Received: 31 May 2021 | Revised: 25 November 2021 | Accepted: 25 November 2021

DOI: 10.1002/jbio.202100167

RESEARCH ARTICLE

JOURNAL OF BIOPHOTONICS

In vivo detection of head and neck tumors by hyperspectral imaging combined with deep learning methods

Dennis Eggert[1*] | **Marcel Bengs**[2] | **Stephan Westermann**[3] | **Nils Gessert**[2] | **Andreas O. H. Gerstner**[4] | **Nina A. Mueller**[3] | **Julian Bewarder**[1] | **Alexander Schlaefer**[2] | **Christian Betz**[1] | **Wiebke Laffers**[1,3,5]

[1]Clinic and Polyclinic for Otolaryngology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Institute of Medical Technology, Hamburg University of Technology, Hamburg, Germany

[3]Department of Otorhinolaryngology/ Head and Neck Surgery, University of Bonn, Bonn, Germany

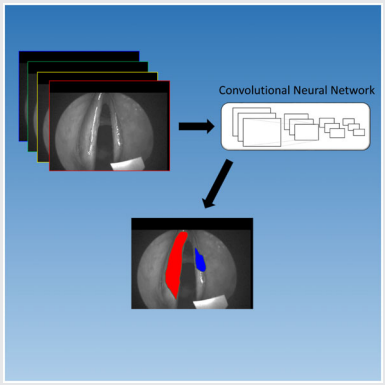
[4]Department of Otorhinolaryngology, Klinikum Braunschweig, Braunschweig, Germany

[5]Department of Otorhinolaryngology, Head and Neck Surgery, Evangelisches Krankenhaus, Carl von Ossietzky-University, Oldenburg, Germany

***Correspondence**
Dennis Eggert, Clinic and Polyclinic for Otolaryngology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany.
Email: d.eggert@uke.de

Funding information
Deutsche Krebshilfe

Abstract



Currently, there are no fast and accurate screening methods available for head and neck cancer, the eighth most common tumor entity. For this study, we used hyperspectral imaging, an imaging technique for quantitative and objective surface analysis, combined with deep learning methods for automated tissue classification. As part of a prospective clinical observational study, hyperspectral datasets of laryngeal, hypopharyngeal and oropharyngeal mucosa were recorded in 98 patients before surgery in vivo. We established an automated data interpretation pathway that can classify the tissue into healthy and tumorous using convolutional neural networks with 2D spatial or 3D spatio-spectral convolutions combined with a state-of-the-art Densenet architecture. Using 24 patients for testing, our 3D spatio-spectral Densenet classification method achieves an average accuracy of 81%, a sensitivity of 83% and a specificity of 79%.

KEYWORDS

convolutional neural network, head and neck cancer, hyperspectral imaging, intraoperative imaging, optical biopsy

1 | INTRODUCTION

According to the global cancer statistics 2018, head and neck cancer is estimated the eighth most common tumor entity worldwide, thereof yet 177 422 new cases of carcinoma of the larynx [1]. Histologically, about 90% of these tumors are squamous cell carcinomas originating from epithelial cells [2]. The overall survival for head and neck cancer is still poor; the 5-year survival rate for carcinomas of the larynx accounting for only approximately 50%, regardless of the improvement in diagnostic modalities and treatment outcome [3]. Prognosis declines with extent of the tumor; therefore, early detection of malignant or even better of premalignant lesions is the best

Dennis Eggert and Marcel Bengs contributed equally to this work.

way to enhance effectiveness in disease management, treatment outcome and quality of life of patients [4].

State-of-the-art examination for the primary onsite diagnosis is performed by visual inspection often in combination with handheld flexible or rigid (video-) laryngoscopes. The introduction of endoscopic tools in the daily practice has improved the detection rate of premalignant lesions. However, the detection of suspect mucosal lesions is limited to the examiner's experience and skills and the description of morphological aspects of the tissue's surface. But even experienced physicians cannot sufficiently differentiate benign hyperplastic or premalignant dysplastic epithelial lesions in the larynx only by their outer appearance [5]. Additionally, the risk for malignant transformation [6] or early invasive growth cannot be determined with certainty.

Therefore, the gold standard to further classify endoscopically detected lesions is invasive biopsies and a histopathological assessment requiring general anesthesia [7, 8]. Besides the common risks of general anesthesia, every biopsy may lead to tissue damage and scar formation. Especially, the function of the larynx is extremely sensitive to epithelial changes and scarring, so that even small tissue biopsies can cause long-term adverse effects such as dysphonia [9], resulting in a poor quality of life. But even if there are studies where biopsies taken from leucoplakia of the larynx showed no dysplasia in about 50% of the cases [6], one has to take into consideration the other half of the patients. Biopsies are only taken from a small part of a suspect area and could still miss the transformed cell area, so mucosal lesions with no dysplasia in the initial biopsy still have an increased risk for transforming into an invasive squamous cell carcinoma in the future.

These facts underline the importance for a better diagnostic risk stratification of mucosal lesions of the head and neck region without invasive biopsies under general anesthesia. In this regard, a tool would be desirable that is on the one hand capable of detecting early invasive growth or high-grade dysplasia in need of surgical intervention, but that could—on the other hand—differentiate lesions with no or low-grade dysplasia thus preventing unnecessary surgeries. At the same time, it should be a reliable tool for follow-up examinations of these latter lesions without performing recurrent invasive biopsies.

Hyperspectral imaging (HSI) is a noninvasive imaging technique for quantitative and objective surface analysis detecting reflecting light [10, 11]. HSI has been developed for satellites and geographical measurements and has been applied in medical research in recent years. In contrast to conventional RGB (red, green, blue) images consisting only of three image bands on red, green, and blue wavelengths, respectively, HSI uses more spectral channels (up to a few hundred) with each channel only having a small bandwidth. HSI generates a three-dimensional dataset of a two-dimensional image on each wavelength called hyperspectral image cube [12]. Acquiring a HSI image cube can be done either by using a special illumination light source that uses a monochromator to illuminate the sample with light of different wavelengths coupled to a standard monochromatic camera or by using a special hyperspectral camera that brakes up the reflected light into different spectral channels in combination with a standard white light source (eg, a halogen light bulb).

HSI has been applied to different medical purposes like enhancement of blood vessel visualization [13], identifying intestinal ischemia [14] or oximetry of the retina [15, 16]. Promising results are achieved in detecting cancer in vivo and ex vivo such as gastric cancer, colon cancer, breast cancer, brain cancer, and head neck cancer, respectively [17]. It has also been used to classify pigmented skin lesions, and to identify malignant melanoma in atypical naevi of the skin [18].

Due to the complexity of HSI datasets, interpretation is not trivial, and usually cannot be done by the examining physician. It has been shown that computer algorithms that use deep learning methods (especially convolutional neural networks [CNNs]) enable an objective analysis of HSI datasets [19–22]. Recently, deep learning methods have been applied in a couple of clinical studies [20, 21, 23–30]. Most of these studies performed the image acquisition ex vivo [21, 24–28]. Only the studies by Fabelo et al [20], Lin et al [29], Tsai et al [23] and one study from the authors of this paper [30] were conducted in vivo but they had different aims and approaches than the current study.

In our current study, we used state-of-the-art deep learning techniques for the classification of HSI images acquired in vivo from laryngeal, hypopharyngeal and oropharyngeal mucosa.

2 | EXPERIMENTAL SECTION

2.1 | Acquisition of HSI datasets

HSI datasets of laryngeal, hypopharyngeal and oropharyngeal mucosa were taken in vivo from 98 patients who were examined due to suspicious lesions of the mucosal membrane. This study was initially performed as a prospective clinical observational study at the Department of Otorhinolaryngology/Head and Neck Surgery, University of Bonn, supported by the Deutsche Krebshilfe as part of the project "Early Detection of Laryngeal Cancer by Hyperspectral Imaging" (project numbers 109825 and 110275). The study was approved by the local ethics committee (176/10061/13, University of Bonn). Written informed consent was gained from all patients undergoing this study. The acquisition procedure is described in detail in Gerstner et al [11]. In brief, data acquisition was

performed simultaneously to tissue collection for diagnostic evaluation under general anesthesia. A monochromator (Polychrome V, TillPhotonics, Graefelfing, Germany) coupled to a rigid 0°-endoscope was used for illumination. The endoscope was fixed in an endoscope holding system (all endoscopic devices: Karl Storz GmbH CoKG, Tuttlingen, Germany) to minimize motion artifacts. HSI datasets were acquired with a spectral range from 390 to 680 nm in 30 steps of 10 nm bandwidth. The HSI datasets were acquired using a dedicated computer as an integrated imaging platform (Xeon 6 core, Zeiss MicroImaging, Göttingen) with an imaging software (AxioVision 4.8.2, Zeiss MicroImaging, Göttingen) and a monochromatic high-resolution monochromatic CCD-camera (AxioCam MRm, Zeiss MicroImaging, Göttingen). The monochromator was synchronously triggered with the camera in order to generate monochromatic images at different wavelengths at 10 nm steps (390, 400, 410, ..., 680 nm; each band with 15 nm FWHM). The acquisition time for one HSI cube was 25 seconds. Only the last 22 channels (from 470 to 680 nm) were used for image processing as the first eight channels were too dark lacking usable information. In the first eight channels, the pixels in the lesions showed too little deviation to pixels outside of the lesion. This was probably based on the combination of the monochromator which does not have the same brightness for all spectral bands as well as the camera which also does not have the same sensitivity for all spectral bands. RGB images were taken by a conventional camera right after performing the HSI. As white balancing was not integrated in the software, we used a white balancing sheet (Zenith Polymer® Folie, Sphereoptics GmbH, Uhldingen, Germany) also which was cut to an appropriate size, sheathed and directly placed in the examined area without concealing the suspicious lesion. This sheet was sterilized prior to use according to medical hygiene standards. Clinically, healthy areas and clinically suspect areas were digitally marked within the recorded RGB-images by medical experts (head neck surgeons) and confirmed by histopathological examinations. For the healthy areas, no tissue extraction was performed, due to ethical reasons. To obtain a well-defined annotation, we focus on classifying the marked areas. While all images contain a marked suspect area, healthy areas could not be marked in 29 cases due to the extent of the tumor.

2.2 | Deep learning methods

Before applying our deep learning methods, we follow the preprocessing pipeline of a previous study [31], which includes aligning and filtering of the hyperspectral image stacks. Due to the breathing movements of the patient during the acquisition (25 seconds for one HSI cube), the different spectral images are laterally not aligned. Hence, we performed an image registration using the ImageJ-implementation of the SIFT-algorithm "Linear Stack Alignment with SIFT" [32].

We frame our classification task, as a two-class supervised learning problem and train our deep learning methods to classify tumor or healthy tissue based on HSI data, while both benign and malignant lesions belong to the class tumor. For this purpose, we consider image patches with a size of $32 \times 32 \times 22$ (high-width-spectral dimension) cropped from the marked areas of the images. The patch size was chosen experimentally based on inspection of the images in our data set. To be able to consider smaller marked areas, we choose a patch size of 32×32 pixels. In addition, the smaller patch size allows for an increased number of patches that can be used for training.

Notably, some areas of the HSI contain specular reflections due to wet tissue surfaces. Therefore, we filter the patches with an intensity threshold and exclude patches that contain primarily extreme specular reflections. The value of the intensity threshold was chosen manually based on inspection of the extracted patches and the entire images. We set the threshold of the pixel intensities to 245 (for 8-bit images the pixel intensity is between 0 and 255).

To learn from the image patches, we employ a state-of-the-art CNN and we follow the idea of densely connected neural networks [33] that enable strong feature propagation and reduce the number of parameters substantially. For our CNN, we use one initial convolutional layer, followed by 3 Densenet blocks which are connected with average pooling layers. In each average pooling layer, we use a stride of two and hence downsample our input dimensions. The Densenet blocks consist of three convolutional layers each, while each layer receives the output of all its preceding layers. We fed the output of the last Densenet block into a global average pooling layer, followed by our final classification layer with two outputs.

Using this baseline CNN, we evaluate three different methods to learn from the HSI data.

First, we evaluate learning with the RGB representations derived from the HSI stacks and employ our baseline CNN with 2D spatial convolutions, which is the standard approach for natural color images (RGB 2D Densenet).

Second, we use our baseline with 2D convolutions and consider the different spectral dimensions as additional color information and stack all the spectral dimensions into the input's channel dimension. Note, this approach has been proposed for ex-vivo head and neck cancer classification using HSI data [34] (HSI 2D Densenet).

Third, we extend our architecture to learn combined features from the spatial and spectral dimensions throughout the network. For this purpose, we consider the different spectral channels as an additional dimension and employ 3D convolutions [22] (HSI 3D Densenet).

To have a balanced test and validation set in terms of classes, we only consider those images that have both areas marked (tumor, healthy tissue). Note that 69 images out of the 98 have both areas marked.

While our data come from a challenging in vivo environment, the quality of the data varies quite significantly (strong reflections and low image quality). Note that in the case of applying our method in vivo for classification, one would take a new image in case of obvious bad quality. Hence, we carefully inspected the images of the 69 patients that have both areas marked and considered images with low quality only for training. We also evaluated to remove the images with bad quality in general. However, this reduces the available training data quite notably and we observed lower performance in this case. Hence, we left images with low quality during training but excluded them from testing.

We split our data into two subsets and apply a cross-fold scheme. Following above procedure, we ended up with images of n = 40 patients that we considered for validation and testing ($k = 2$, testing (n) = 12, validation (n) = 8). Note, our test and validation splits are balanced in terms of classes. Using all patches of a marked area leads to redundant information, that is, strong overlap between subsequent patches of an area. Hence, for training we choose a distance of 10 pixels between subsequent patches of a marked area and for evaluation we consider a distance of 5 pixels between subsequent patches of a marked area. For the first fold this leads to 100 265, 16 422 and 23 824 patches for training, validation and testing, respectively. For the second fold, this leads to 83 595, 55 367 and 51 001 patches for training, validation and testing, respectively. For evaluation, we average all patch predictions for a marked area to obtain one final prediction for the corresponding area of a patient.

Using Adam [35] for optimization, we train all our networks with a cross-entropy loss for 700 epochs with a batch size of 1170. To counter our class imbalance, we weight the loss of the individual classes inversely proportional to the class frequency. We use a learning rate of lr = 0.0001 and every 100 epochs we reduce the learning rate by a factor of two. For data augmentation, we use random horizontal and vertical flipping of the patches during training. All our methods are implemented in PyTorch (v1.7.0) and trained on a NVIDIA TITAN RTX.

Training takes around 30 minutes for the 3D CNN and around 20 minutes for the 2D CNNs, using a NVIDIA TITAN RTX graphic card.

Using the same graphic card and performing 100 repetitions, the inference times of a single patch for the different methods are 2.49 ± 0.021 milliseconds (RGB 2D CNN), 2.50 ± 0.027 milliseconds (HSI 2D CNN) and 3.0117 ± 0.157 milliseconds (HSI 3D CNN). Using a batch size of 1000, the throughput, that is, the number of patches that can be estimated within 1 second are 190.187 (RGB 2D CNN), 171.526 (HSI 2D CNN) and 10.595 (HSI 3D CNN).

Hence, for all methods the classification during testing can be performed within seconds.

3 | RESULTS

Figure 1 shows the averaged accuracy, sensitivity, specificity, the F1-Score and the area under the curve (AUC) for each of our tissue classification methods across our two folds with SD. Note, for some metrics we obtain the

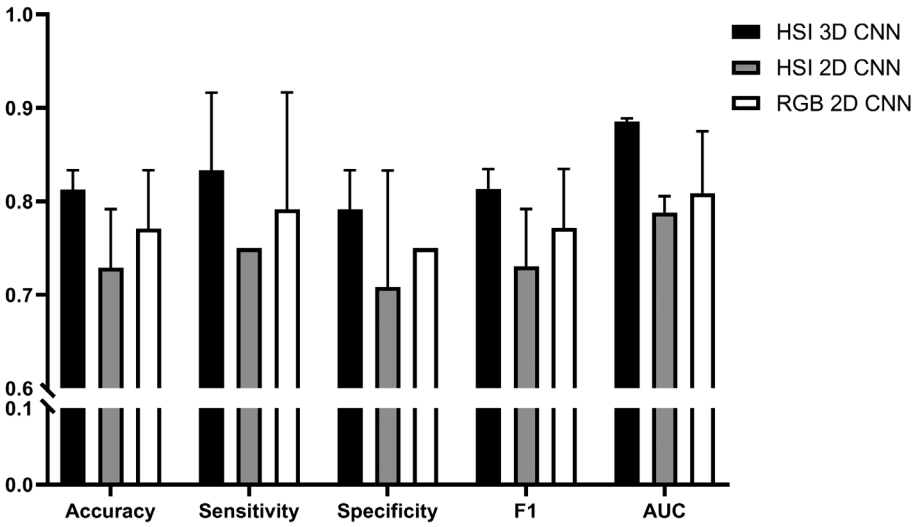


FIGURE 1 Averaged accuracy, sensitivity, specificity, AUC and F1-Score for the tissue classification of the HSI and RGB datasets. HSI datasets were classified using either a 3D Densenet algorithm or a 2D Densenet algorithm, RGB images were analyzed by a 2D Densenet algorithm

exact same performance for each fold, hence the SD is zero. Figure 2 shows the receiver operating characteristic (ROC) curve.

HSI datasets analyzed by our 3D Densenet algorithm showed the best classification results. Accuracy for the tissue classification was 0.813 ± 0.021 for HSI 3D Densenet compared to 0.729 ± 0.063 for HSI 2D Densenet and 0.771 ± 0.063 for RGB 2D Densenet. Also the sensitivity (0.833 ± 0.083), the specificity (0.792 ± 0.042), the F1 score (0.812 ± 0.021) and the AUC (0.872 ± 0.004) were best for the HSI 3D Densenet approach. Also the ROC curve (Figure 2) shows that the HSI 3D Densenet approach performs better for the tissue classification than the HSI 2D Densenet and the RGB 2D Densenet approach.

Figures 1 and 2 show that tissue classification based on HSI image cubes was superior to RGB images, when the full number of 22 spectral channels was used for the 3D Densenet approach. We then tested the dependency of the classification accuracy on the number of spectral channels of the HSI image cubes. For this experiment, we gradually exclude the margins of our spectral range, for example, for 18 wavelengths we consider the range of 490 to 660 nm, for 14 wavelengths we consider the range of 510 up to 640 nm and for 10 wavelengths we consider the range of 530 up to 620 nm. Figure 3 shows the accuracy of the tissue classification with an increasing number of spectral channels.

The accuracy of the tissue classification improves with an increasing number of spectral channels. For 10 spectral channels, the accuracy was 0.708 ± 0.000, for

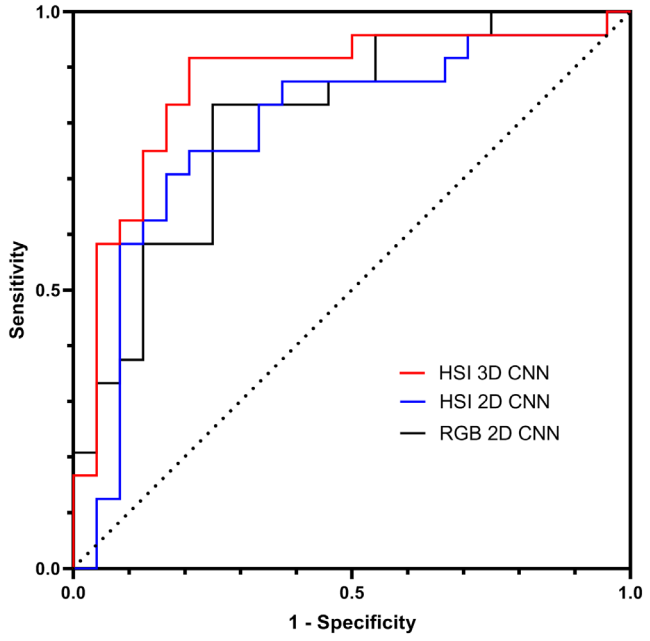


FIGURE 2 ROC curve for the classification of the HSI datasets by a 3D Densenet algorithm, a 2D Densenet algorithm, and the classification of the RGB images by a 2D Densenet algorithm

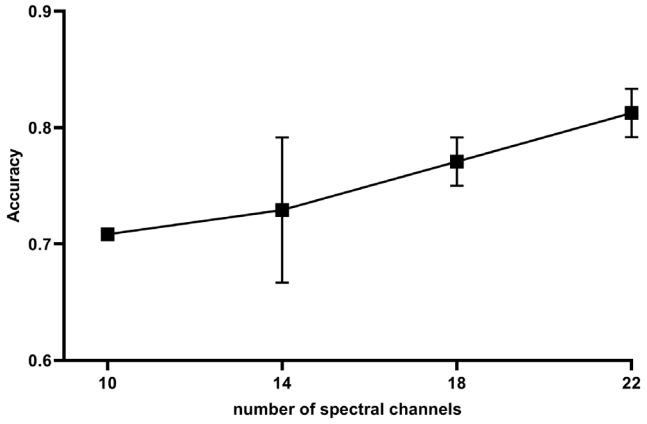


FIGURE 3 Dependency of the classification accuracy on the number of spectral channels of the HSI image cubes

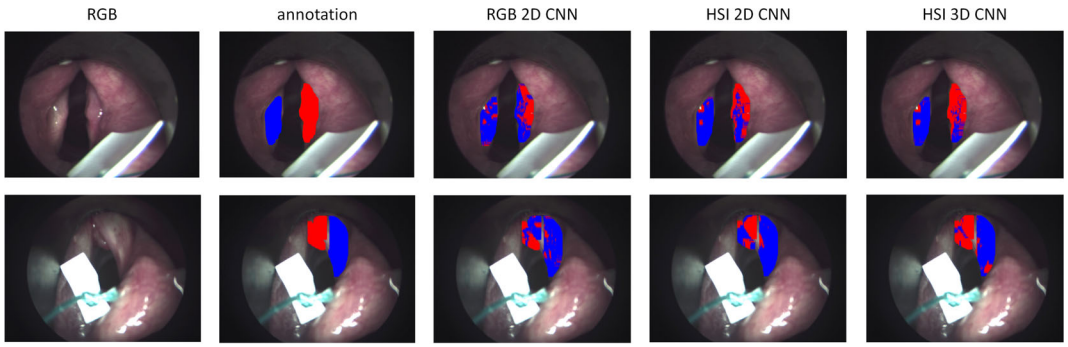


FIGURE 4 Visual prediction by our automated tissue classification pipeline if a given tissue area was healthy or tumorous (blue = healthy, red = tumorous). Shown is the endoscopic RGB image, the annotation that was based on the histopathological examinations, the prediction based on the RGB image and the HSI image cube analyzed by our 2D Densenet algorithm and the prediction based on the HSI image cube analyzed by our 3D Densenet algorithm (using all 22 spectral channels). The white sheet in the lower part of the images was used for white balance

14 spectral channels 0.729 ± 0.063, for 18 spectral channels 0.771 ± 0.021, and for 22 spectral channels 0.813 ± 0.021.

In the final step, we wanted our automated tissue classification pipeline to visually predict if a given area was healthy or tumorous (Figure 4).

Figure 4 shows that the visual prediction based on the HSI image cube analyzed by our 3D Densenet algorithm comes closest to the annotation with two small false positive regions at the left vocal fold and a false negative region at the lower part of the right vocal fold. There are more false positive and false negative regions present in the RGB image and the HSI image cube by our 2D Densenet algorithm.

4 | DISCUSSION AND OUTLOOK

In our study, we demonstrated automated tissue classification in the head and neck region with special regards to the larynx in vivo by combining HSI with state-of-the-art deep learning techniques.Most studies combining HSI and deep learning that have recently been published were performed on ex vivo tissue [21, 24–28], which makes a huge difference to in vivo studies. Important parameters like blood flow and oxygenation as well as limiting factors like movements of the patients and reflections mainly from saliva are not present in an ex vivo setting. Therefore, the results from ex vivo studies cannot directly be applied as an onsite screening method whereas in vivo studies come close to a setting that could be used in medical screening routine.

The studies performed in vivo had different approaches than our study. In the study by Fabelo et al [20], the aim was glioblastoma detection. In the study by Lin et al [29], the aim was the development of an instrument and a software pipeline for 3D reconstruction of the tissue surface and intraoperative HSI, but no cancer detection was performed. The study by Tsai et al [23] uses a system to convert white light endoscopic images and narrow band images to hyperspectral images and compares the classification results. So the hyperspectral images were generated virtually and not directly by the light source.

In our study, we used a monochromator as the light source to acquire HSI datasets in vivo. Using a deep learning pipeline developed and optimized for this purpose, we were able to automatically distinguish between healthy and tumorous tissue with an average accuracy of 0.813, a sensitivity of 0.833 and a specificity of 0.792. This is similar to the results of Halicek et al [21, 34] who developed a method using CNNs and HSI images acquired by a CRI Maestro imaging system. The CNN could distinguish squamous cell carcinoma from normal oral tissues with an accuracy of 0.81. Other authors applied HSI in vivo for the analysis of tongue tumors. They used an acousto-optic tunable filter and a classifier based on the sparse representation method comparing it to support vector machine and relevance vector machine gaining an accuracy of 0.965 with their method [36]. In former studies, part of our group analyzed a small set of identical HSI datacubes by classification using random forest algorithm to distinguish between healthy and cancerous tissues with an accuracy of 0.88 using five patients for testing [31]. Altogether, the method used in our study currently performs at the lower edge of the range reported in comparable trials which focused on automated HSI-based tissue classification in the upper aerodigestive tract. Possible reasons for this finding are presented in the following.

Interference factors and limitations of the present study were motion artifacts, image noise and surface reflection which were also observed in other studies [21, 31]. Some artifacts like the patients' pulse cannot be eliminated when performing in vivo studies. But as shown by Regeling et al [31], this impediment can be overcome by the calculation of corrections which can also be used for the correction of image noise. The so-called preprocessing of HSI images is performed in many studies and also showed good results in removing specular glare in the work of Halicek et al [21].

Another factor that could have compromised the performance of the tissue classification in our study was the quality of the labeling of the RGB images that were taken as annotation. Biopsies were only taken from certain parts of the suspect area, but if the histopathological examinations confirmed a tumor the whole suspect area was marked "tumorous." So the size of the tumor area might have been overestimated. In Figure 4, the lower part of the right vocal fold was recognized as "healthy" by all three analysis methods, but the annotation classifies it as "tumor." As no biopsy was taken from this area, it remains unclear if the tissue classification by our deep learning approach was wrong or if the labeling of the annotation was wrong for that part of the tumor. For future studies, more care should be taken of the precise labeling of the RGB images.

The imbalanced values obtained for the sensitivity and specificity are mainly based on the fact that we have an imbalanced training data set, where we have more samples of the class tumor. Training with this data set leads to a bias toward predicting the tumor class, which in turn leads to the higher sensitivity (tumor recognition rate). Note that during training, we weight the loss of the individual classes inversely proportional to the class frequency to counter this effect. However, this does not

completely remove the classification bias. Moreover, the trade-off between sensitivity and specificity can be adjusted during testing by using different decision threshold, this is visualized in the ROC curve, see Figure 2. For a patient, it is most important that every tumor is recognized, so for the medical practice a higher sensitivity is more important than a higher specificity. That is why we have chosen the sensitivity to be higher than the specificity.As described in the review by Lu and Fei [12], HSI generally covers a contiguous portion of the light spectrum with more spectral bands (up to a few hundred) and higher spectral resolution than multispectral imaging (such as RGB color cameras). We compared automated tissue classification results based on the HSI image cubes with RGB images taken with the same optical setup. Our results demonstrate that high performance can be achieved by only using RGB information. This highlights that spatial information with few bands already contains valuable information for the classification task. Also the annotation was performed on the RGB images. Comparing RGB 2D CNN and HSI 2D CNN, the only difference between those two networks is that for HSI 2D CNN 22 input channels are given instead of three. The reduced performance of HSI 2D CNN suggests that the additional HS information cannot effectively be learned by the network, reducing the performance compared to the compact RGB representation. This indicates that more complex processing of the spectral information is required, which is confirmed by our results for our 3D approach, which learns features jointly from the spectral and spatial dimensions. Tissue classification was notably better for HSI image cubes using the 3D Densenet approach. According to our data, the classification accuracy correlates with the number of spectral channels of the HSI data cubes used. More spectral channels resulted in improved tissue classification accuracy and eventually outperformed the three channels of RGB images. HSI sensors are capable of capturing a very large number of contiguous spectral bands across the electromagnetic spectrum. Each pixel of an HSI image has associated a specific spectral signature combined in the data cube [17]. The larger amount of data and the spectral signatures might be beneficial for the interpretation by deep learning techniques. Also specific physiological properties that are associated with tumor growth like lower oxygenation and a lower concentration of hemoglobin can be determined by HSI [37].

In future studies, we want to use a HSI system with 100 spectral channels in 10 nm increments which might lead to even better automated tissue classification results. However, in this respect the study of Halicek et al [22] has to be taken in account, showing only an accuracy of less than 0.80 using a spectral bandwidth from 450 to 900 nm in 5-nm increments producing a hypercube with 91 spectral bands, in an ex-vivo examination of excised tissue samples of patients with squamous cell carcinomas. This may be due to the fact that only 29 patients were included and that spectral redundancy may occur when very small spectral increments like 5 nm are used.

Further challenges in HSI involve the fast processing of the vast amount of data. In the future, fast processing could be performed on modern GPUs with pretrained deep learning networks. To achieve a reliable pretrained deep learning network, large amounts of training data (HSI images + annotation) have to be acquired. Up to date, there is no library of HSI data cubes available that can be used for training. So one of the major challenges for the future is the creation of such a library. Furthermore, the computer programs for tissue analysis have to be designed "user-friendly" so that they can be operated by physicians instead of computer experts.

In this paper, the authors describe a tool that discriminates tumors from healthy tissue. In a different study, the authors also developed a tool for the discrimination of benign and malignant lesions using deep learning techniques and HSI [30]. However, the performance in discriminating between benign and malignant lesions was lower (maximum sensitivity = 0.745, maximum specificity = 0.689) than the performance that was achieved in discriminating tumors from healthy tissue described in the current study.

The usage of HSI shows several advantages, being noninvasive and without the need of applying special markers (like fluorescent or radioactive markers). Furthermore, image acquisition only takes a few seconds. Therefore, it could be applied in an outpatient setting. With future developments like the integration of an HSI into flexible laryngoscopes [38] and algorithm development that enables real-time tissue classification, HSI might have the potential to revolutionize head and neck tumor detection, especially in the laryngeal region, and become one of the standard methods in this field.

5 | CONCLUSION

The current study demonstrates the potential of HSI combined with deep learning methods as tool for the classification of head and neck tumors. Using our automated tissue classification pipeline, we were able to classify the tissue into healthy and tumorous with an average accuracy of over 80%.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Krebshilfe as part of the project "Early Detection of Laryngeal Cancer

by Hyperspectral Imaging" (project numbers 109825 and 110275). This work was also supported by the Forschungszentrum Medizintechnik Hamburg (fmthh) Förderlinie 2019 (project: Verbesserte Diagnostik von Tumoren des oberen Luft-Speisewegs durch Kombination von hyperspektraler Bildgebung mit Methoden der KI). Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no financial or commercial conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Dennis Eggert https://orcid.org/0000-0002-6464-9373
Christian Betz https://orcid.org/0000-0002-7717-7387

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, *CA Cancer J. Clin.* **2018**, *68*, 394.
[2] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, M. J. Thun, *CA Cancer J. Clin.* **2007**, *57*, 43.
[3] S. B. Thavarool, G. Muttath, S. Nayanar, K. Duraisamy, P. Bhat, K. Shringarpure, P. Nayak, J. P. Tripathy, A. Thaddeus, S. Philip, B. Satheesan, *World J. Surg. Oncol.* **2019**, *17*, 15.
[4] S. S. Napier, P. M. Speight, *J. Oral Pathol. Med.* **2008**, *37*, 1.
[5] N. Gale, D. R. Gnepp, M. Poljak, P. Strojan, A. Cardesa, T. Helliwell, R. Sifrer, M. Volavsek, A. Sandison, N. Zidar, *Adv. Anat. Pathol.* **2016**, *23*, 84.
[6] J. S. Isenberg, D. L. Crozier, S. H. Dailey, *Ann. Otol. Rhinol. Laryngol.* **2008**, *117*, 74.
[7] N. Gale, R. Blagus, S. K. El-Mofty, T. Helliwell, M. L. Prasad, A. Sandison, M. Volavsek, B. M. Wenig, N. Zidar, A. Cardesa, *Histopathology* **2014**, *65*, 456.
[8] N. Gale, L. Michaels, B. Luzar, M. Poljak, N. Zidar, J. Fischinger, A. Cardesa, *Histopathology* **2009**, *54*, 639.
[9] M. D. Weller, P. C. Nankivell, C. McConkey, V. Paleri, H. M. Mehanna, *Clin. Otolaryngol.* **2010**, *35*, 364.
[10] A. O. Gerstner, W. Laffers, G. Schade, F. Göke, R. Martin, B. Thies, *HNO* **2012**, *60*, 1047.
[11] A. O. Gerstner, W. Laffers, F. Bootz, D. L. Farkas, R. Martin, J. Bendix, B. Thies, *J. Biophotonics* **2012**, *5*, 255.
[12] G. Lu, B. Fei, *J. Biomed. Opt.* **2014**, *19*, 10901.
[13] H. Akbari, Y. Kosugi, K. Kojima, N. Tanaka, *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2009**, 1461.
[14] H. Akbari, Y. Kosugi, K. Kojima, N. Tanaka, *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2008**, 1238.
[15] J. Kaluzny, H. Li, W. Liu, P. Nesper, J. Park, H. F. Zhang, A. A. Fawzi, *Curr. Eye Res.* **2017**, *42*, 629.
[16] L. Gao, R. T. Smith, T. S. Tkaczyk, *Biomed. Opt. Express* **2012**, *3*, 48.
[17] M. Halicek, H. Fabelo, S. Ortega, G. M. Callico, B. Fei, *Cancer* **2019**, *11*, 756.
[18] D. L. Farkas, D. Becker, *Pigment Cell Res.* **2001**, *14*, 2.
[19] P. R. Jeyaraj, E. R. S. Nadar, *J. Cancer Res. Clin. Oncol.* **2019**, *145*, 829.
[20] H. Fabelo, M. Halicek, S. Ortega, M. Shahedi, A. Szolna, J. F. Piñeiro, C. Sosa, A. J. O'Shanahan, S. Bisshopp, C. Espino, M. Márquez, M. Hernández, D. Carrera, J. Morera, G. M. Callico, R. Sarmiento, B. Fei, *Sensors* **2019**, *19*, 920.
[21] M. Halicek, J. V. Little, X. Wang, A. Y. Chen, B. Fei, *J. Biomed. Opt.* **2019**, *24*, 1.
[22] M. Halicek, G. Lu, J. V. Little, X. Wang, M. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, B. Fei, *J. Biomed. Opt.* **2017**, *22*, 60503.
[23] C. L. Tsai, A. Mukundan, C. S. Chung, Y. H. Chen, Y. K. Wang, T. H. Chen, Y. S. Tseng, C. W. Huang, I. C. Wu, H. C. Wang, *Cancer* **2021**, *13*, 4593.
[24] S. Ortega, M. Halicek, H. Fabelo, R. Camacho, M. L. Plaza, F. Godtliebsen, G. M. Callicó, B. Fei, *Sensors* **1911**, *2020*, 20.
[25] S. Ortega, M. Halicek, H. Fabelo, R. Guerra, C. Lopez, M. Lejaune, F. Godtliebsen, G. M. Callico, B. Fei, *Proc. SPIE Int. Soc. Opt Eng.* **2020**, *11320*, 113200V.
[26] L. Ma, M. Halicek, X. Zhou, J. Dormer, B. Fei, *Proc. SPIE Int. Soc. Opt. Eng.* **2020**, *11320*, 113200W.
[27] M. Halicek, H. Fabelo, S. Ortega, J. V. Little, X. Wang, A. Y. Chen, G. M. Callico, L. Myers, B. D. Sumer, B. Fei B. *J. Med. Imag.* **2019**, *6*, 035004.
[28] M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, B. Fei, *Cancers* **2019**, *11*, 1367.
[29] J. Lin, N. T. Clancy, J. Qi, Y. Hu, T. Tatla, D. Stoyanov, L. Maier-Hein, D. S. Elson, *Med. Image Anal.* **2018**, *48*, 162.
[30] M. Bengs, N. Gessert, W. Laffers, D. Eggert, S. Westermann, N. A. Mueller, A. O. H. Gerstner, C. Betz, A. Schlaefer Medical, *Image Computing Comput. Assist. Interven. - MICCAI* **2020**, MICCAI 2020. Lecture Notes in Computer Science, vol 12263. Springer, Cham 690.
[31] B. Regeling, W. Laffers, A. O. Gerstner, S. Westermann, N. A. Müller, K. Schmidt, J. Bendix, B. Thies, *J. Biophotonics* **2016**, *9*, 235.
[32] D. G. Lowe, *Int. J. Comput. Vis.* **2004**, *60*, 91.
[33] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, *arXiv: 1608.06993* **2016**.
[34] M. Halicek, J. V. Little, X. Wang, M. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, B. Fei, *Proc. SPIE Int. Soc. Opt. Eng.* **2018**, *10469*, 104690X.
[35] D. P. Kingma, J. Ba, *arXiv:1412.6980* **2014**.
[36] Z. Liu, H. Wang, Q. Li, *Sensors* **2012**, *12*, 162.
[37] A. Kulcke, A. Holmer, P. Wahl, F. Siemers, T. Wild, G. A. Daeschlein, *Biomed. Tech. (Berl.)* **2018**, *63*, 519.
[38] B. Regeling, B. Thies, A. O. Gerstner, S. Westermann, N. A. Müller, J. Bendix, W. Laffers, *Sensors* **2016**, *16*, 1288.

How to cite this article: D. Eggert, M. Bengs, S. Westermann, N. Gessert, A. O. H. Gerstner, N. A. Mueller, J. Bewarder, A. Schlaefer, C. Betz, W. Laffers, *J. Biophotonics* **2022**, *15*(3), e202100167. https://doi.org/10.1002/jbio.202100167