Research Article | Vol. 12, No. 4 / 1 April 2021 / Biomedical Optics Express 2236
Biomedical Optics EXPRESS

Auto-detection of cervical collagen and elastin in Mueller matrix polarimetry microscopic images using K-NN and semantic segmentation classification

CAMILO ROA,[1,5] V. N. DU LE,[2,5,6] MALA MAHENDROO,[3] ILYAS SAYTASHEV,[4] AND JESSICA C. RAMELLA-ROMAN[2,4,7]

[1]*Department of Biological Sciences, College of Arts, Sciences and Education, Florida International University, 11200 SW 8th Street, Miami, FL 33199, USA*
[2]*Department of Biomedical Engineering, College of Engineering and Computing, Florida International University, 10555 West Flagler Street, Miami, FL 33174, USA*
[3]*Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA*
[4]*Department of Ophthalmology, Herbert Wertheim College of Medicine, Florida International University, 11200 SW 8th Street, Miami, FL 33199, USA*
[5]*These authors contributed equally*
[6]*vidule@fiu.edu*
[7]*jramella@fiu.edu*

Abstract: We propose an approach for discriminating fibrillar collagen fibers from elastic fibers in the mouse cervix in Mueller matrix microscopy using convolutional neural networks (CNN) and K-nearest neighbor (K-NN) for classification. Second harmonic generation (SHG), two-photon excitation fluorescence (TPEF), and Mueller matrix polarimetry images of the mice cervix were collected with a self-validating Mueller matrix micro-mesoscope (SAMMM) system. The components and decompositions of each Mueller matrix were arranged as individual channels of information, forming one 3-D voxel per cervical slice. The classification algorithms analyzed each voxel and determined the amount of collagen and elastin, pixel by pixel, on each slice. SHG and TPEF were used as ground truths. To assess the accuracy of the results, mean-square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) were used. Although the training and testing is limited to 11 and 5 cervical slices, respectively, MSE accuracy was above 85%, SNR was greater than 40 dB, and SSIM was larger than 90%.

1. Introduction

Preterm birth (PTB), defined as any birth prior to 37 completed weeks of gestation, is responsible for 35% of the annual 3.1 million global neonatal deaths [1–3]. Many survivors will face life-long challenges including neurological disorders, long-term cognitive impairment, defects in hearing, vision, and digestion, as well as respiratory disease.

Unfortunately, there is an absence of clinical tools for early and accurate detection of spontaneous preterm birth risk, in part due to a lack in understanding of the molecular events that drive a term or preterm birth. Understanding the cervical remodeling process in a term or preterm pregnancy is critical to define therapeutic targets and to develop clinical tools. Given the demonstrated reorganization of the cervical extracellular matrix through pregnancy [4] and its direct correlation with mechanical function of the cervix [5], we aim in this study to develop improved methodology to characterize collagen and elastic fibers in the mouse cervix.

An extracellular matrix (ECM) lies under the epithelium of the cervix and consists of roughly 70% collagen fibers [6–8]. Numerous researchers have studied cervical collagen [9–16] and

#420079 https://doi.org/10.1364/BOE.420079
Journal © 2021 Received 19 Jan 2021; revised 10 Mar 2021; accepted 17 Mar 2021; published 23 Mar 2021

Biomedical Optics EXPRESS

its anisotropic alignment surrounding the cervical canal [5,8,17–20] through second harmonic generation and transmission electron microscopy (TEM) [21]. Collagen is the main load bearing component of the cervical stroma which undergoes substantial remodeling during pregnancy [5]. Collagen molecules are organized into fibrils through enzymatic cross-links. Mature cross-links density has been shown to decrease substantially in the mouse model between day 12 and day 18 of gestation resulting in increased collagen solubility and reduced tissue stiffness as pregnancy progresses [22].

Elastic fibers are also important components of the cervix ECM [5] and consist of elastin polymer cross-linked into microfibrils. In the cervix, elastin fibers are intertwined with collagen fibers and are primarily located in the subepithelial stroma. Elastic fibers are modified during pregnancy [21] becoming more dispersed and less ordered. While not well studied at the molecular level, a reduction in cross-link density of elastic fibers is correlated with preterm birth due to cervical insufficiency [23]. Assessment of elastin is complicated by its paucity compared to collagen in the cervix. TPEF, TEM, and fluorescence confocal microscopy with immunofluorescence [24] are among the most effective modalities for elastin assessment in the cervical environment.

Previous studies have shown that polarization-based microscopy techniques enhanced by Sirius Red staining offer similar imaging capability of cervical collagen [12] as compared to SHG based modalities. Mueller matrix polarimetry has also been used extensively to image the cervix and its collagen arrangement [20,24],–27. To our knowledge polarization optical modalities have not been explored in the assessment of the cervical elastic fibers. The use of the Mueller matrix modality, that does not require complex instrumentation, or specialized staining could accelerate the development of clinical tools for assessment of cervical function and risk of malfunction. This study utilizes machine learning and a Self-validating Mueller matrix Micro–Mesoscope (SAMMM) for the characterization of elastic fibers in the mouse cervix.

Analyzing and diagnosing large numbers of samples require the use of machine learning techniques. Although different classification methods have been applied to Mueller matrix imaging [28–30], these methods have not been used to visually separate elastin from collagen in the cervix. We propose two frequently used machine vision methods [31] to detect and classify collagen and elastin fibers, based on the Mueller matrix components: one method is the use of a supervised K-nearest neighbor (K-NN) algorithm [32–34] to associate pixels with similar characteristics into defined groups using the Mueller matrix components as features, the other one is a deep learning semantic segmentation neural network that assigns pixels into different layers depending on their intensity. Deep learning segmentation is commonly used in biomedical imaging for a wide variety of diagnostics [35–39]. Such networks usually have higher accuracy and are more robust than other classifiers but require more time and samples for training [40]. To reduce training time, a knowledge transfer from a well-tested CNN (Resnet18) is used.

2. Material and methods

2.1. Self-validating Mueller matrix micro-mesoscope (SAMMM) system

In this study, the reflectance measurements of SHG, TPEF, and near-infrared total reflectance images are performed using the SAMMM system described in [41,42]. Note that only total reflectance was used to study polarimetry in this work. The excitation source is a pre-compensated mode locked laser beam from a Ti-Sapphire broadband femtosecond laser with the central wavelength at 800 nm (FWHM=100 nm). SHG signal at 400 (FWHM=30 nm) and TPEF at 500 (FWHM=20 nm), and total reflectance images are collected by appropriated photo-multiplier tube detectors and a single data acquisition board sampled up to 125MHz (Vidrio Technologies LLC, VA). Each Mueller matrix is constructed utilizing a polarization state analyzer (PSA of four polarization states) and a polarization state generator (PSG of six polarization states), resulting in a set of 24 images in each of the three channels (SHG, TPEF and total reflectance). All SAMMM

Biomedical Optics EXPRESS

raw images of the mice cervix had a resolution of 1000×1000 pixels and were taken with a 5X objective.

2.2. Image decomposition

The Mueller matrix M of the medium were decomposed using the Lu-Chipman decomposition method in which M can be expressed as the product of three basic matrices [43]:

$$M = M_{\Delta} M_R M_D \tag{1}$$

M_{Δ}, M_{R} and M_{D} are depolarization, retardance and diattenuation Mueller matrices, respectively. Following the decomposition, scalar terms such as diattenuation (d), depolarization coefficient (Δ), linear retardation (δ) of the medium are determined:

$$d = \frac{\sqrt{M_D^2(1,2) + M_D^2(1,3) + M_D^2(1,4)}}{M_D(1,1)} \tag{2}$$

$$\Delta = 1 - \frac{|\mathrm{Tr}\, M_{\Delta} - 1|}{3} \tag{3}$$

$$\delta = cos^{-1}\left(\sqrt{\left(M_R^2(2,2) + M_R^2(3,3)\right)^2 + \left(M_R^2(3,2) - M_R^2(2,3)\right)^2} - 1\right) \tag{4}$$

In addition to polar decomposition, differential matrix formalism of the Mueller calculus was used to retrieve total retardance (linear and circular) [44]. In this case, the medium polarization properties are contained in a single differential matrix m which relates the Mueller matrix M and its spatial derivative along the propagation of light [45,46]:

$$\frac{dM}{dz} = mM \tag{5}$$

Applying Lorentz symmetric matrices, L_m and Lorentz antisymmetric matrices, L_u to Eq. (5), we have:

$$\ln M = mz = L = L_u + L_m \tag{6}$$

where

$$L_m = \frac{1}{2}(L - GL^T G) \tag{7}$$

$$L_u = \frac{1}{2}(L + GL^T G) \tag{8}$$

In Eqs. (7–8), G is Minkowski metric tensor. For a depolarizing medium, the off-diagonal elements of L_m represents mean values of the elementary medium polarization properties over the path-length z and the off-diagonal elements of L_u express their respective uncertainties. Lorentz components of the matrices is used to retrieve both linear (δ_L), circular (δ_C) and total (R) retardation, and angle of orientation (θ)

$$\delta_C = \frac{L_m(2,3)}{2} \tag{9}$$

$$\delta_L = \sqrt{(L_m(2,4))^2 + (L_m(3,4))^2} \tag{10}$$

$$R = \sqrt{{\delta_L}^2 + 4{\delta_C}^2} \tag{11}$$

$$\theta = \frac{1}{2}\tan^{-1}\frac{\{L_m(2,4)\}}{\{L_m(3,4)\}} \tag{12}$$

Biomedical Optics EXPRESS

2.3. Section preparation for SAMMM imaging

In this paper, a total of 16 cervical slices were obtained from 5 mice, including one non-pregnant mice, two pregnant mice at gestation day 6 and two at gestation day 18 samples in accordance with the Institutional Animal Care and Use Committee protocol. The tissue was snapped frozen at -80° C in optimal cutting temperature (O.C.T.) compound (Tissue Tek, Elkhart, Indiana). The entire length of the cervix was cryosectioned transversely at -20° C using a cryostat (Leica CM3050). Sections were mounted on glass slides and left dry for 1 hour at room temperature. Unwanted residues were washed away with phosphate buffered saline (PBS).

2.4. Confirmation of elastin on TPEF images with immunofluorescence

While SHG of collagen in the cervix is optimal at wavelength region of 400 nm region [47,48] with the excitation at 800 nm, elastin and NADH has an overlapping fluorescence emission at 500 nm region. Other glycosaminoglycans (GAGs) and proteoglycans could also contribute to the TPEF signal. Therefore, in order to confirm that cervical elastin is responsible for TPEF signal in our samples, we performed indirect immunofluorescence of a selected slice using rabbit anti-mouse tropoelastin antibody (Elastin Products Company, PR385) as the primary antibody and Alexa Fluor 546–conjugated antibody (Life Technologies, A11035) as the secondary antibody. Slides were washed with 20mM Tris (pH = 8.0) for 15 minutes followed by treating the section with 100 mM iodoacetamide (Sigma-Aldrich Inc., I5161) in the dark for 15 minutes.

Diluted goat serum at 10% (ThermoFisher Scientific, 31872) was used to block the section for 1 hour at room temperature. The section was then incubated with 1:250 dilution of primary antibody in 1% goat serum overnight at 4°C. Finally, the section was washed with PBS and incubated in 1:500 dilution of secondary antibody in 1% goat serum for 30 minutes at room temperature. The section was imaged using a commercial linear microscope (Olympus BX61) with laser at 550 nm and a 10X objective. As shown in Fig. 1, TPEF signal (Fig. 1(a)) obtained with SAMMM shows strong correlation to the immunofluorescence images of elastin (Fig. 1(b)-(e)). Following this study, TPEF images of the mice cervix in this study were used as ground truth for elastin.

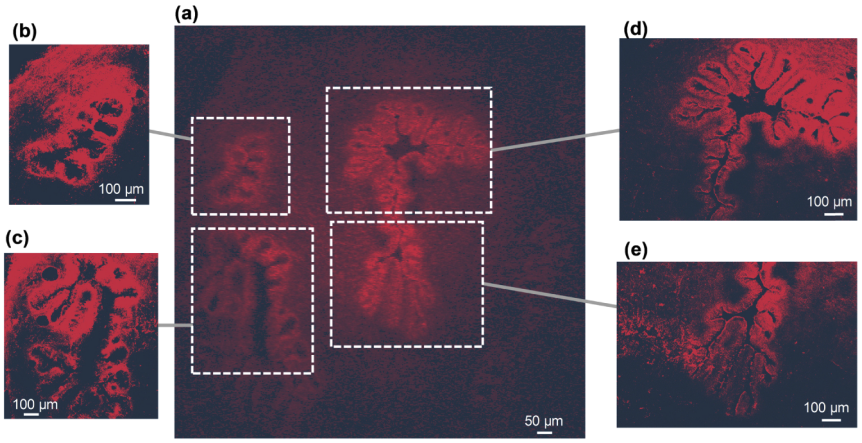

Fig. 1. Immunofluorescence confirms elastin on TPEF image: (a) TPEF image of an unstained mice cervix at day 18 pregnancy using SAMMM, (b-e) immunofluorescent images of the same slice using a commercial linear microscope.

2.5. Data processing

In Fig. 2, the process to detect collagen or elastin from Mueller matrix data and its decompositions is presented. The process extracts the ground truth (GTruth), reduces the cervix images

Biomedical Optics EXPRESS

to a 200×200 pixels images, normalizes them, extracts features, applies two classification methodologies, and compares the results for accuracy. Two commonly used machine vision classifiers [31] were chosen: K-NN and semantic segmentation neural networks. Both methods are extensively used to classify images [32,49], being neural networks especially useful for imaging diagnostics [35,36,39].

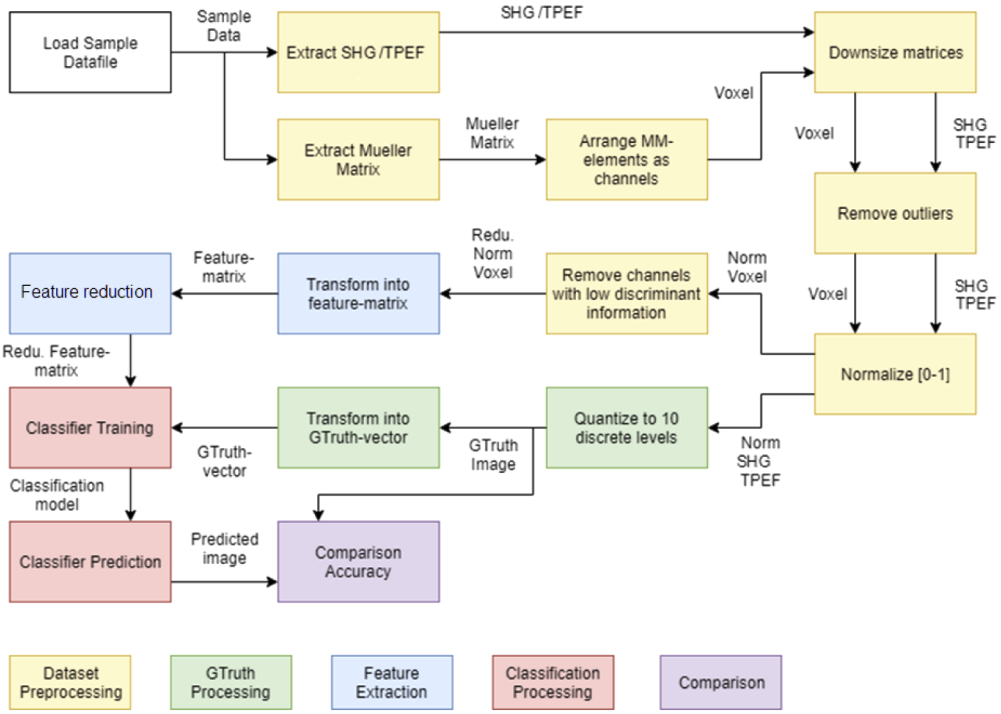

Fig. 2. Block diagram of an algorithm to detect and classify collagen/elastin samples.

2.5.1. Preprocessing

For each data sample, the corresponding SHG, and TPEF are obtained. SHG data corresponds to the ground truth of collagen, while TPEF corresponds to the one of elastin. Each ground truth is represented by one image in which the intensity is related with the density of its respective tissue.

The Mueller matrix elements (M_{11}, M_{12}, ..., M_{44}) of data samples, along with its decomposition values, are arranged as a voxel were the channels of information correspond to the elements or the decompositions. All elements and decompositions are aligned so a specific pixel has 26 initial channels of information. Some of the voxel's channels do not bring useful discriminating information. Channels with very low standard deviation are removed since they are mostly noise. Additionally, information on some channels is very similar to others; redundant channels were also removed after comparing them with each other using structural similarity index [50] (Fig. 3). The following variables were used as data channels of the voxel: diattenuation (Eq. (2)), depolarization coefficient (Eq. (3)), linear retardation (Eq. (4)), total retardation (Eq. (11)), orientation (Eq. (12)), and all Mueller matrix elements except M_{22}, M_{33}, and M_{44}.

SHG, TPEF, and data voxels are reduced in resolution to 200×200 pixels images, so the computation is more efficient. Due to errors in the sampling process or data gathering, outlier pixels are present. Pixels with very high values compared to their neighbors can skew the classification process. Outliers were removed from the voxel, the SHG, and the TPEF data using

Biomedical Optics EXPRESS

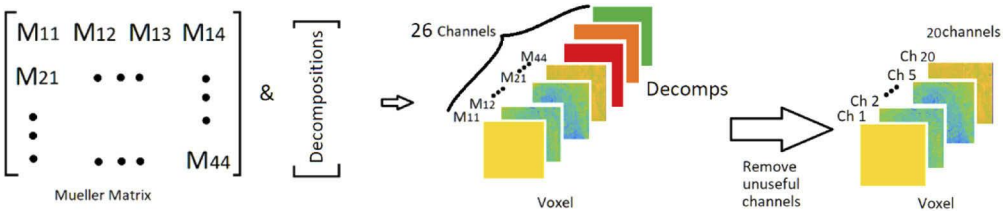

Fig. 3. From Mueller matrix and decompositions to voxel with data channels.

the *Grubbs* method. The outlier pixels were replaced by the mean of their neighbors. Raw values among images varied greatly, so, to make them comparable to each other, all images were normalized independently, based on the maximum and minimum values of each one. The value of any pixel within an image was then between 0 and 1.

2.5.2. Ground truth processing

Although a binary (0 or 1) classification is possible, the SHG and TPEF were quantized into 10 relative discrete levels representing degrees of density of collagen or elastin. These levels were associated to normalized values between [0-1] and can be scaled to fit any range. High levels (from 6 to 10) were associated with a detection of collagen or elastin at different densities. Maintaining this resolution was important to preserve image details. In addition, background, or pixels from other tissues are associated to low levels (from 1-4). All values of SHG and TPEF were rounded to one of the10 discrete values to obtain an appropriate ground truth for classification (Fig. 4). A non-uniform quantization was used to account for the imbalance between the number of low intensity pixels and the number of high intensity pixels and to introduce bias to the classifiers. The 10-level quantization grid favored high intensity pixels by starting high levels at lower intensity values. The grid's thresholds were [0, 0.05, 0.11, 0.19, 0.29, 0.36] for low levels and [0.47, 0.60, 0.85, 0.96, 1] for high levels.

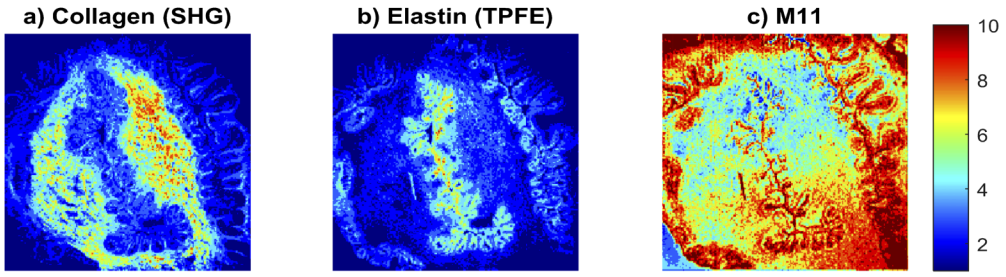

Fig. 4. Cervical cross-section from a mouse gestation on day 18. Images for a) collagen ground truth (SHG), b) elastin ground truth (TPEF), and c) normalized total reflectance intensity M_{11}. Ground truth densities are represented on a 10-level heat scale; normalized M_{11} values, that are between [0-1], are represented on the same 10-level scale for comparison purposes.

For the K-NN case, the ground truth image was arranged as a vector of GTruth levels, the vector position's index was associated with the position of a pixel in the ground truth image (Fig. 5).

For the semantic segmentation case, each level of the ground truth acted as a layer superimposed on the sample data. The ground truth was an image of the same size as the input data with categorical values (levels) for each pixel, in essence, equal to the ground truth of the K-NN classifier.

Biomedical Optics EXPRESS

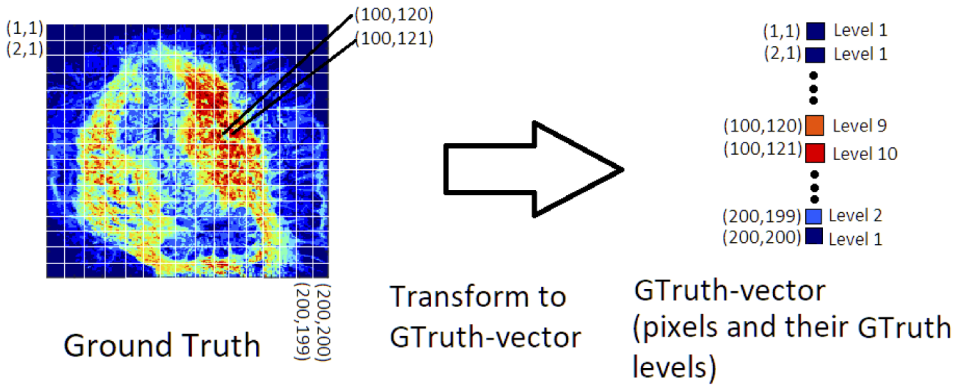

Fig. 5. Transform ground truth image to GTruth-vector. Collagen data sample.

2.5.3. Feature extraction

K-NN classifiers take several observations and try to group them together based on the similarity of their features. A standard input for a K-NN classifier is a single matrix in which the rows represent independent observations, and the columns are features associated with those observations. The voxel was arranged into a matrix of features where a row of the matrix was linked to a pixel and the columns were its corresponding channels (features). The number of features was reduced further using principal component analysis (PCA). By taking the features that compose 95% of the discrimination information the number of the features reduced to 4 (Fig. 6).

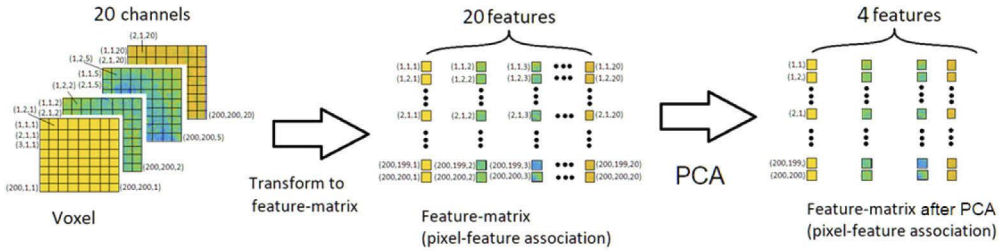

Fig. 6. Transformation from voxel to feature-matrix for K-NN classifier.

For the semantic segmentation classifier, data channels were reduced to only Linear retardation, and Mueller matrix elements M_{34} and M_{43} since they were the most dissimilar among channels and carried more discriminant (classification) information. Thousands of features were extracted based on the pretrained *Resnet18* and *DeepLab v3+* [51] architectures (Fig. 7).

2.5.4. Classification training

There was a total of 16 mice cervical slices, 11 slices were used for training and the remaining for testing. However, since the classifiers take each pixel vector (pixels at the same coordinate position in all channels) as one observation and there are 40,000 pixel vectors per cervical slice, the set of observations for the classifiers to train and test on was close to 320,000 and 120,000, respectively. Those two sets were independent from each other.

Two K-NN (k=5 neighbors) classifiers were trained based on the collagen and elastin ground truths, respectively. The algorithm produced two independent models, one to detect collagen and one to detect elastin.

Biomedical Optics EXPRESS

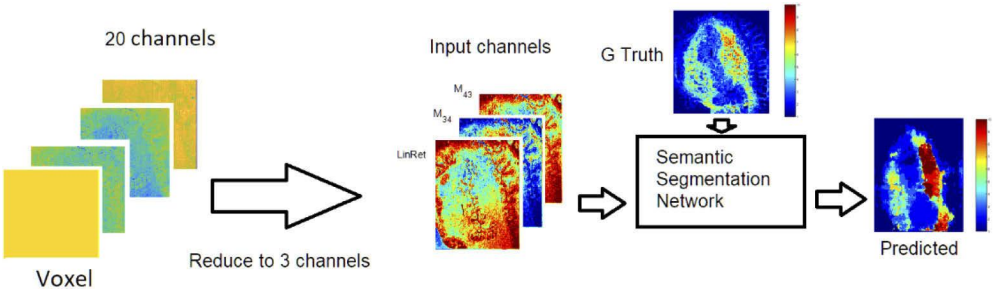

Fig. 7. Transformation from voxel to RGB for Semantic Segmentation CNN classifier.

To save computational time and reduce the amount of training data in the semantic segmentation case, transfer information from a pre-trained *Resnet18* into a *DeepLab v3+* [51] semantic segmentation architecture was done. Two models, one to detect collagen and the other elastin, were trained interpedently with 11 cervical slices. The networks used encoder-decoder architectures, dilated convolutions, and skip connections to segment images. Data was artificially augmented to increase the number of samples (images) to a similar number than in the K-NN case; random left/right reflection, and random X/Y translation of +/- 10 pixels were applied.

2.6. Collagen and elastin predictions

The two classifying methodologies were applied to each of the cervical slices. Intensity levels from 1-10 represent the amount of tissue predicted, where 1 is no amount and 10 is a high amount. The outputs of both, the K-NN model and the semantic segmentation model, were a predicted pixel-level image of the same dimensions than the original inputs. Some classification errors were seen as static noise (salt and pepper noise).

A median filter was used to smooth the noise on the K-NN prediction in postprocessing. The image was more or less granular depending on the size of the filtering window. Two window sizes of $[3 \times 3]$ and $[10 \times 10]$ pixels were chosen. For the semantic segmentation case, no filter was used due to low noise level in the output images.

2.7. Accuracy

The original ground truth and the predicted image were compared for measuring image quality, using three different methods: mean-squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) index [50]. MSE and PSNR are purely mathematical metrics and may not agree with human perception of the image quality. SSIM considers local contrast and luminance and agrees more closely with a subjective metric.

Utilizing the MATLAB built-in function "immse.m", MSE metric compared the data matrices of the original and predicted images, pixel by pixel, then calculated the mean-squared errors for each pixel pair, averaged them, and subtracted the value from 1. Meanwhile, function "psnr.m" was used to find the peak signal-to-noise ratio which indicates the ratio of the maximum pixel intensity to the power of the distortion. It compares the original and predicted images as if the classifier were a transmission system introducing noise, the original images were a transmitted signal, and the predicted images were a received signal. An SNR > 40 dB is considered good, SNR <10 dB is considered poor.

MATLAB built-in function "ssim.m" was used to find the structural similarity index which combines local image structure, luminance, and contrast into a single local quality score given as a percentage. Structures are selected as patterns of pixel intensities, particularly among neighboring pixels.

Biomedical Optics EXPRESS

3. Results

Section 3.1 presents visual results for 2 tested slices from day 18 (D18) and day 6 (D6); for collagen and elastin predictions; and for K-NN and Semantic Segmentation classifiers. Section 3.2 presents the average of all the test samples for each of the three metrics, for the semantic segmentation and the K-NN classifiers with filters. Overall accuracies, involving all the samples, have similar results but are not presented.

On an Intel Core i7, 3.40 GHz with 32 GB of RAM, the K-NN model took 5 minutes to train and 30 minutes to predict a sample. On the contrary, semantic segmentation took substantially more time to train (around 60 hours) but much faster prediction time (3 seconds per sample).

Data used for classification consisted of the normalized Mueller matrix component and the decomposition described in section 2.5.1. Ground truths SHG and TPEF, for samples D18 and D6, along with linear retardation, M_{11}, M_{34} and M_{43} features are presented in Fig. 8. These features carried the most discriminant information and are presented on a 10-level heat scale.

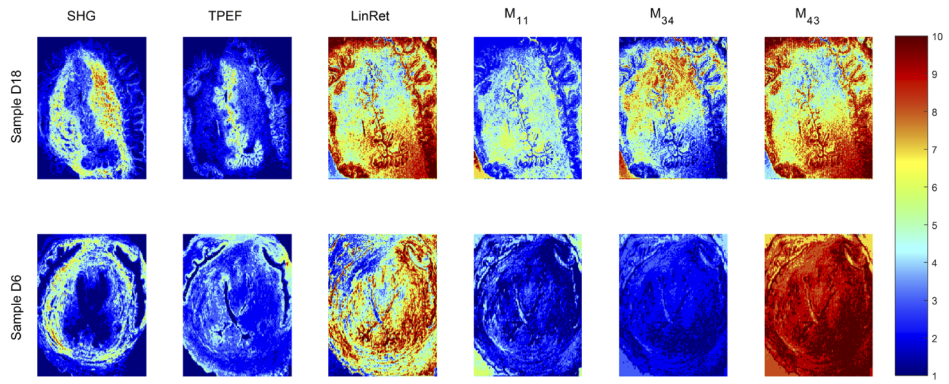

Fig. 8. Ground truth for collagen (SHG), ground truth for elastin (TPEF), normalized linear retardation (LinRet), and normalized Mueller matrix component M_{11}, M_{34}, and M_{43}, for samples D18 and D6. Ground truth densities are represented on a 10-level heat scale. Normalized LinRet, M_{11}, M_{34}, and M_{43} values, that are between [0-1], are represented on the same 10-level scale for comparison.

3.1. Classifiers qualitative results

Classification results of two tested slices for collagen and elastin are presented in Fig. 9 and Fig. 10, respectively. In each figure, the first column presents the image that was used as a ground truth (SHG for collagen and TPEF for elastin); the second column presents the K-NN prediction pixel-by-pixel where some classification errors are seen as salt-and-pepper noise. The following two columns show the K-NN prediction after two different filters. The last column presents the semantic segmentation prediction. All images were normalized, and pixels scaled from 1 to 10 levels, with 1 defined as low density of the tissue and 10 a high density of the tissue.

In general, structures on collagen predictions present less noise than structures on elastin predictions. K-NN predictions had significant salt-and-pepper noise that was removed by a mean $[3 \times 3]$ filter. The K-NN prediction with mean filter of $[10 \times 10]$ provides general areas in which the tissue structures are contained, but lacks resolution. The semantic segmentation prediction does not require filtering, presenting general areas containing the tissue while providing good resolution.

Biomedical Optics EXPRESS

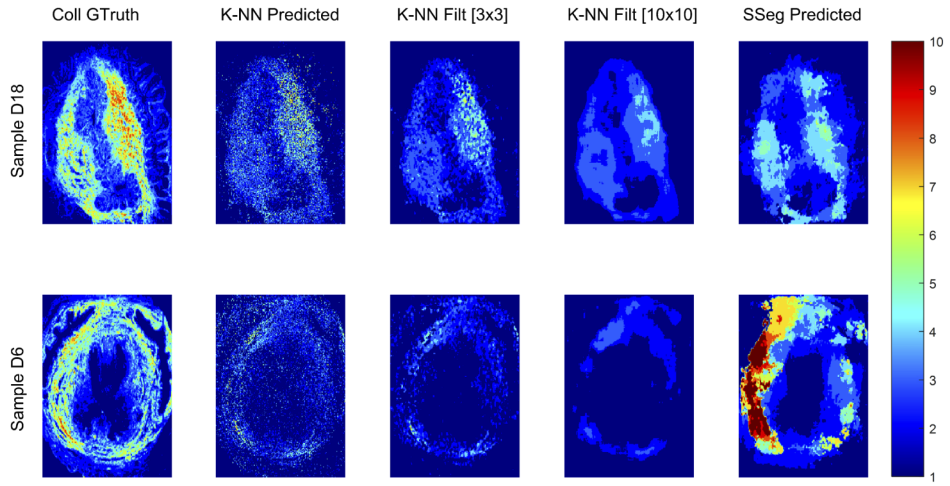

Fig. 9. Collagen classification of two samples (D18 and D6) using a K-NN algorithm and semantic segmentation classification ground truth images (GTruth column), K-NN predicted images (K-NN Predicted column), K-NN predicted images with a median filter $[3 \times 3]$ applied (K-NN Filt $[3 \times 3]$ column), K-NN predicted images with a median filter $[10 \times 10]$ applied (K-NN Filt $[10 \times 10]$ column) and semantic segmentation predicted images (SSeg Predicted column) are presented. Intensity reflects the density of the respective tissue in the sample and is presented on a 10-level heat scale.

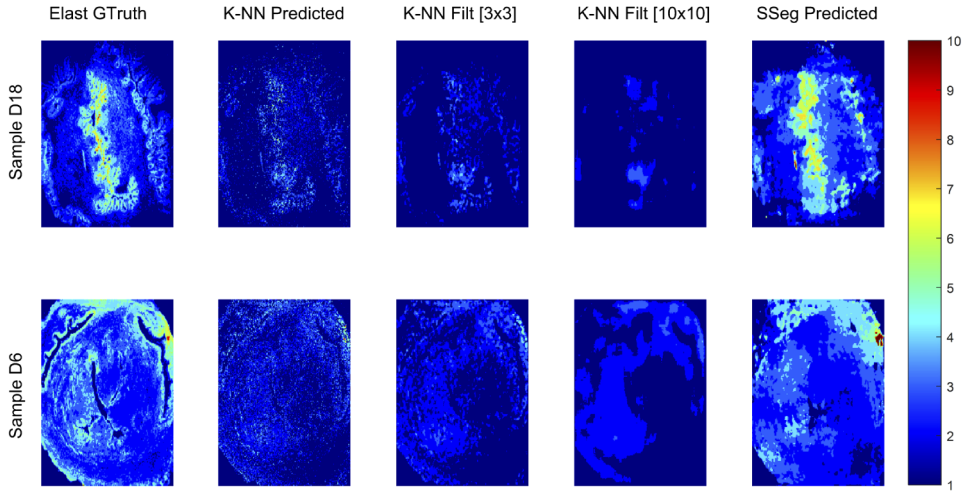

Fig. 10. Elastin classification of two samples (D18 and D6) using a K-NN algorithm and semantic segmentation classification ground truth images (GTruth column), K-NN predicted images (K-NN Predicted column),K-NN predicted images with a median filter $[3 \times 3]$ applied (K-NN Filt $[3 \times 3]$ column), K-NN predicted images with a median filter $[10 \times 10]$ applied (K-NN Filt $[10 \times 10]$ column) and semantic segmentation predicted images (SSeg Predicted column) are presented. Intensity reflects the density of the respective tissue in the sample and is presented on a 10-level heat scale.

Biomedical Optics EXPRESS

3.2. Accuracy

Table 1 summarizes the classification results for independent classification of collagen and elastin on the test set (6 cervical slices for testing). All accuracies are above 80% for collagen cases, while elastin cases have accuracies close to 90%. In overall, semantic segmentation gives slightly better results, especially for the SSIM metric.

Table 1. Accuracy of K-NN and semantic segmentation classifiers

Fiber	Classifier method	Accuracy metrics		
		MSE (%)	PSNR (dB)	SSIM (%)
Collagen	K-NN	82.33	41.28	85.25
	K-NN + [3 × 3] filter	84.56	41.84	83.95
	K-NN + [10 × 10] filter	84.40	41.79	83.06
	Semantic Seg.	87.60	42.93	92.89
Elastin	K-NN	91.08	44.05	87.34
	K-NN + [3 × 3] filter	91.31	44.12	85.56
	K-NN + [10 × 10] filter	91.48	43.68	83.84
	Semantic Seg.	91.27	43.76	90.27

4. Discussion and conclusions

We proposed a methodology that integrates Mueller matrix polarimetry, and convolutional neural networks (CNN) and K-nearest neighbor (K-NN) techniques for successfully detecting and classifying cervical collagen and elastin. The methodology is reliable and low cost.

Original digitalized samples have a large size that can cause a significant increase in computing time, especially for K-NN classifiers. We demonstrated that these images can be reduced to 200×200 pixels images without sacrificing accuracy. The reduction makes processing, training, and classification more manageable. Some Mueller matrix components or decompositions do not provide relevant classification information (depolarization with differential method, total retardation, and Mueller matrix components M_{22}, M_{33} and M_{44}). Simple feature reduction techniques, like average standard deviation or PCA, can reduce the components used and the size of the samples to no detriment in the accuracy. The K-NN model is quick to train but significantly slower in making predictions. On the contrary, semantic segmentation takes substantially more time to train but is very fast predicting. Depending on the problem, each method can be used independently to mutually confirm their results. A CNN-K-NN hybrid is also proposed as a next step. Additionally, with a larger sample size, a more complex neural network, like U-net [52], can be trained.

The use of images of SHG and TPEF as ground truths proved to be effective. However, outlier pixels from the initial sampling process can skew the quantization and introduce error to the classification. In addition, the ratio of high-level pixels versus low-level pixels is very small, even after normalization, representing a bias towards low-level pixels detections. A non-uniform quantization grid was used to compensate for this bias. Further work can be focused on improving the imbalance between high-level and low-level pixels by or applying class noise reduction methodologies [53].

Predicted collagen images have greater qualitative similarity to the ground truth than elastic fiber images. This is in part because there is more collagen protein in the samples than elastin, rendering more pixels with higher concentration associated to collagen than ones associated to elastin. In addition, collagen structures occupy a larger area making them easier to detect. However, accuracies for collagen and elastin, based on independently detecting them on the

Biomedical Optics EXPRESS

same sample, are comparable. Each pixel of a sample would have a collagen and an elastin level associated after classification; those levels are not mutually exclusive and sometimes overlap. That means that there could be a discrepancy on the classification which would need a joint classification approach (assigning likelyhood of collagen or elastin to each pixel), or it could mean that both tissues are present at the same time since collagen and elastin fibers are frequently intertwined. This is a limitation of independently classifiying collagen and elastic fibers and will be addressed in future work.

Measuring image quality is difficult since some metrics do not match subjective perception. Furthermore, the number of pixels representing collagen or elastin is significantly smaller compared to the "background" pixels which is why some accuracy metrics are relatively high, but the ground truth and predicted images do not look that similar. The classifiers are good at detecting the low-intensity "background" pixels, which are the majority, but they are not as good at properly classifying high-intensity pixels. Accuracy metrics that consider this imbalance or that favor subjective quality (like SSIM) are better suited for this kind of classification assessment.

A larger number of cervix slices is strongly recommended, especially for the semantic segmentation case, however, estimating the exact number of data samples needed to successfully train a neural network is very difficult if not impossible. The minimum number of samples depends directly on the characteristics of the problem and the chosen CNN architecture. Some magnetic resonance imaging (MRI) diagnostic studies required more than 3,000 samples [52] while others close to 300 [54,55]. Although a small number of samples is a limitation, the three accuracy metrics used in this study show overall good results compared to other approaches [30]. Notably, comparing to K-NN, the semantic segmentation classifiers are more robust and less sensitive to noise.

Mueller matrix microscopy presents several advantages compared to other modalities used for the quantification of collagen and elastin such as nonlinear microscopy or SHG. The modality is relatively low cost, easy to use, fast and can be designed with low encumbrance. Combined with machine learning techniques, this modality could expand the toolkit for researchers studying the reproductive system and particularly preterm labor. Training of the system with SHG and TPEF is necessary with our proposed approach, yet beyond the training phase, classification of cervical elastin and collagen can be achieved through a Mueller matrix system alone. Future work will focus on expanding this approach to a standalone system, i.e. systems that are not co-registered as the SAMMM.

Funding. National Science Foundation (DMR 1548924).

Acknowledgments. This work was also supported by Herbert Wertheim College of Medicine and the Herbert and Nicole Wertheim Professorship Endowment. We are grateful to Dr. Timothy Allen for access to the Olympus microscope. We thank Dr. Mariano Colon-Caraballo at UT Southwestern Medical Center for guidance on elastin immunohistochemistry.

Disclosures. The authors declare no conflicts of interest.

References

1. H. Blencowe, S. Cousens, M. Z. Oestergaard, D. Chou, A.-B. Moller, R. Narwal, A. Adler, C. V. Garcia, S. Rohde, and L. Say, "National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications," Lancet **379**(9832), 2162–2172 (2012).
2. A. Dbstet, "WHO: recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths," Acta. Obstet Gynecol. Scand. **56**, 247–253 (1977).
3. L. Liu, H. L. Johnson, S. Cousens, J. Perin, S. Scott, J. E. Lawn, I. Rudan, H. Campbell, R. Cibulskis, and M. Li, "Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000," Lancet **379**(9832), 2151–2161 (2012).
4. S. Nallasamy and M. Mahendroo, *Distinct Roles of Cervical Epithelia and Stroma in Pregnancy and Parturition* (Thieme Medical Publishers), pp. 190–200.
5. K. Yoshida, C. Jayyosi, N. Lee, M. Mahendroo, and K. M. Myers, "Mechanics of cervical remodelling: insights from rodent models of pregnancy," Interface Focus. **9**(5), 20190026 (2019).
6. M. House, D. L. Kaplan, and S. Socrate, "Relationships between mechanical properties and extracellular matrix constituents of the cervical stroma during pregnancy," Semin Perinatol. **33**(5), 300–307 (2009).

Biomedical Optics EXPRESS

7. K. Myers, S. Socrate, D. Tzeranis, and M. House, "Changes in the biochemical constituents and morphologic appearance of the human cervical stroma during pregnancy," Eur. J. Obstet. Gynecol. Reprod. Biol. **144**(1), S82–89 (2009).
8. W. Yao, Y. Gan, K. M. Myers, J. Y. Vink, R. J. Wapner, and C. P. Hendon, "Collagen fiber orientation and dispersion in the upper cervix of non-pregnant and pregnant women," PLoS One **11**, e0166709 (2016).
9. L. Shi, S. Q. Shi, G. R. Saade, K. Chwalisz, and R. E. Garfield, "Changes in cervical resistance and collagen fluorescence during gestation in rats," J. Perinat. Med. **27**(3), 188–194 (1999).
10. H. Maul, G. Olson, C. T. Fittkow, G. R. Saade, and R. E. Garfield, "Cervical light-induced fluorescence in humans decreases throughout gestation and before delivery: Preliminary observations," Am. J. Obstet. Gynecol. **188**(2), 537–541 (2003).
11. M. L. Akins, K. Luby-Phelps, R. A. Bank, and M. Mahendroo, "Cervical softening during pregnancy: regulated changes in collagen cross-linking and composition of matricellular proteins in the mouse," Biol. Reprod. **84**(5), 1053–1062 (2011).
12. S. Bancelin, A. Nazac, B. H. Ibrahim, P. Dokladal, E. Decenciere, B. Teig, H. Haddad, H. Fernandez, M. C. Schanne-Klein, and A. De Martino, "Determination of collagen fiber orientation in histological slides using Mueller microscopy and validation by second harmonic generation imaging," Opt. Express **22**(19), 22561–22574 (2014).
13. R. A. Word, X. H. Li, M. Hnat, and K. Carrick, "Dynamics of cervical remodeling during pregnancy and parturition: mechanisms and current concepts," Semin. Reprod. Med. **25**(1), 069–079 (2007).
14. R. Holt, B. C. Timmons, Y. Akgul, M. L. Akins, and M. Mahendroo, "The molecular mechanisms of cervical ripening differ between term and preterm birth," Endocrinology **152**(3), 1036–1046 (2011).
15. L. M. Reusch, H. Feltovich, L. C. Carlson, G. Hall, P. J. Campagnola, K. W. Eliceiri, and T. J. Hall, "Nonlinear optical microscopy and ultrasound imaging of human cervical structure," J. Biomed. Opt. **18**(3), 031110 (2013).
16. K. M. Myers, S. Socrate, A. Paskaleva, and M. House, "A study of the anisotropy and tension/compression behavior of human cervical tissue," J. Biomech. Eng. **132**(2), 021003 (2010).
17. R. M. Aspden, "Collagen organisation in the cervix and its relation to mechanical function," Collagen Relat. Res. **8**(2), 103–112 (1988).
18. K. M. Myers, H. Feltovich, E. Mazza, J. Vink, M. Bajka, R. J. Wapner, T. J. Hall, and M. House, "The mechanical role of the cervix in pregnancy," J. Biomech. **48**(9), 1511–1523 (2015).
19. M. Fernandez, M. House, S. Jambawalikar, N. Zork, J. Vink, R. Wapner, and K. Myers, "Investigating the mechanical function of the cervix during pregnancy using finite element models derived from high-resolution 3D MRI," Comput. Methods Biomech Biomed Eng. Imaging Vis. **19**(4), 404–417 (2016).
20. K. Yoshida, H. Jiang, M. Kim, J. Vink, S. Cremers, D. Paik, R. Wapner, M. Mahendroo, and K. Myers, "Quantitative evaluation of collagen crosslinks and corresponding tensile mechanical properties in mouse cervical tissue during normal pregnancy," PLoS One **9**(11), e112391 (2014).
21. A. Blaustein and R. J. Kurman, *Blaustein's Pathology of the Female Genital Tract* (Springer, 2011).
22. M. Mahendroo, "Cervical remodeling in term and preterm birth: insights from an animal model," Reproduction **143**(4), 429–438 (2012).
23. S. W. Lee, J. Y. Yoo, J. H. Kang, M. S. Kang, S. H. Jung, Y. Chong, D. S. Cha, K. H. Han, and B. M. Kim, "Optical diagnosis of cervical intraepithelial neoplasm (CIN) using polarization-sensitive optical coherence tomography," Opt. Express **16**(4), 2709–2719 (2008).
24. S. Nallasamy, K. Yoshida, M. Akins, K. Myers, R. Iozzo, and M. Mahendroo, "Steroid hormones are key modulators of tissue mechanical function via regulation of collagen and elastic fibers," Endocrinology **158**(4), 950–962 (2017).
25. A. Pierangelo, S. Manhas, A. Benali, C. Fallet, J. L. Totobenazara, M. R. Antonelli, T. Novikova, B. Gayet, A. De Martino, and P. Validire, "Multispectral Mueller polarimetric imaging detecting residual cancer and cancer regression after neoadjuvant treatment for colorectal carcinomas," J. Biomed. Opt. **18**(4), 046014 (2013).
26. M. Anastasiadou, A. D. Martino, D. Clement, F. Liège, B. Laude-Boulesteix, N. Quang, J. Dreyfuss, B. Huynh, A. Nazac, L. Schwartz, and H. Cohen, "Polarimetric imaging for the diagnosis of cervical cancer," Phys. Status Solidi C **5**(5), 1423–1426 (2008).
27. A. Nazac, S. Bancelin, B. Teig, B. H. Ibrahim, H. Fernandez, M. C. Schanne-Klein, and A. De Martino, "Optimization of Picrosirius red staining protocol to determine collagen fiber orientations in vaginal and uterine cervical tissues by Mueller polarized microscopy," Microsc. Res. Tech. **78**(8), 723–730 (2015).
28. C. Heinrich, J. Rehbinder, A. Nazac, B. Teig, A. Pierangelo, and J. Zallat, "Mueller polarimetric imaging of biological tissues: classification in a decision-theoretic framework," J. Opt. Soc. Am. A **35**(12), 2046–2057 (2018).
29. X. Li, R. Liao, J. Zhou, P. T. Y. Leung, M. Yan, and H. Ma, "Classification of morphologically similar algae and cyanobacteria using Mueller matrix imaging and convolutional neural networks," Appl. Opt. **56**(23), 6520–6530 (2017).
30. M. Ma, Y. Zou, and Z. Huang, "Deep learning-based automated morphology classification of electrospun ultrafine fibers from M44 element image of Muller matrix," Optik **206**, 164261 (2020).
31. N. Rajini and B. R, "Classification of MRI brain images using K-nearest neighbor and artificial neural network," *International Conference on Recent Trends in Information Technology, ICRTIT 2011* (2011).
32. R. Ramteke and M. Khachane, "Automatic medical image classification and abnormality detection Using K- nearest neighbour," International Journal of Advanced Computer Research **2**(4), 190–196 (2012).

Biomedical Optics EXPRESS

33. S. Wagle, J. A. Mangai, and V. S. Kumar, "An improved medical image classification model using data mining techniques," in *2013 7th IEEE GCC Conference and Exhibition (GCC)* (2013), pp. 114–118.
34. F. Yang, M. Hamit, C. B. Yan, J. Yao, A. Kutluk, X. M. Kong, and S. X. Zhang, "Feature extraction and classification on esophageal X-ray images of Xinjiang Kazak nationality," J. Healthc. Eng. **2017**, 1–11 (2017).
35. C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," IEEE Trans. Med. Imaging **38**(1), 280–290 (2019).
36. K. Pathak, M. Pavthawala, N. Patel, D. Malek, V. Shah, and B. Vaidya, "Classification of brain tumor using convolutional neural network," in *2019 3rd International conference on Electronics, Communication and Aerospace Technology* (ICECA) (2019), pp. 128–132.
37. J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," IEEE Trans. Med. Imaging **37**(2), 491–503 (2018).
38. F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 Challenge," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Springer, 2018).
39. Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," ArXiv abs/1606.06650 (2016).
40. A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. V. Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, P. Bilic, P. F. Christ, R. K. G. Do, M. Gollub, J. Golia-Pernicka, S. H. Heckers, W. R. Jarnagin, M. K. McHugo, S. Napel, E. Vorontsov, L. Maier-Hein, and M. J. Cardoso, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," arXiv:1902.09063 (2019).
41. V. N. Du Le, I. Saytashev, S. Saha, P. F. Lopez, M. Laughrey, and J. C. Ramella-Roman, "Depth-resolved Mueller matrix polarimetry microscopy of the rat cornea," Biomed. Opt. Express **11**(10), 5982–5994 (2020).
42. I. Saytashev, S. Saha, J. Chue-Sang, P. Lopez, M. Laughrey, and J. C. Ramella-Roman, "Self validating Mueller matrix micro-mesoscope (SAMMM) for the characterization of biological media," Opt. Lett. **45**(8), 2168–2171 (2020).
43. S. Y. Lu and R. A. Chipman, "Interpretation of Mueller matrices based on polar decomposition," J. Opt. Soc. Am. A **13**(5), 1106–1113 (1996).
44. J. Qi and D. S. Elson, "Mueller polarimetric imaging for surgical and diagnostic applications: a review," J. Biophotonics **10**(8), 950–982 (2017).
45. R. M. A. Azzam, "Propagation of partially polarized-light through anisotropic media with or without depolarization - differential 4×4 matrix calculus," J. Opt. Soc. Am. **68**(12), 1756–1767 (1978).
46. N. Ortega-Quijano and J. L. Arce-Diego, "Mueller matrix differential decomposition," Opt. Lett. **36**(10), 1942–1944 (2011).
47. M. L. Akins, K. Luby-Phelps, and M. Mahendroo, "Second harmonic generation imaging as a potential tool for staging pregnancy and predicting preterm birth," J. Biomed. Opt. **15**(2), 026020 (2010).
48. Y. Zhang, M. L. Akins, K. Murari, J. Xi, M. J. Li, K. Luby-Phelps, M. Mahendroo, and X. Li, "A compact fiber-optic SHG scanning endomicroscope and its application to visualize cervical remodeling during pregnancy," Proc. Natl. Acad. Sci. U. S. A. **109**(32), 12878–12883 (2012).
49. A. Bhandare, M. V. Bhide, P. Gokhale, and R. Chandavarkar, "Applications of convolutional neural networks," Int. J. Comp. Sci. Info. Tech. **7**, 2206–2215 (2016).
50. W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Process. **13**(4), 600–612 (2004).
51. L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Proceedings of the European Conference on Computer Vision (ECCV), abs/1802.02611* (2018).
52. M. M. Badža and M. Č. Barjaktarović, "Classification of brain tumors from MRI images using a convolutional neural network," Appl. Sci. **10**(6), 1999 (2020).
53. D. Anyfantis, M. Karagiannopoulos, S. Kotsiantis, and P. Pintelas, "Robustness of learning techniques in handling class noise in imbalanced datasets," in *Artificial Intelligence and Innovations 2007: from Theory to Applications*, C. Boukis, A. Pnevmatikakis, and L. Polymenakos, eds., (Springer US, 2007), pp. 21–28.
54. X. Qin, Y. Zhou, Z. He, Y. Wang, and Z. Tang, "A faster R-CNN based method for comic characters face detection," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (IEEE), pp. 1074–1080.
55. J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for MR image reconstruction," in M. M. Niethammer, *et al.*, eds., *Information Processing in Medical Imaging*, vol 10265 (Springer, 2017), pp. 647–658.