional neural networks, the hierarchical architecture displayed in Figure 2A are tailored toward imaging data. In fact, a complete image with an arbitrary number of spectral channels is taken as input, rather than a single spectrum as in a conventional neural network. Specifically, morphology is taken into account by hierarchically merging information from neighboring pixel spectra. At the first hidden layer L_2 , information is merged from the immediate pixel neighborhood. Layer L_3 further integrates information from a broader neighborhood by merging the areas represented in layer L_2 . In this example, the spatial information in the tensors in level L_3 is flattened into a single feature vector in layer L_3' . The following layers L_4 to output layer L_6 represent conventional convolutional layers. In L_6 , each of the 2 vertices represents 1 class to classify cells into either healthy or cancerous. Moreover, Figure 2B indicates the architecture of our neural network MN_i introduced in detail below. It shows how more spectral information can be included in the neural network: the input layer L_1 can be easily extended to collect information from more than just 2 of the spectral bands. The network is thus capable of integrating spatial information as well as spectral information.

Including more spectral as well as more spatial information in a DCNN comes at the cost of increasing the number of edges in the resulting network. In other words, the number of parameters to be optimized during training will increase. Technically, large numbers of parameters are of limited concern, as current implementations of DCNNs can deal with neural networks with tens of millions or more

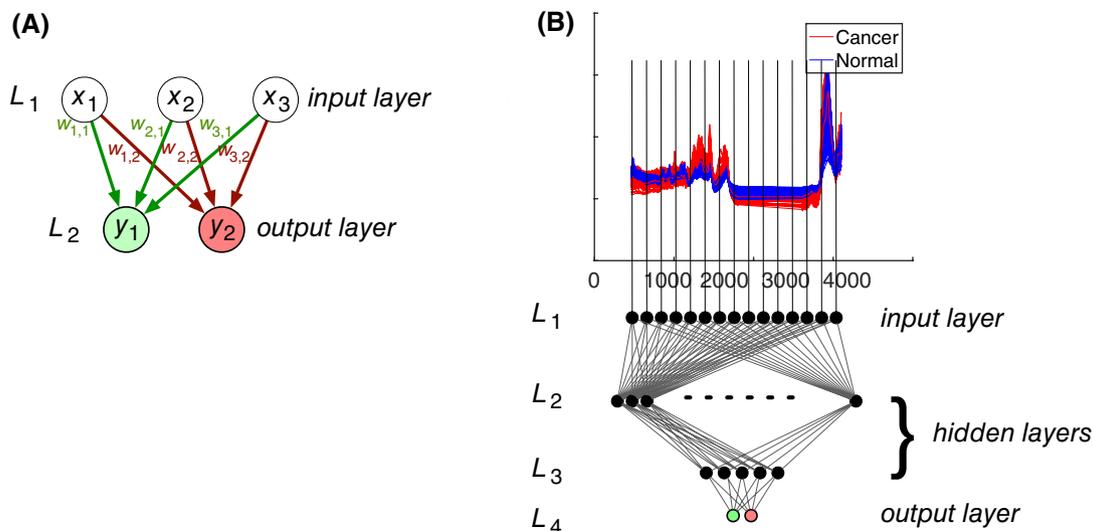


FIGURE 1 Conventional convolutional neural networks. Panel A indicates a minimal neural network with 3-dimensional input and 2 output classes. Panel B displays the topology of typical conventional convolutional neural network with 2 hidden layers

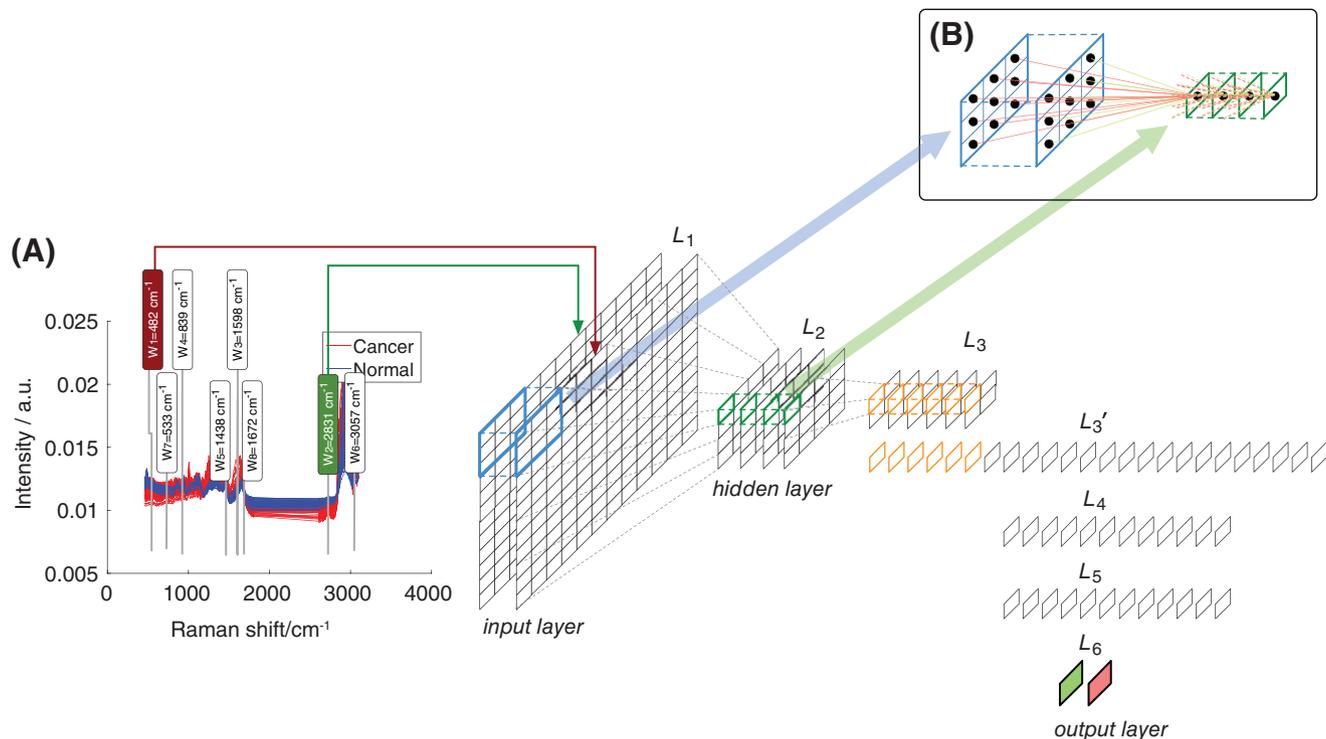


FIGURE 2 Conventional vs hierarchical deep convolutional neural networks. Panel A displays a reduced version of the 6-layered MN_i network. Input layer L_1 takes $i = 2$ wavenumbers of all pixel spectra of a Raman spectral image as input. (The assignment of the selected wavenumbers can be found in Table 3.) Consecutive layers reduce spatial resolution through hierarchically connecting subregions of the image as displayed in panel B. With increasing layers, information is carried through deeper tensors until a 1-dimensional representation is achieved in layer L_3' . Layers L_3' through L_6 are fully connected as in a conventional neural network. Spatial dimension and depth have been reduced for illustrative reasons; the actual dimensions implemented in networks MN_i are shown in Table 1

parameters. The main technical limitation in many cases is graphics hardware memory, which needs to fit the complete network topology along with a certain fraction of the training data into graphics memory. More importantly, however, large numbers of parameters constitute a challenge in terms of the neural networks obtained after training. If networks with many parameters are trained on too few or too homogeneous training data, they tend to exhibit either poor classification accuracy or strong overfitting toward the training data [41].

This general challenge in fact sets the frame of our contribution. We address the question of which DCNN topologies can be combined with the training data that are commonly available in spectral cytopathological studies, and which training conditions will lead to robust cell classifiers. Specifically, different topologies deal with trade-offs between the amount of spectral vs spatial information, and the depth of network vs the amount of data available for training. We investigate both training DCNNs from scratch using cytopathology data as well as utilizing pretrained networks. We further investigate feature extraction for controlling the amount of spectral information used for classification by the means of pretrained networks. Furthermore, so-called *transfer learning* [42, 43] of partially pretrained networks can be employed. Both approaches, relying on pretrained networks and using transfer learning, utilize

DCNNs that have been trained on large image datasets and are readily available, for example, the well known *AlexNet* [8] based on the *ImageNet* database [44] with several million pictures from thousands of classes. Even though the images underlying the *ImageNet* dataset display everyday objects rather than cytopathological data, the *AlexNet* DCNN produces low-level features that are potentially useful and discriminative for cytopathological images.

3 | MATERIALS AND METHODS

3.1 | Wet lab

In this study we have reanalyzed data from a recent study [7, 15]. For the sake of completeness, we briefly describe the previously described sample preparation and image acquisition [7, 15].

3.1.1 | Urine sampling

This study was conducted within the framework of the Protein research Unit Ruhr at the Ruhr-University Bochum, Germany. Prior to human urine sampling, institutional review board approval (IRB 3674-10) and written informed consent from all patients has been obtained. Urine samples were collected at Marienhospital Herne, Germany, from 10 patients diagnosed with high-grade urothelial bladder

cancer and from 10 more patients with pathologically confirmed urocystitis but without cancer. Collected urine samples were spun at 3700 rpm (10 minutes, 10°C). The supernatant urine was discarded and followed by suspension and fixation of the precipitated pellets of urine cells using 1 mL of 4% formaldehyde solution (Roti-Histofix, Carl Roth GmbH, Karlsruhe, Germany) and stored at 4°C. In case of haematuria, urine cell solution was filtered using a Millipore nylon-net filter of 11 µm pore size (Merck Chemicals GmbH, Darmstadt, Germany) and washed with phosphate-buffered saline (PBS, Life Technologies, Darmstadt, Germany), in order to remove blood cells, bacteria, yeast and other coloring contaminants.

Calcium fluoride (CaF₂) slides (Korth Kristalle, Kiel, Germany) were used as cell substrate for spectroscopic measurements. CaF₂ slides were coated with 0.01% (wt/vol) poly-L-Lysine (Sigma-Aldrich GmbH, Munich, Germany) in order to facilitate the attachment of cells [45]. An amount of 100 µL of urine cell solution was added to a cytofunnel assembly (Shandon, ThermoFisher GmbH, Dreieich, Germany) which contains a CaF₂ slide. The cytofunnel assembly was spun using cytospin centrifuge (Cytospin 4, ThermoFisher GmbH) at 1500 rpm for 20 minutes. Next, the CaF₂ slide carrying a spot of adhered urine cells was removed from the assembly and subsequently immersed in PBS buffer to proceed with spectroscopic measurements.

3.1.2 | Spectroscopy

A confocal Raman microscope (alpha300 AR, WITec, Ulm, Germany) was implemented for Raman measurements of urine cells, as described previously [18, 25]. Raman excitation source is a frequency-doubled Nd:YAG laser 532 nm (Crystal laser, Reno, Nevada) with the output power of approximately 40 mW. The excitation laser beam is directed into a Zeiss microscope by a wavelength-specific single-mode optical fiber, which is followed by collimation of the laser beam and focused on the sample by a Nikon NIR APO (×60/1.00 NA, Nikon, Düsseldorf, Germany) water immersion objective. The urine cell slide is fixed on a piezoelectrically driven microscope scanning stage. The collected Raman-scattered light is directed to a back-illuminated deep-depletion charge-coupled device camera that is operated at -60°C, which can detect the Raman signal. In this study, Raman imaging is conducted using a raster scanning laser beam over cells, in order to acquire the full Raman spectra at

speed of 0.5 seconds per pixel and a pixel resolution of 500 nm. From the 20 patients 60 high-grade cancer urothelial cells and 61 noncancerous cells were selected, which resulted in 375 203 and 161 937 Raman spectra from cancerous and noncancerous cells, respectively. The images in this dataset consist of 4439 pixel spectra on average.

3.1.3 | Staining

The staining of cells with HE was conducted after Raman measurements. The cells were fixed via spraying of a commercially fixative solution Merkofix (Merck KGaA, Darmstadt, Germany) that was left for 10 minutes to dry. Next, the cell slide was immersed in Haris hematoxylin solution (Merck KGaA) for 1 minute and subsequently washed in water stream for 1 minute. Afterward, cells were immersed in eosin Y (0.5% alcoholic, Merck KGaA) for 1 minute and then washed by a water stream for 1 minute. The cell slide was immersed subsequently in multiple washing solvents (15 seconds each): ethanol (96%), ethanol (100%) and Xylol (100%) (Sigma-Aldrich GmbH, Munich, Germany). Lastly, a liquid cover glass (Merkoglass, Merck KGaA) was added on the cell slide and then covered with a glass coverslip (Servoprax GmbH, Wesel, Germany) and left overnight to dry. For imaging of cells stained with HE, a Nikon upright microscope (Eclipse Ni-U, Düsseldorf, Germany) was employed. The microscopic imaging was conducted using a Nikon Plan APO (×60/1.4 NA, Nikon) oil immersion objective. All collected images were sent for cytopathologic annotation.

3.2 | Data analysis

A complete sample consists of thousands of cells on a slide. From each sample, a small number of cells relevant for classification is selected visually. These selected cells are first measured by Raman microscopy and then stained with HE (see section 3.1.3). A pathologist characterizes these stained cells as cancer or normal urothelial cells.

3.2.1 | Spectral preprocessing

Raman hyperspectral image data were exported to Matlab 2017b (The MathWorks, Natick, Massachusetts) and preprocessed using in-house Matlab scripts. All Raman spectra where the C-H band at 2850 to 3000 cm⁻¹ was absent were treated as background and erased. To remove cosmic spikes, a pulse noise filter was used and the Raman spectra were

TABLE 1 Metaparameters of the MN_i network architectures

MN _i Layer	L ₁	L ₂	L ₃	L _{3'}	L ₄	L ₅	L ₆
Dimension	48 × 48 × i	21 × 21 × 32	17 × 17 × 16	1 × 4624	1 × 256	1 × 256	1 × 2
Transfer function	(input)	ReLU	ReLU	(linear)	ReLU	ReLU	Softmax
Kernel		7 × 7	5 × 5	(flatten)	Inner product	Inner product	Softmax
Stride		2	1				
Type		Convolution	Convolution	Flatten	Inner product dropout	Inner product dropout	Output

TABLE 2 Hyperparameters used for training the 3 different kinds of DCNNs (transfer learning, AlexNet and minimal net)

	TL	AN	MN8
Maximum epochs	65	70	90
Mini-batch size	10	5	6
Initial learn rate (LR)	0.0001	0.0001	0.0002

interpolated to a reference wavenumber scale. In addition, all spectra have been vector normalized. We have carried out the classification on spectra without baseline correction, as we had previously shown that the baseline correction (with a third-order polynomial [46]) had only a very small influence on the classification results [7].

3.2.2 | Spectral band selection

The basis for training our morphological classifiers is formed by the identification of individual Raman wavenumbers that will be most informative toward distinguishing cancer cells from healthy cells. For this kind of feature selection problem numerous approaches have been proposed [47, 48]. Here, it will be particularly relevant to identify a small number of wavenumbers which on the one hand are highly distinctive between normal and cancer cells, while on the other hand should also be uncorrelated. A well-established approach to achieve this is the *max-relevance min-redundancy* (MRMR) algorithm [49], which optimizes information theoretic measures in order to identify features that will be informative to distinguish different classes. A crucial property of this approach is the identification of spectral bands that tend to carry uncorrelated information, so that the morphology for each selected band can be expected to carry different and thus additional information for morphological feature-based classification. We followed the workflow from [7] to select $k = 1$ to $k = 100$ many wavenumbers. For each k , a DCNN topology taking input images with k channels in the input layer L_1 was established.

3.2.3 | Deep learning

We derived 3 different approaches to use DCNNs from the *AlexNet* [8] topology. We used the *AlexNet* topology as implemented in *caffe* [50], which also allows to integrate the parameters obtained in [8] from training the network on the *ImageNet* dataset. Additionally, we established a new topology different from the *AlexNet* pattern.

Specifically, we assess the following set of DCNN-based classifiers:

Feature extraction

We utilized the pretrained *AlexNet* [8] as a *feature extractor*. To this end, we reduced each of the Raman images I_1, \dots, I_N to 3 wavenumbers using feature selection. Interpreting the resulting images as RGB images allows to apply the network from ref. [8] along with all parameters obtained on the *ImageNet* dataset to each image I_j in the training dataset.

We considered the tensor $T(I_j)$ resulting from the output layer L_7 as a feature representing each image I . We then trained a *support vector machine* (SVM) on the tensors $T(I_1), \dots, T(I_N)$, with each I_j labeled as displaying a cancer or a normal cell.

Transfer learning

As a second strategy, we applied so called *transfer learning*. To obtain network transfer learning (TL), we took the *AlexNet* pretrained on the *ImageNet* dataset as a starting point and ran further training iterations on the cytopathology data set I_1, \dots, I_N reduced to 3 channels as in network feature extraction (FE). During this transfer learning step, we fixed the layers L_1, \dots, L_4 and allowed only the parameters of the last 3 layers (a fully connected, softmax and classification layer) to be further optimized.

AlexNet topology

Our third strategy was to train the *AlexNet* from scratch on the cytopathology dataset I_1, \dots, I_N reduced to 3 channels. To obtain network AlexNet topology (AN), we did not use the *ImageNet* parameters at all, but started training the network from random initialization.

Minimal net

Due to its origins in photographic image recognition, the *AlexNet* is limited to utilizing 3 wavenumbers only. We devised a minimal topology with $i = 1, \dots, 30$ spectral bands in the input layer and including 6 hidden layers. The topology of minimal net (MN i) is displayed in Figure 2. Varying the number i of spectral bands allows to investigate the impact of including different amounts of spectral information for classification which can be seen in Figure 4.

These methods involve a number of metaparameters where Table 2 shows the values we used.

As a reference to compare the classifiers FE, TL and AN to conventional approaches, we took into account 2 classifiers previously established:

Spatial bagging

The classifier established in ref. [15] implements a conventional random forest [51] classifier that categorizes individual pixel spectra, combined with *spatial bagging*, where a majority vote over the classification outcomes of individual spectra within a complete spectral image leads to assigning the cancer vs normal to a cell [7].

Conventional morphology

As a purely morpho-textural classifier, we utilized the classifier proposed in ref. [7] which computes the morphotextural features proposed in ref. [52] and classified images based on these features using a conventional Random Forest classifier.

TABLE 3 List of the 8 most discriminative wavenumbers selected by the MRMR algorithm for training DCNNs, including their chemical explanation [39, 40]

Wavenumber	Chemical explanation
482	Glycogen, skeletal modes, in-plane bending (CCC)
533	Glycogen, skeletal modes, (CCC) and (CCO) in-plane, bending, glycosidic linkage, disulfide bridge of cystine
839	Glycogen, (COC) deformation, polysaccharide, structure
1438	CH ₂ bending deformation, CH ₂ , CH ₃ scissoring of lipids
1598	Carbonyl stretching, amide I, C=N, NH ₂ , adenine
1672	Carbonyl stretching, amide I
2831	C–H stretching
3057	N–H amide vibration, C ₂ C–H ₂ Aromatic stretching

3.3 | Implementation

The DCNNs FE, TL and AN were implemented in *Matlab 2017b* using the *Neural Network toolbox*. The networks MN_{*i*} were implemented in *Python 2.7* using the *Keras 2.0.9* and *TensorFlow 1.4.0* deep learning frameworks. Preprocessing of Raman spectra was based on implementations described previously [7]. Experiments were run on a server with 20 CPUs and 4 Nvidia GeForce GTX 1080 Ti graphics cards running *Ubuntu 16.04*.

3.3.1 | Validation and testing

Training neural networks involves 2 different stages at which the dataset needs to be divided into a training and a validation set. The first stage precedes training, where part of the dataset is withheld for validation after the training of the classifier is completed. The second stage occurs repeatedly during training, where in each iteration of the training, a gradient needs to be computed on the training set, and the effect of adjusting the parameters in the network by following the gradient is assessed on the withheld second stage validation set.

In both levels, we implemented withholding the validation set on a *per patient* basis, where 2 patients (1 healthy and 1 affected by cancer) were randomly selected. For every round of cross-validation, all cells and samples associated with these 2 patients were held back for testing, and the remaining patients' cells were used for training as previously suggested and discussed [17]. Dividing training set from validation set on the patient level as the highest hierarchical level possible follows the standards established in ref. [53] for reliable validation of the robustness of classifiers.

Dividing the validation set at the patient level provides a further advantage already discussed in a previous contribution [7]. The classification of an individual cell can be interpreted as a vote, so that we can classify an individual patient through majority vote over all cells associated with that patient. If, for example, 7 out of 12 cells from 1 patient are classified as cancerous, the whole patient is put into this category. This allows to calculate a second, patient based, accuracy that measures correctly diagnosed patients.

4 | RESULTS AND DISCUSSION

The wavenumbers identified by the MRMR feature selection algorithm as being the most important ones are shown in Table 3 together with their chemical explanation. While the minimal net classifier (MN_{*x*}) is able to handle any number of channels, the other DCNNs studied here are limited to a maximum of 3 wavenumbers, in this case 482, 2831 and 1598 cm⁻¹ being the 3 most discriminative ones as previously described [7, 15].

Table 4 shows the accuracies we achieved according to per-patient cross validation with the deep learning methods described in section 3.2.3. Table 4 also compares accuracies with the results on the same dataset from our previous publications [7, 15] using conventional classifiers. As the dataset gets balanced, an accuracy of, for example, 0.925 per cell means, that 111 of 120 cell images were correctly classified. This results in an accuracy of 1 based on patients, as for every patient the majority decision is made over all cells of this patient as described in section 3.3.1. In general, one may consider to lower the threshold from *50% of the cells* to, for example, *at least one cancerous cell*, which might be a more appropriate diagnostic strategy in the clinical practice. However, in our setting it did not affect results in any significant manner.

Table 4 shows that both the pretrained FE and the TL classifier nearly match the accuracy of the previously proposed spectral classifier [15]. Remarkably, the MN8 achieves significantly better accuracy than any conventional

TABLE 4 Overview of accuracies with different machine learning techniques on the same urothelial cell dataset

	Stained	Raman	# Channels
Spatial bagging (SB) [15]			
Per cell	/	0.96	755
Per patient	/	1	755
Conventional morphology (CM) [7]			
Per cell	0.84	0.89	3
Per patient	0.9	1	3
FE			
Per cell	0.86	0.93	3
Per patient	1	1	3
TL			
Per cell	0.87	0.93	3
Per patient	1	1	3
AN			
Per cell	0.78	0.88	3
Per patient	0.9	0.95	3
Minimal net (MN8)			
Per cell	/	0.99	8
Per patient	/	1	8

The last column indicates how many wavenumbers were used for classification. All accuracies were calculated through cross validation at patient level as both per-cell and per-patient accuracies. Results for classifiers SB and CM have been obtained from previous work [7, 15] and shown for comparison.

classifier while involving only 8 out of the 755 wavenumbers of the dataset. In other words, less than 2% of the spectral information is required to exceed the accuracy accomplished through the conventional spatial bagging (SB) classifier which requires the complete spectral information.

The DCNNs also exceed the accuracy obtained from the conventional morphological (CM) classifier, which uses the identical dataset reduced to the same 3 wavenumbers rather than the complete Raman spectrum. Two factors may contribute to the improved accuracy of DCNNs over conventional morphological classifiers. First, the features provided through the DCNN-derived tensor utilized by the FE classifier may represent morphological information more suitable toward classification. Second and more importantly, the topology of the DCNNs involves spectral classification in the sense that features from different bands of the spectrum are interlinked by the topology of the DCNN. Conversely, the conventional morphological features used by classifier CM are computed within isolated bands of the spectrum.

The 2 approaches FE and TL take into account the training parameters from the network trained on the noncytopathological *ImageNet* dataset. Training from scratch purely relying on the cytopathology data in the AN, on the other hand, reduces the accuracy (see Table 4) and manifests in a noisier and lengthier learning process (see Figure 3).

Another practically relevant aspect to be observed are significant differences in running time. Transfer learning for classifier TL on the full dataset requires less than 1 minute of computing time, while training the random forest of classifier SB involves more than 8 hours on the same machine. Classifying all 121 images with either of the DCNNs requires roughly half a second compared to 45 seconds for classifier SB. In general, DCNNs require less computing time during both training and classification in practice. This advantage is certainly largely due to the highly optimized graphics hardware implementations that are available for DCNNs.

5 | CONCLUSIONS

We could demonstrate that DCNNs facilitate the reliable identification of cancerous cells in a Raman-microscopy-based cytopathology study and exceed the accuracy of conventional pixel-by-pixel full-spectrum-based classifiers while using only a small fraction of the Raman spectrum. At the same time, DCNNs exceed the classification accuracy of conventional morphological classifiers which extract morpho-textural features separately in different spectral bands, indicating that the ability to link spatial information across different spectral bands positively affects classification. This allows the

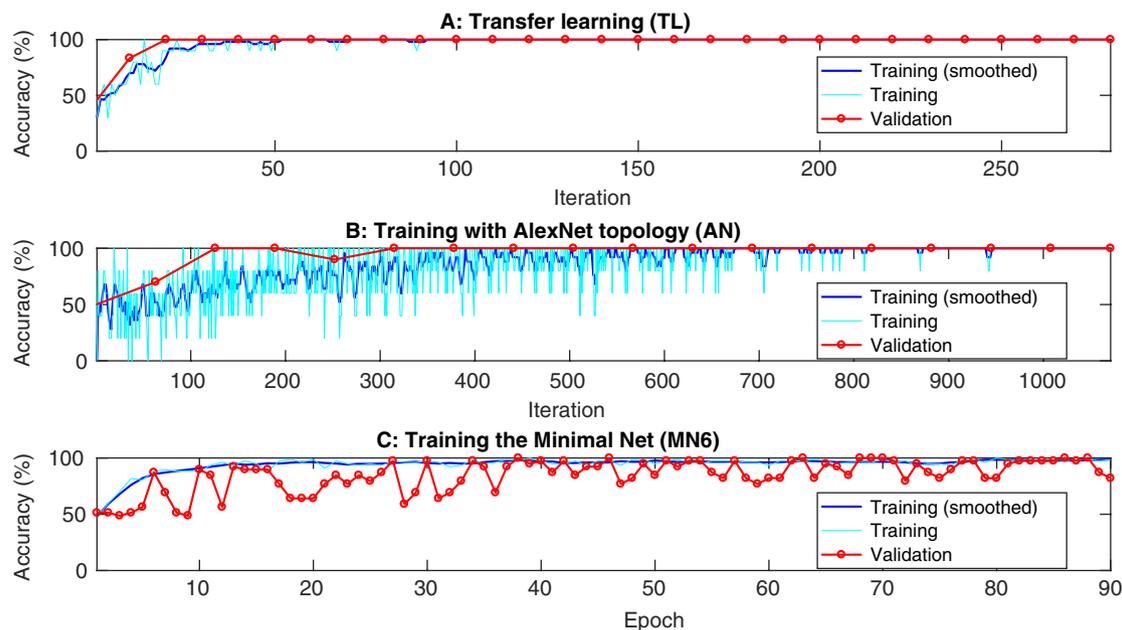


FIGURE 3 Visualization of the learning process showing the training and validation accuracy during (A) the transfer learning TL, (B) the training from scratch with the AlexNet topology AN and (C) the Minimal Net with 6 wavenumbers MN6. (A) The exemplary transfer learning lasted nearly 40 seconds and ended with meeting the validation criterion, where the loss on the validation set was larger than or equal to the previously smallest loss for 5 times in a row (*validation patience*). The *validation frequency*—meaning the number of iterations between evaluations of the validation loss—was set to 10 iterations. As the stop criterion was met after 280 of 585 possible iterations, only 32 of the maximal 65 epochs were executed. An iteration uses a mini-batch, an epoch is the full pass of the training algorithm over the entire training set—according to our mini-batch size of 10 images and a training set of approximately 90 images in this run, this results in 9 iterations per epoch. As the training accuracy is measured every iteration—in contrast to the validation accuracy—a smoothed curve is shown additionally. (B) The training of the AlexNet topology-based DCNN from scratch achieved 100% accuracy, too, but it lasted twice as long compared to transfer learning. It did not reach the stop criterion and had to fulfill the maximum of 70 epochs and 1120 iterations. (C) It was done using a different framework, where the accuracies are measured every epoch instead of every x iterations. For this example 6 wavenumbers were used, it took around 10 seconds

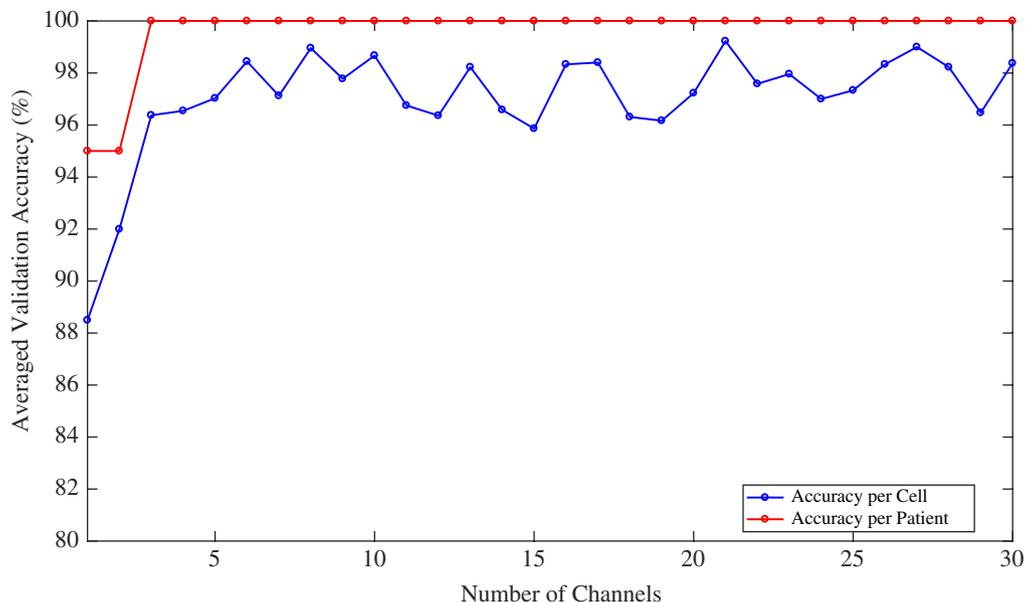


FIGURE 4 Impact of an increasing amount of wavenumbers on the accuracy achieved by training the minimal net

conclusion that DCNNs simultaneously take into account spectral as well as spatial information for classification.

The main challenge in applying DCNNs is to determine a topology that matches the type and the complexity of the data to be classified, and where the number of parameters to be trained matches the amount of data available for training. The training data set in our present study is by orders of magnitude smaller than the datasets in typical image recognition DCNNs [8, 44, 54]. Considering that some of these studies involve thousands of images per class, and that in some cases they can distinguish hundreds or thousands of different classes [8, 44, 54], it comes as a positive surprise that our comparatively small dataset allows to train relatively robust classifiers, as demonstrated by classifier AN based on the AlexNet topology. This forebodes that collecting more data for training classifiers will not just allow to train DCNNs with more layers and more parameters taking into account more spectral information, but that such DCNNs with more complex topologies may allow to distinguish cell types with less obvious differences, such as cells from early stage tumors or different subtypes of tumors. The challenge of requiring sufficient amounts of training data for such studies clearly translates into a challenge for the design of underlying clinical studies.

The potential of DCNNs has by far not been fully exploited by our present contribution. While some parameters such as the learning rate or the *mini batch* size have undergone minimal optimization, there is plenty of room for further improvement through adding *dropout* layers [55], performing regularization or taking optimization toward further modifying the network topology.

We also believe that the rise of DCNNs will impact not just microscopy of cellular material, and not just in the context Raman microscopy. Being able to classify on parts of

the spectrum is clearly favorable toward stimulated Raman scattering (SRS) and coherent anti-stokes Raman scattering (CARS) microscopy [56–58], where selected wavenumbers of the spectrum can be recorded at high speed. The significantly reduced recording time of SRS or CARS microscopy promises to relieve the challenge of acquiring sufficient amounts of training data.

Finally, it is easily conceivable that DCNNs will also impact the classification of infrared microscopic images, where large amounts of training data can be collected relatively easily [59, 60]. Infrared microscopy recently gained even more traction from the availability of quantum cascade laser systems [61], which increased the rate of data acquisition by at least one order of magnitude. As an overall conclusion, we see no reason to doubt that the groundbreaking impact of DCNNs on numerous fields of object recognition will come to halt at Raman or infrared imaging in its various forms.

ACKNOWLEDGMENTS

This research was supported by the Protein Research Unit Ruhr within Europe (PURE), Ministry of Innovation, Science and Research (MIWF) of North-Rhine Westphalia, Germany, and the European Regional Development Fund, European Union and North-Rhine Westphalia, Germany.

AUTHOR BIOGRAPHIES

Please see Supporting Information online.

ORCID

Sascha D. Krauß  <http://orcid.org/0000-0001-6271-9940>

Hesham K. Yosef  <http://orcid.org/0000-0002-2385-0046>

Tatjana Lechtonen  <http://orcid.org/0000-0002-8118-4715>

Samir F. El-Mashtoly  <http://orcid.org/0000-0001-6087-8817>

Klaus Gerwert  <http://orcid.org/0000-0001-8759-5964>

REFERENCES

- [1] F. M. Lyng, D. Traynor, I. R. M. Ramos, F. Bonnier, H. J. Byrne, *Anal. Bioanal. Chem.* **2015**, 407(27), 8279.
- [2] A. Sahu, S. Tawde, V. Pai, P. Gera, P. Chaturvedi, S. Nair, C. Murali Krishna, *Anal. Methods* **2015**, 7(18), 7548.
- [3] Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, H. Zeng, *Int. J. Cancer* **2003**, 107(6), 1047.
- [4] L. Mavarani, D. Petersen, S. F. El-Mashtoly, A. Mosig, A. Tannapfel, C. Kötting, K. Gerwert, *Analyst* **2013**, 138(14), 4035.
- [5] D. Petersen, L. Mavarani, D. Niedieker, E. Freier, A. Tannapfel, C. Kötting, K. Gerwert, S. F. El-Mashtoly, *Analyst* **2017**, 142(8), 1207.
- [6] F. B. Legesse, A. Medyukhina, S. Heuke, J. Popp, *Comput. Med. Imaging Graph.* **2015**, 43, 36.
- [7] S. D. Krauß, H. K. Yosef, T. Lechtonen, H. Jütte, A. Tannapfel, H. U. Kafferlein, T. Brüning, F. Roghmann, J. Noldus, S. F. El-Mashtoly, K. Gerwert, A. Mosig, *J. Chemom.* **2018**, 32(1), e2973.
- [8] A. Krizhevsky, I. Sutskever, G. E. Hinton, *Adv. Neural Inf. Process. Syst.* **2012**, 1097.
- [9] H. Mhaskar, Q. Liao, T. A. Poggio, *AAAI* **2017**, 2343.
- [10] G. J. S. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, C. I. Sánchez, *CoRR* **2017** <http://arxiv.org/abs/1702.05747>.
- [11] J. M. Schubert, B. Bird, K. Papamarkakis, M. Miljković, K. Bedrossian, N. Laver, M. Diem, *Lab. Investig.* **2010**, 90(7), 1068.
- [12] L. T. Kerr, K. Domijan, I. Cullen, B. M. Hennelly, *Photon. Lasers Med.* **2014**, 3(3), 193.
- [13] K. C. Schuster, E. Urlaub, J. R. Gapes, *J. Microbiol. Methods* **2000**, 42(1), 29.
- [14] U. Neugebauer, J. H. Clement, T. Bocklitz, C. Krafft, J. Popp, *J. Biophotonics* **2010**, 3(8–9), 579.
- [15] H. K. Yosef, S. D. Krauß, T. Lechtonen, H. Jütte, A. Tannapfel, H. U. Kafferlein, T. Brüning, F. Roghmann, J. Noldus, A. Mosig, S. F. el-Mashtoly, K. Gerwert, *Anal. Chem.* **2017**, 89(12), 6893.
- [16] S. F. El-Mashtoly, D. Niedieker, D. Petersen, S. D. Krauß, E. Freier, A. Maghnouj, A. Mosig, S. Hahn, C. Kötting, K. Gerwert, *Biophys. J.* **2014**, 106(9), 1910.
- [17] S. D. Krauß, D. Petersen, D. Niedieker, I. Fricke, E. Freier, S. F. El-Mashtoly, K. Gerwert, A. Mosig, *Analyst* **2015**, 140(7), 2360.
- [18] S. F. El-Mashtoly, D. Petersen, H. K. Yosef, A. Mosig, A. Reinacher-Schick, C. Kötting, K. Gerwert, *Analyst* **2014**, 139(5), 1155.
- [19] M. Harz, R. Petra, K.-D. Peschke, O. Ronneberger, H. Burkhardt, J. Popp, *Analyst* **2005**, 130(11), 1543.
- [20] M. Hedegaard, C. Krafft, H. J. Ditzel, L. E. Johansen, S. Hassing, J. Popp, *Anal. Chem.* **2010**, 82(7), 2797.
- [21] M. Hedegaard, C. Matthäus, S. Hassing, C. Krafft, M. Diem, J. Popp, *Theor. Chem. Accounts* **2011**, 130(4–6), 1249.
- [22] P. Sajda, S. Du, L. Parra, et al., *Proc. SPIE* **2003**, 321.
- [23] H. Nawaz, A. Garcia, A. D. Meade, F. M. Lyng, H. J. Byrne, *Analyst* **2013**, 138(20), 6177.
- [24] T. Tolstik, C. Marquardt, C. Matthäus, N. Bergner, C. Bielecki, C. Krafft, A. Stallmach, J. Popp, *Analyst* **2014**, 139(22), 6036.
- [25] S. F. El-Mashtoly, H. K. Yosef, D. Petersen, L. Mavarani, A. Maghnouj, S. Hahn, C. Kötting, K. Gerwert, *Anal. Chem.* **2015**, 87(14), 7297.
- [26] M. Isabelle, J. Dorney, A. Lewis, G. R. Lloyd, O. Old, N. Shepherd, M. Rodriguez-Justo, H. Barr, K. Lau, I. Bell, S. Ohrel, G. Thomas, N. Stone, C. Kendall, *Faraday Discuss.* **2016**, 187, 87.
- [27] I. P. Santos, P. J. Caspers, T. C. Bakker Schut, R. van Doorn, V. N. Hegt, S. Koljenovic, G. J. Puppels, *Anal. Chem.* **2016**, 88(15), 7683.
- [28] I. W. Schie, R. Kiselev, C. Krafft, J. Popp, *Analyst* **2016**, 141(23), 6387.
- [29] H. K. Yosef, L. Mavarani, A. Maghnouj, S. Hahn, S. F. El-Mashtoly, K. Gerwert, *Anal. Bioanal. Chem.* **2015**, 407(27), 8321.
- [30] D. Zink, A. H. Fischer, J. A. Nickerson, *Nat. Rev. Cancer* **2004**, 4(9), 677.
- [31] P. M. J. Moonen, L. A. L. M. Kiemeny, J. A. Witjes, *Eur. Urol.* **2005**, 48(6), 951.
- [32] R. Goodacre, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell, P. J. Rooney, et al., *Microbiology* **1998**, 144(5), 1157.
- [33] M. Gniadecka, P. A. Philipsen, S. Wessel, R. Gniadecki, H. C. Wulf, S. Sigurdsson, O. F. Nielsen, D. H. Christensen, J. Hercogova, K. Rossen, H. K. Thomsen, L. K. Hansen, *J. Investig. Dermatol.* **2004**, 122(2), 443.
- [34] A. Naumann, *Analyst* **2009**, 134(6), 1215.
- [35] Z. Lin, G. W. Small, A. S. Haka, L. H. Kidder, E. Neil Lewis, *Appl. Spectrosc.* **2003**, 57(1), 14.
- [36] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, *Deep Learning*, Vol. 1, MIT press, Cambridge **2016**.
- [37] M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, **2015**.
- [38] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Nature* **1986**, 323(6088), 533.
- [39] I. u. Rehman, Z. Movasaghi, S. Rehman, *Vibrational Spectroscopy for Tissue Analysis*, CRC Press, New York, USA, **2012**.
- [40] E. Wiercigroch, E. Szafraniec, K. Czamara, M. Z. Pacia, K. Majzner, K. Kochan, A. Kaczor, M. Baranska, K. Malek, *Spectrochim. Acta A* **2017**, 185, 317.
- [41] C. Sun, A. Shrivastava, S. Singh, A. Gupta, *2017 I.E. International Conference on Computer Vision (ICCV)*, IEEE, **2017**, p. 843.
- [42] L. Y. Pratt, *Adv. Neural Inf. Process. Syst.* **1993**, 204.
- [43] S. J. Pan, Q. Yang, *IEEE Trans. Knowl. Data Eng.* **2010**, 22(10), 1345.
- [44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, *CVPR 2009*, IEEE, **2009**, p. 248.
- [45] D. Mazia, G. Schatten, W. Sale, *J. Cell Biol.* **1975**, 66(1), 198.
- [46] T. Dieing, O. Hollricher, J. Toporski, *Confocal Raman microscopy*, Vol. 158, Springer Science & Business Media, Heidelberg, Germany, **2011**.
- [47] I. Guyon, A. Elisseeff, *J. Mach. Learn. Res.* **2003**, 3(March), 1157.
- [48] Y. Saeyns, I. Inza, P. Larrañaga, *Bioinformatics* **2007**, 23(19), 2507.
- [49] H. Peng, F. Long, C. Ding, *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, 27(8), 1226.
- [50] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, *Proceedings of the 22nd ACM International Conference on Multimedia*, ACM, **2014**, p. 675.
- [51] L. Breiman, *Mach. Learn.* **2001**, 45(1), 5.
- [52] M. V. Boland, R. F. Murphy, *Bioinformatics* **2001**, 17(12), 1213.
- [53] S. Guo, T. Bocklitz, U. Neugebauer, J. Popp, *Anal. Methods* **2017**, 9(30), 4410.
- [54] O. Russakovsky, J. Deng, S. Hao, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., *IJCV* **2015**, 115(3), 211.
- [55] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, *J. Mach. Learn. Res.* **2014**, 15(1), 1929.
- [56] F. Dan, J. Zhou, W. S. Zhu, P. W. Manley, Y. Karen Wang, T. Hood, A. Wylie, X. Sunney Xie, *Nat. Chem.* **2014**, 6(7), 614.
- [57] W. J. Tipping, M. Lee, A. Serrels, V. G. Brunton, A. N. Hulme, *Chem. Soc. Rev.* **2016**, 45(8), 2075.
- [58] C. Krafft, M. Schmitt, I. W. Schie, D. Cialla-May, C. Matthäus, T. Bocklitz, J. Popp, *Angew. Chem. Int. Ed.* **2017**, 56(16), 4392.
- [59] C. Krafft, G. Steiner, C. Beleites, R. Salzer, *J. Biophotonics* **2009**, 2(1–2), 13.
- [60] A. Kallenbach-Thieltges, F. Großerüschkamp, A. Mosig, M. Diem, A. Tannapfel, K. Gerwert, *J. Biophotonics* **2013**, 6(1), 88.
- [61] P. Bassan, M. J. Weida, J. Rowlette, P. Gardner, *Analyst* **2014**, 139(16), 3856.

How to cite this article: Krauß SD, Roy R, Yosef HK, et al. Hierarchical deep convolutional neural networks combine spectral and spatial information for highly accurate Raman-microscopy-based cytopathology. *J. Biophotonics*. 2018;e201800022. <https://doi.org/10.1002/jbio.201800022>