



# Article AI-Based Detection of Oral Squamous Cell Carcinoma with **Raman Histology**

Andreas Weber <sup>1,2,\*,†</sup>, Kathrin Enderle-Ammour <sup>1,†</sup>, Konrad Kurowski <sup>1,3,4</sup>, Marc C. Metzger <sup>5</sup>, Philipp Poxleitner <sup>5,6,7</sup>, Martin Werner <sup>1,3</sup>, René Rothweiler <sup>5</sup>, Jürgen Beck <sup>6,8</sup>, Jakob Straehle <sup>6,8</sup>, Rainer Schmelzeisen <sup>5,6</sup>, David Steybe <sup>5,6,7,‡</sup> and Peter Bronsert <sup>1,3,4,‡</sup>

- 1 Institute for Surgical Pathology, Medical Center, University of Freiburg, 79106 Freiburg, Germany
- 2 Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany
- 3 Tumorbank Comprehensive Cancer Center Freiburg, Medical Center, University of Freiburg, 79106 Freiburg, Germany
- Core Facility for Histopathology and Digital Pathology, Medical Center, University of Freiburg, 79106 Freiburg, Germany
- Department of Oral and Maxillofacial Surgery, Medical Center, University of Freiburg, 79106 Freiburg, Germany
- Center for Advanced Surgical Tissue Analysis (CAST), University of Freiburg, 79106 Freiburg, Germany
- 7 Department of Oral and Maxillofacial Surgery and Facial Plastic Surgery, University Hospital, LMU Munich, 80337 Munich, Germany
- 8 Department of Neurosurgery, Medical Center, University of Freiburg, 79106 Freiburg, Germany
- Correspondence: and reas.weber.pathologie@uniklinik-freiburg.de
- These authors contributed equally to this work. ‡
  - These authors contributed equally to this work.

Simple Summary: Stimulated Raman Histology (SRH) is a technique that uses laser light to create detailed images of tissues without the need for traditional staining. This study aimed to use deep learning to classify oral squamous cell carcinoma (OSCC) and different non-malignant tissue types using SRH images. The performances of the classifications between SRH images and the original images obtained from stimulated Raman scattering (SRS) were compared. A deep learning model was trained on 64 images and tested on 16, showing that it could effectively identify tissue types during surgery, potentially speeding up decision making in oral cancer surgery.

Abstract: Stimulated Raman Histology (SRH) employs the stimulated Raman scattering (SRS) of photons at biomolecules in tissue samples to generate histological images. Subsequent pathological analysis allows for an intraoperative evaluation without the need for sectioning and staining. The objective of this study was to investigate a deep learning-based classification of oral squamous cell carcinoma (OSCC) and the sub-classification of non-malignant tissue types, as well as to compare the performances of the classifier between SRS and SRH images. Raman shifts were measured at wavenumbers  $k_1 = 2845$  cm<sup>-1</sup> and  $k_2 = 2930$  cm<sup>-1</sup>. SRS images were transformed into SRH images resembling traditional H&E-stained frozen sections. The annotation of 6 tissue types was performed on images obtained from 80 tissue samples from eight OSCC patients. A VGG19-based convolutional neural network was then trained on 64 SRS images (and corresponding SRH images) and tested on 16. A balanced accuracy of 0.90 (0.87 for SRH images) and F1-scores of 0.91 (0.91 for SRH) for stroma, 0.98 (0.96 for SRH) for adipose tissue, 0.90 (0.87 for SRH) for squamous epithelium, 0.92 (0.76 for SRH) for muscle, 0.87 (0.90 for SRH) for glandular tissue, and 0.88 (0.87 for SRH) for tumor were achieved. The results of this study demonstrate the suitability of deep learning for the intraoperative identification of tissue types directly on SRS and SRH images.

Keywords: machine learning; neural networks; pathology; computational biology; head and neck neoplasms



Citation: Weber, A.: Enderle-Ammour. K.: Kurowski, K.: Metzger, M.C.: Poxleitner, P.; Werner, M.; Rothweiler, R.; Beck, J.; Straehle, J.; Schmelzeisen, R.; et al. AI-Based Detection of Oral Squamous Cell Carcinoma with Raman Histology. Cancers 2024, 16, 689. https://doi.org/10.3390/ cancers16040689

Academic Editor: Dania Cioni

Received: 15 January 2024 Revised: 2 February 2024 Accepted: 2 February 2024 Published: 6 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland, This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

# 1. Introduction

Due to their significant influence on recurrence-free survival, the assessments of surgical margins play a pivotal role in the operative treatment of oral squamous cell carcinoma (OSCC) [1–4]. The conventional approach for the intraoperative evaluation of resection margins is based on the preparation of hematoxylin and eosin (H&E)-stained frozen sections [5]. To generate these sections, tissue samples are frozen, sectioned into thin slices, stained, and then microscopically examined by a board-certified pathologist. This technique provides intraoperative real-time feedback on the resection status, enabling surgeons to extend the resection in case of tumor-positive margins [6].

Stimulated Raman Histology (SRH) [7,8] addresses challenges associated with conventional frozen section analysis, which utilizes fiber laser-based Stimulated Raman Scattering (SRS) microscopy to generate images resembling the appearance of H&E-stained tissue sections from fresh tissue specimens without the need for preprocessing. The NIO Laser Imaging System (Invenio Imaging Inc., Santa Clara, CA, USA) is a movable, standalone clinical SRS microscope that facilitates the application of this technique in an intraoperative setting. It measures energy shifts via Raman scattering [9] at wavenumbers of  $k_1 = 2845$  cm<sup>-1</sup> and  $k_2 = 2930$  cm<sup>-1</sup>. As photons with a wavenumber of  $k_1$  scatter mostly at CH<sub>2</sub> bonds, abundant in lipids, and photons with a wavenumber of  $k_2$  scatter at CH<sub>3</sub> bonds, predominant in proteins and DNA, these spectral data depict the spatial distribution of lipids, proteins, and DNA in the tissue sections. To facilitate visual assessment, the data are subsequently subjected to a coloring algorithm included in the NIO Laser Imaging System software version 1.6.0, providing images similar to conventional H&E-stained sections.

In recent years, the field of medicine has witnessed a paradigm shift with the advent of deep learning (DL) applications. Deep learning is a subfield of machine learning that involves the use of artificial neural networks composed of multiple layers of interconnected nodes that are capable of representing complex patterns in data and automatically make predictions without explicit programming [10]. DL has revolutionized various aspects of medicine such as radiology [11], neurology [12], and cardiology [13]. As a promising technique, DL has found its way into the evaluation of histopathological images [14–16]. In this context, DL-based algorithms have exhibited promising outcomes in assessing various tissue and tumor types, including OSCC, in conventional histopathology sections [17]. In the domain of SRH, encouraging results have already been reported when applying this approach to the evaluation of SRH images obtained from brain tumor tissue [18]. Considering the potential of SRH in the assessment of tissue sections from patients with OSCC and the possibilities of DL-driven image evaluations, the combination of these two approaches could offer promising options in an accelerated intraoperative assessment of tissue samples from OSCC patients.

Thus, the objective of the present study was to assess a DL-based approach for the identification of OSCC in SRS and SRH images and for the further sub-classification of non-neoplastic tissues.

## 2. Materials and Methods

#### 2.1. Study Protocol and Sample Acquisition

This prospective study received ethical approval from the Ethics Committee of the University of Freiburg (Reference: #22-1037). Prior to inclusion into the study, informed written consent was obtained from all patients. The inclusion criteria comprised individuals of legal age (>18 years) with biopsy-confirmed OSCC, without prior neoadjuvant therapy and with an indication for surgical resection. In total, 80 tissue specimens were collected from 8 patients as part of this prospective study during the period of May to July 2022.

## 2.2. SRH Image Acquisition

The acquisition of SRH images was performed as previously reported [19]. In brief, native tissue samples, each with a maximum size of  $0.4 \times 0.4 \times 0.2$  cm, were extracted from areas macroscopically suspected of tumor presence, as well as from areas that appeared

macroscopically non-neoplastic. The selection of areas suspected of tumor presence was guided by morphologic criteria such as ulcerative and exophytic alterations. The rationale behind this procedure was to achieve a balanced distribution of tumor-positive and tumorfree tissue samples for subsequent analyses. Each tissue specimen was positioned on a custom microscope slide. The resulting SRS images portray the spatial distribution of the  $CH_2$  and  $CH_3$  bonds across the tissue samples via Raman scattering. These images were obtained through multiple line scans, each spanning a width of 1000 pixels (with a pixel size of 467 nm) and were taken at a depth of 10  $\mu$ m below the coverslip. The pixel size and scanning depth are predetermined settings in the NIO Laser Imaging System. With the scanning time increasing as the pixel size decreases, 467 nm was the preset value chosen to strike a favorable balance between scanning time and image resolution. In terms of scanning depth, the objective was to reduce image noise while simultaneously addressing surface irregularities. In this context, a scanning depth of 10 µm was identified as providing a favorable tradeoff. In the workflow of the NIO Laser Imaging System, the two-channel SRS images were further converted with a special look-up table to provide images reminiscent of conventional H&E-stained slides. Hereinafter, we will refer to images showing the energy shift at the two aforementioned wavenumbers as SRS images, and to images generated with the vendor-specific look-up table, resembling H&E-stained images, as SRH images.

## 2.3. Histopathological Evaluation

In total, 80 SRH images were obtained from 8 patients diagnosed with OSCC. All SRH images were imported into QuPath Version v0.4.3 [20], and the regions of interest within these images were categorized into one of the following six groups: tumor, stroma, adipose tissue, muscle, squamous epithelium, and glandular tissue. Experienced pathologists performed annotations on all 80 images using the wand and brush tools, as well as manual annotations. Annotations were conducted only if the tissue could be clearly classified into one of the aforementioned groups. A total of 877 annotations were made across all 80 specimens, ranging from 1 to 52 annotations per specimen and varying in size. The annotations were directly generated from the SRS images, the annotations made on the SRH images could be seamlessly transferred to the corresponding SRS images (see Figure 1).

#### 2.4. Generation of the Dataset

The SRS images are pixel arrays with two channels, one for storing the scattering values of each of the two wavenumbers. Since DL algorithms suitable for image processing mainly require images with three channels, we populated the first two channels of an empty array with scattering values representing the CH<sub>2</sub> and CH<sub>3</sub> bonds. The third channel of the array was populated with the spectral difference CH<sub>3</sub>-CH<sub>2</sub> for each pixel following the approach of [18]. Subsequently, all labeled SRS and SRH images were divided into tiles measuring  $250 \times 250$  pixels. A tile was considered labeled if at least 99% of its area overlapped with an annotated region; tiles that did not meet this criterion were excluded from the dataset. A threshold of 99% ensures that tiles contain almost only labeled pixels. A lower threshold would include more unlabeled pixels (and therefore non-relevant) or even contradictory information about the label into the respective tile. In effect, this would increase the difficulty of the prediction task for the neural network. Additionally, tiles at the periphery of an annotation where some pixels at a certain edge of the tile are not within the annotation were considered to include sufficient labeled information. The final dataset comprised 21,703 tiles, distributed as follows: 4892 tumor, 4902 stroma, 1471 adipose tissue, 756 muscle, 8461 squamous epithelium, and 1221 glandular tissue tiles.



**Figure 1.** Annotations of tissue classes "Squamous epithelium", "Stroma", and "Tumor" on an SRH image (**A**) and transferred annotations on a corresponding SRS image (**B**) as well as tiles generated from the annotations with class labels "Squamous epithelium", "Stroma", and "Tumor" on a SRH image (**C**) and on the corresponding SRS image (**D**). Only tiles that intersect with an annotation by 99% were kept for the generation of the dataset.

# 2.5. Data Split and Class Distribution

The dataset was divided into a training set comprising 64 images (80% of the dataset) and a test set consisting of 16 images (20% of the dataset). Within the training set, 10% was allocated as a validation set. Table 1 provides an overview of the relative class distribution of the training, validation, and test sets, as well as the entire dataset. Achieving an identical class distribution for all subsets was not possible due to variations in class abundance across different images. To approximate a similar class distribution among all subsets, we iteratively computed the Jensen–Shannon Distance between the class distributions of the subsets. During each iteration, random images from the dataset were selected and added to a subset. An image remained in the respective subset if the Jensen–Shannon Distance between the class distribution of the subset and that of the whole dataset decreased; otherwise, it was discarded, and the next image was considered. This iterative process continued until the desired number of images was included in each subset, with the boundary condition that every class had to be present in each of the subsets.

	Tumor	Stroma	Adipose Tissue	Muscle	Squamous Epithelium	Glandular Tissue
Total	0.23	0.23	0.07	0.03	0.39	0.05
Training set	0.25	0.26	0.06	0.03	0.37	0.03
Validation Set	0.31	0.28	0.17	0.03	0.16	0.05
Test Set	0.24	0.22	0.07	0.04	0.30	0.13

**Table 1.** Relative class distributions for the entire dataset (total), training set, validation set, and test set.

# 2.6. Deep Learning-Based Evaluation of Images

A convolutional neural network (CNN) based on the architecture of VGG19 [21] with randomly initialized weights was used. The choice of a VGG19 architecture was based on the fact that it outperformed other architectures like GoogLeNet [22] or ResNet50 [23] on histology slides in the study presented by Kather et al. [14]. The input dimension of the CNN was (batch size, 250, 250, 3), and the output dimension was (batch size, 6). To account for rotational invariance, all tiles were randomly flipped horizontally and vertically before being fed to the neural network. Two fully connected layers were added to the end of the CNN with 1000 and 100 neurons, respectively. A dropout layer was inserted between the last two fully connected layers with a dropout probability of 0.5 active during training. The CNN was trained for 100 epochs with a batch size of 100 and a learning rate of 0.0001. All computations were performed with Python 3.9.16 and Tensorflow 2.6.0 [24] on a NVIDIA Geforce RTX 4090. Class imbalance was addressed by weighting the loss function during training according to the overall class distribution.

## 2.7. Statistical Evaluation

The statistical evaluation of the CNN performance was conducted using the metrics of precision, recall, and F1-score for each class. Precision is defined as the number of true positives divided by the sum of true positives and false positives, representing the proportion of tiles predicted to belong to a certain class that actually belong to that class. Recall is defined as the number of true positives divided by the sum of true positives and false negatives, representing the proportion of tiles of a certain class correctly predicted. The F1-score, which is the harmonic mean of precision and recall [25], is defined as follows:

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

All metrics have a value range in the closed interval between 0 and 1, where 1 indicates the best performance, and 0 the worst. To assess the overall performance across all classes, we report the balanced accuracy score, which accounts for imbalanced datasets and is the average of the recall scores per class. Additionally, the confusion matrix illustrates the number of tiles predicted to belong to a certain class over the true classes for each class of the test set.

# 3. Results

The neural network demonstrated the ability to detect different tissue types on SRS images (and corresponding SRH images) with an overall balanced accuracy of 0.90 (0.87). Stromal tissue tiles exhibited a precision of 0.90 (0.90) and a recall of 0.91 (0.92). Adipose tissue tiles displayed a precision of 0.97 (0.99) and a recall of 0.98 (0.94). The performance for squamous epithelium, while slightly lower, still achieved a precision of 0.89 (0.82) and a recall of 0.90 (0.84). Tiles belonging to the muscle class were identified with a precision of 0.95 (0.79) and a recall of 0.89 (0.73). For glandular tissue, the precision was 0.92 (0.96), with a recall of 0.82 (0.85). The identification of squamous cell carcinoma resulted in a precision of 0.86 (0.92) and a recall of 0.90 (0.82), leading to an F1-score of 0.88 (0.87).

An example of ground truth class labels for each tile and corresponding predictions for SRS and SRH image can be seen in Figure 2. For stroma tissue, the metrics reveal a better performance of the neural network on the SRH images compared to SRS images, and for glandular tissue, a similar performance was observed. However, an overall comparison of the balanced accuracy reveals a slight outperformance of the neural network on the SRS images compared to the performance on the SRH images. Regarding the detection of OSCC, the performances on both data modalities show similar results.



**Figure 2.** Ground truth class labels for each tile (**A**) and predicted class labels for each tile (**B**) on a sample SRS image. Both true tiles with class label "Stroma" were classified correctly, whereas 6 tiles with class label "Tumor" were incorrectly classified as "Squamous epithelium" (5 tiles) and "Stroma" (1 tile). Ground truth class labels for each tile (**C**) and predicted class labels for each tile (**D**) on a sample SRH image. Both true tiles with class label "Stroma" were classified correctly, whereas 8 tiles with class label "Tumor" were incorrectly classified as "Squamous epithelium" (5 tiles) and "Stroma" (3 tiles).

The dataset's class imbalance appears to have been effectively addressed by the weighting of the loss function during training, as underrepresented classes did not exhibit worse performances than overrepresented ones. Detailed performance metrics for all classes on the SRS images (and SRH images, respectively) can be found in Table 2. Figure 3

illustrates the confusion matrices, displaying the number of predicted and true labels for each tile in the test.

**Table 2.** Performance metrics of the CNN on the test set on SRS images (SRH images) as well as the number of tiles for each class in the test set.

	Precision	Recall	F1-Score	Number of Tiles
Stroma	0.90 (0.90)	0.91 (0.92)	0.91 (0.91)	1035
Adipose tissue	0.97 (0.99)	0.98 (0.94)	0.98 (0.96)	351
Squamous epithelium	0.89 (0.82)	0.90 (0.94)	0.90 (0.87)	1393
Muscle	0.95 (0.79)	0.89 (0.73)	0.92 (0.76)	206
Glandular tissue	0.92 (0.96)	0.82 (0.85)	0.87 (0.90)	607
Tumor	0.86 (0.92)	0.90 (0.82)	0.88 (0.87)	1138



**Figure 3.** Confusion matrices for the classification of the CNN on the SRS test dataset (**left**) and the corresponding SRH test dataset (**right**). The diverging colormap shows small values in dark blue with increasing brightness according to increasing values. Large values are shown in dark red with decreasing brightness according to increasing values.

For both SRS and SRH images, the confusion matrices highlight that the primary source of error lies in the confusion between squamous epithelium and tumor. Specifically, for SRS images, the neural network misclassified 8% (112 out of 1393) of squamous tissue tiles as tumor and 7% (85 out of 1138) of tumor tiles as squamous tissue. For SRH images, the neural network misclassified 4% (58 out of 1393) of squamous tissue tiles as tumor tiles and 17% (190 out of 1138) of tumor tiles as squamous tissue tiles.

# 4. Discussion

This study aimed to assess the performance of a DL-based identification of oral squamous cell carcinoma and the subclassification of non-malignant tissue on SRS and SRH images. SRH is a novel technology that provides images reminiscent of conventional H&E-stained slides without the need for sectioning and staining. By leveraging the inherently digital nature of these images, DL-based evaluation has the potential to expedite diagnoses and facilitate intraoperative surgical decision making.

Deep learning applications in pathology have made significant advancements in recent years, encompassing various cancer types and supporting diagnosis, classification, grading, and staging methods [26–29]. In the field of pathohistological diagnosis, promising results have emerged for the automated detection of numerous cancers in whole-slide images of conventional histology slides. This includes OSCC, where current investigations have reported F1-scores > 0.90 [17,30].

In the present study, a convolutional neural network based on the VGG19 architecture [21] with randomly initialized weights was used. The CNN was trained to assign one of six class labels to tiles of SRS and SRH images. The performance on the hold-out test set demonstrates the neural network's ability to identify OSCC and to subclassify non-malignant tissue types such as stroma, adipose tissue, squamous epithelium, muscle, and glandular tissue. The CNN's performance varied across tissue types, with the highest performance observed for adipose tissue (F1-score of 0.98 on SRS images and 0.96 on SRH images), and the lowest for glandular tissue (F1-score of 0.87 on SRS images) and muscle tissue (F1-score of 0.76 on SRH images). Possible explanations for this finding could be the variations in tissue composition, heterogeneity of tissue structure, and the fact that certain tissue types are represented more characteristically in SRH than others (e.g., the characteristic representation of adipose tissue with deep-purple vacuoles).

Regarding the confusion between squamous epithelium and tumor, it must be considered that normal mucosa can be difficult to distinguish from carcinoma cells, especially from precursor lesions at the cellular level in SRH morphology.

While this study represents the first exploration of a DL-based evaluation of SRH images from OSCC patients, results are available from previous investigations in the field of neurosurgery, where SRH originated [18]. These findings include the use of patch-level CNN predictions for glioma recurrence detection, achieving a diagnostic accuracy of 95.8% and the use of the Inception-ResNet-v2 architecture for detecting common central nervous system tumors, with a CNN-based diagnosis of SRH images shown to be non-inferior to a pathologist-based interpretation of conventional histologic images (overall accuracy, 94.6% vs. 93.9%). Additionally, the automated analysis of skull base tumor specimens for the detection of various tumor types has applied ResNet architectures that resulted in overall diagnostic accuracies of 91.5% with cross-entropy, 83.9% with self-supervised contrastive learning, and 96.6% with supervised contrastive learning.

The CNN-based prediction of brain tumor diagnosis reported by [18] yielded a mean class accuracy of 89.2% (at the patient level), which is comparable to the mean class accuracy (equivalent to the mean of the recall values) of 90.0% for SRS images (and 86.7% for SRH images) presented in our study. The authors of [18] predicted 13 class labels using a training dataset comprising over 2.5 million labeled tiles from 415 patients. In contrast, our study predicted six class labels with a training dataset comprising a number of labeled tiles several orders of magnitude lower (16,973). Considering the relatively low number of tiles in the training set, the performance of the CNN is remarkable. Factors that complicate a direct comparison between the results reported by [18] and our study include differences in CNN architecture (Inception-ResNet-v2 in [18] vs. VGG19-based in our study), the covered inter-patient variety of the dataset (415 patients in [18] vs. 8 patients in our study), and the disease-specific complexity of the mapping between morphology and histological classification (brain tumor in [18] vs. OSCC in our study).

While the present study yielded promising results, the limitations of this method could include its use on rare tumor entities, tumor precursors, or distinctly large inflammatory reactive changes. Further the training of DL algorithms is needed for tumor entities not included in the study so far, such as rare subtypes of adenocarcinomas or hematological illnesses or lymphomas.

To enhance the neural network's performance, larger and more diverse datasets are imperative. On the algorithmic front, aggregating predictions from multiple neural networks through ensemble learning could lead to increased stability and performance improvements. However, high throughput is not yet possible due to the described restrictions and monitoring.

A clear advantage of SRH is that, in contrast to H&E and other routinely used staining procedures and the associated laboratory-triggered deviations, the technique, and thus the image to be assessed, do not differ. This can provide internationally standardized, homogeneous modularity. In the future, a globally accessible open-source database with scattering spectra for comparable histology is conceivable.

For a DL-based classifier to be effectively deployed in clinical practice, it must not only demonstrate high performance metrics but also exhibit stability across diverse datasets and remain resilient to performance errors caused by slight variations in input data. Moreover, issues of trustworthiness and transparency arise, especially in the medical field, where incorrect predictions can lead to erroneous diagnoses with potentially severe consequences for patients.

# 5. Conclusions

In recent years, there has been a concerted effort to digitize and automate histopathological workflows. Within this context, Raman scattering has emerged as a promising technology, and the inherently digital format of images obtained from Stimulated Raman Scattering and Stimulated Raman Histology provides an ideal foundation for deep learning-based image analyses. The results of this study illustrate the significant potential of integrating SRH and deep learning to advance the digitization of workflows in the surgical treatment of oral squamous cell carcinoma.

Author Contributions: Conceptualization, A.W., K.E.-A., K.K., M.C.M., P.P., R.R., J.B., J.S., R.S., D.S. and P.B.; methodology, A.W., K.E.-A., K.K., M.C.M., P.P., R.R., J.B., J.S., R.S., D.S. and P.B.; software, A.W.; validation, A.W., K.E.-A., D.S. and P.B.; formal analysis, A.W., K.E.-A., K.K., D.S. and P.B.; investigation, A.W. and K.E.-A.; resources, M.C.M., P.P., R.R., J.B., J.S., R.S., D.S. and P.B.; data curation, A.W. and K.E.-A.; writing—original draft preparation, A.W., K.E.-A., D.S. and P.B.; writing—review and editing, A.W., K.E.-A., K.K., M.C.M., P.P., M.W., R.R., J.B., J.S., D.S. and P.B.; visualization, A.W.; supervision, M.W., R.S., D.S. and P.B.; project administration, D.S. and P.B.; funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF), grant number 13GW0571D.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the University of Freiburg (protocol code #22-1037, 22 February 2022).

**Informed Consent Statement:** Written informed consent was obtained from the patient(s) to publish this paper.

**Data Availability Statement:** Data supporting the findings of the study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors would like to acknowledge Florian Khalid for his valuable technical assistance.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

### References

- Dillon, J.K.; Brown, C.B.; McDonald, T.M.; Ludwig, D.C.; Clark, P.J.; Leroux, B.G.; Futran, N.D. How Does the Close Surgical Margin Impact Recurrence and Survival When Treating Oral Squamous Cell Carcinoma? J. Oral Maxillofac. Surg. 2015, 73, 1182–1188. [CrossRef]
- Hinni, M.L.; Ferlito, A.; Brandwein-Gensler, M.S.; Takes, R.P.; Silver, C.E.; Westra, W.H.; Seethala, R.R.; Rodrigo, J.P.; Corry, J.; Bradford, C.R.; et al. Surgical Margins in Head and Neck Cancer: A Contemporary Review. *Head Neck* 2013, 35, 1362–1370. [CrossRef] [PubMed]
- Loree, T.R.; Strong, E.W. Significance of Positive Margins in Oral Cavity Squamous Carcinoma. Am. J. Surg. 1990, 160, 410–414. [CrossRef] [PubMed]
- 4. Li, M.M.; Puram, S.V.; Silverman, D.A.; Old, M.O.; Rocco, J.W.; Kang, S.Y. Margin Analysis in Head and Neck Cancer: State of the Art and Future Directions. *Ann. Surg. Oncol.* **2019**, *26*, 4070–4080. [CrossRef] [PubMed]
- Gal, A.A.; Cagle, P.T. The 100-year anniversary of the description of the frozen section procedure. JAMA 2005, 294, 3135–3137. [CrossRef] [PubMed]
- Ord, R.A.; Aisner, S. Accuracy of Frozen Sections in Assessing Margins in Oral Cancer Resection. J. Oral Maxillofac. Surg. 1997, 55, 663–669; discussion 669–671. [CrossRef] [PubMed]

- 7. Freudiger, C.W.; Min, W.; Saar, B.G.; Lu, S.; Holtom, G.R.; He, C.; Tsai, J.C.; Kang, J.X.; Xie, X.S. Label-free biomedical imaging with high sensitivity by stimulated raman scattering microscopy. *Science* **2008**, *322*, 1857–1861. [CrossRef] [PubMed]
- Orringer, D.A.; Pandian, B.; Niknafs, Y.S.; Hollon, T.C.; Boyle, J.; Lewis, S.; Garrard, M.; Hervey-Jumper, S.L.; Garton, H.J.L.; Maher, C.O.; et al. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nat. Biomed. Eng.* 2017, 1, 0027. [CrossRef] [PubMed]
- 9. Raman, C.V.; Krishnan, K.S. The optical analogue of the Compton effect. *Nature* 1928, 121, 711. [CrossRef]
- 10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436-444. [CrossRef]
- 11. Heiliger, L.; Sekuboyina, A.; Menze, B.; Egger, J.; Kleesiek, J. Beyond Medical Imaging—A Review of Multimodal Deep Learning in Radiology. *TechRxiv* 2022. [CrossRef]
- 12. Fernández, I.S.; Peters, J.M. Machine learning and deep learning in medicine and neuroimaging. *Ann. Child Neurol. Soc.* 2023, 1, 102–122. [CrossRef]
- 13. Wehbe, R.M.; Katsaggelos, A.K.; Hammond, K.J.; Hong, H.; Ahmad, F.S.; Ouyang, D.; Shah, S.J.; McCarthy, P.M.; Thomas, J.D. Deep Learning for Cardiovascular Imaging. *JAMA Cardiol.* **2023**, *8*, 1089. [CrossRef] [PubMed]
- Kather, J.N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.-A.; Gaiser, T.; Marx, A.; Valous, N.A.; Ferber, D.; et al. Predicting Survival from Colorectal Cancer Histology Slides Using Deep Learning: A Retrospective Multicenter Study. *PLoS Med.* 2019, 16, e1002730. [CrossRef] [PubMed]
- 15. Rawat, R.R.; Ortega, I.; Roy, P.; Sha, F.; Shibata, D.; Ruderman, D.; Agus, D.B. Deep Learned Tissue "Fingerprints" Classify Breast Cancers by ER/PR/Her2 Status from H&E Images. *Sci. Rep.* **2020**, *10*, 7275. [CrossRef] [PubMed]
- Calderaro, J.; Kather, J.N. Artificial Intelligence-Based Pathology for Gastrointestinal and Hepatobiliary Cancers. *Gut* 2021, 70, 1183–1193. [CrossRef] [PubMed]
- 17. Yang, S.; Li, S.; Liu, J.; Sun, X.; Cen, Y.; Ren, R.; Ying, S.; Chen, Y.; Zhao, Z.; Liao, W. Histopathology-Based Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning. *J. Dent. Res.* **2022**, *101*, 1321–1327. [CrossRef] [PubMed]
- Hollon, T.C.; Pandian, B.; Adapa, A.R.; Urias, E.; Save, A.V.; Khalsa, S.S.S.; Eichberg, D.G.; D'amico, R.S.; Farooq, Z.U.; Lewis, S.; et al. Near Real-Time Intraoperative Brain Tumor Diagnosis Using Stimulated Raman Histology and Deep Neural Networks. *Nat. Med.* 2020, 26, 52–58. [CrossRef]
- Steybe, D.; Poxleitner, P.; Metzger, M.C.; Rothweiler, R.; Beck, J.; Straehle, J.; Vach, K.; Weber, A.; Enderle-Ammour, K.; Werner, M.; et al. Stimulated Raman Histology for Histological Evaluation of Oral Squamous Cell Carcinoma. *Clin. Oral Investig.* 2023, 27, 4705–4713. [CrossRef]
- 20. Bankhead, P.; Loughrey, M.B.; Fernández, J.A.; Dombrowski, Y.; McArt, D.G.; Dunne, P.D.; McQuaid, S.; Gray, R.T.; Murray, L.J.; Coleman, H.G.; et al. QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci. Rep.* **2017**, *7*, 16878. [CrossRef]
- Karen, S.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* 2015, arXiv:1409.1556.
   [CrossRef]
- 22. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.; Liu, W.; et al. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]
- 23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
- Martín, A.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2015. Available online: http://download.tensorflow.org/ paper/whitepaper2015.pdf (accessed on 11 October 2023).
- 25. Yutaka, S. The truth of the F-measure. Teach. Tutor Mater. 2007, 1, 1–5.
- 26. Liu, M.; Hu, L.; Tang, Y.; Wang, C.; He, Y.; Zeng, C.; Lin, K.; He, Z.; Huo, W. A Deep Learning Method for Breast Cancer Classification in the Pathology Images. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5025–5032. [CrossRef] [PubMed]
- 27. Echle, A.; Rindtorff, N.T.; Brinker, T.J.; Luedde, T.; Pearson, A.T.; Kather, J.N. Deep Learning in Cancer Pathology: A New Generation of Clinical Biomarkers. *Br. J. Cancer* **2021**, *124*, 686–696. [CrossRef] [PubMed]
- Farahani, H.; Boschman, J.; Farnell, D.; Darbandsari, A.; Zhang, A.; Ahmadvand, P.; Jones, S.J.M.; Huntsman, D.; Köbel, M.; Gilks, C.B.; et al. Deep Learning-Based Histotype Diagnosis of Ovarian Carcinoma Whole-Slide Pathology Images. *Mod. Pathol.* 2022, 35, 1983–1990. [CrossRef] [PubMed]
- Xie, W.; Reder, N.P.; Koyuncu, C.F.; Leo, P.; Hawley, S.; Huang, H.; Mao, C.; Postupna, N.; Kang, S.; Serafin, R.; et al. Prostate Cancer Risk Stratification via Nondestructive 3D Pathology with Deep Learning–Assisted Gland Analysis. *Cancer Res.* 2022, 82, 334–345. [CrossRef] [PubMed]
- 30. Liu, Y.; Bilodeau, E.; Pollack, B.; Batmanghelich, K. Automated Detection of Premalignant Oral Lesions on Whole Slide Images Using Convolutional Neural Networks. *Oral Oncol.* **2022**, *134*, 106109. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.