Understanding Bias in Multispectral Autofluorescence Lifetime Imaging: Are Models Sensitive to Oral Location?

Kayla Caughlin[1], Rodrigo Cuenca Martinez[2], Gabriel P. Tortorelli[2],
Kathleen E. Higgins, DDS, MS[3], Ronald Faram, DDS[3], Javier A. Jo[2], and Carlos Busso[1]

***Abstract*— While bias in artificial intelligence is gaining attention across applications, model fairness is especially concerning in medical applications because a person's health may depend on the model outcome. Sources of bias in medical applications include age, gender, race, and social history. However, in oral cancer diagnosis, the oral location may be a source of bias. Variability in performance based on the oral location has been reported but is not well understood. To help ensure that models perform equitably regardless of location, we design three experiments to study the effect of oral location on model performance. We show that multispectral autofluorescence images retain tissue-type characteristics, but that the tissue-specific information is degraded in lesion images. Furthermore, we show that the tissue-specific features are not disentangled from the disease-associated features. Our results show that automated diagnosis models need to be thoughtfully designed to remove bias from the oral location to ensure equitable performance. Based on these insights, we propose a tissue-specific fine-tuning approach that increases overall performance and lowers the fairness gap by over 5%.**

***Clinical relevance*— This paper explores sources of off-target variance in multispectral autofluorescence images. By understanding sources of bias in multispectral autofluorescence images, fairer and more robust models for oral cancer diagnosis and margin delineation can be developed, leading to greater clinical acceptance and more equitable patient outcomes.**

I. INTRODUCTION

As machine learning and deep learning models have increasingly become integrated into our daily lives, questions of bias and fairness in modeling have also emerged as important considerations across disciplines [1]–[4]. Particularly in the medical device industry, questions of fairness and interpretability are important for clinical translation. Patients want to know that their gender, race, or social history will not decrease the likelihood of accurate diagnosis and successful treatment. Likewise, clinicians want assurance that a model result is not skewed by training data bias or data collection techniques. Based on these concerns, researchers are exploring bias in automated diagnosis applications. For example, Larrazabal et al. [5]showed that gender imbalance in x-ray training data sets can negatively impact model performance. The authors show changes in performance based on the ratio between males and females in the training data, leading to the worst performance when training with data from one gender and testing with data from the other gender [5]. Lin et al. [6] explore age, race, and sex biases using a pairwise fairness difference metric in several automated diagnosis tasks, including COVID-19 detection from x-rays. Though differences in age, gender, and race are more commonly discussed as sources of bias, numerous factors may contribute to bias. Drukker et al. [7] describe 29 categories of potential bias in medical applications of AI.

Despite promising performance in oral cancer diagnosis and margin delineation [8]–[11], sources of bias in *multispectral autofluorescence imaging* (maFLIM) have not been well-studied. Marsden et al. [12] showed different classification performances based on the tissue type. However, the reason for a difference in performance based on the tissue type remains an open question. Several sources of bias could result in different performances based on the tissue type. For example, oral cancer is more common in the tongue than in the cheek [13] so the difference in performance could be due to the underlying disease process being more common in a specific tissue type, rather than information about the tissue type. However, the images could retain information about the tissue type that is unrelated to the disease process. While other sources of bias may also include gender, tobacco history, alcohol history, race, or lesion size, we begin to narrow these down by focusing on the tissue type. This study explores the following questions:

- Do maFLIM images retain information about the tissue type?
- Is tissue type information degraded in lesion images?
- Are the features that separate diseased tissue from normal tissue different from the features that indicate the oral location? i.e., can we remove oral location information by picking the right subset of features?

By answering the above questions, we hope to guide the development of fairer models for oral cancer and margin delineation that perform equitably regardless of the tissue type where the lesion arises.

Our analysis indicates that maFLIM images do retain information about the tissue type and that we cannot easily remove this information by carefully selecting features. However, our proposed tissue-specific fine-tuning method improves overall performance by 1.76% to 5.16% on a disease-related task. Our proposed model also shows less bias towards a specific oral location.

II. IMAGING SYSTEM

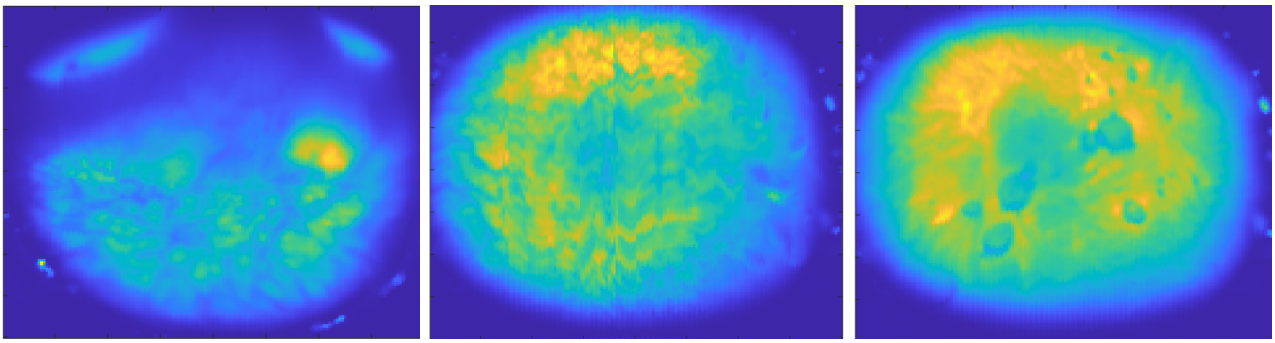
The maFLIM system is a dual-excitation, frequency domain imaging system [14], [15]. The system produces dif-

*This work was supported by NIH R01:5R01CA218739

[1]The Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA. email@school.edu

[2]The School of Electrical and Computer Engineering, The University of Oklahoma, Norman, OK, USA 73019.

[3]Ronald Faram, DDS are with The University of Oklahoma College of Dentistry, Oklahoma City, OK, USA 73117.



(a) gingiva (b) tongue (c) buccal mucosa

Fig. 1: Tissue-Based Differences and Image Artifacts. Left - Gingiva with tooth base, free gingiva, and attached gingiva. Center - Tongue with motion artifact. Right - Buccal mucosa with saliva bubbles.

ferent outputs than the data produced by the time-domain-based system used in our previous studies [8]–[10], [16]. Previously, we collected a time series at each pixel. In the frequency domain system, each pixel contains a complex number for each excitation-emission pair. The imaging system used for data collection uses two excitation wavelengths and collects the resulting autofluorescence in three spectral bands for the lower excitation and four spectral bands for the higher excitation, resulting in a total of seven excitation-emission combinations. The justification for the choice of excitation and emission bands in relation to the development of oral cancer has been established by multiple previous studies [17]–[25]. From the complex representation at each pixel, a total of fifty features were calculated, including intensity, modulation, and phase features. The excitation-emission bands were designed to target structural and metabolic properties of oral tissue that may be altered in disease processes. However, these bands may also contain information about the tissue type. Visually, we can see three different regions in gingiva images (see Fig. 1a). The three regions appear to correspond with the base of two teeth (top of image), a narrow band of free gingiva (unkeratinized tissue), then the attached gingiva (keratinized tissue) that takes up the majority of the image. Similarly, different tissue types tend to have different imaging artifacts. The tongue is difficult to hold still during imaging, and motion is commonly visualized as blurring in the intensity image (see Figure 1b and the vertical line towards the center of the tongue image). Saliva bubbles can accumulate on the buccal mucosa and appear as circular regions in the intensity image (see Figure 1c). These qualitative observations hint that maFLIM images vary based on location to some extent. However, these images only show the intensity information and do not quantitatively establish that the maFLIM images retain tissue type information.

III. DATA

The data was collected at The University of Oklahoma with *institutional review board* (IRB) approval (IRB:12637). Data collection at this clinical site is ongoing and recruitment has already increased since this analysis began. For this research, we include images from the first 80 patients. The data collection protocol differs based on the type of patient. For patients without a lesion, images were taken from all oral locations: lip, cheek, palate, tongue, and gingiva. For patients with a lesion, at least two images are taken: one or more lesion images and a normal image from the contralateral side of the mouth. If possible, multiple images are taken from the lesion, with one at the center and several at the edges of the lesion. For this work, we exclude the images at the edges of the lesion because these images contain both normal and lesion pixels as visually assessed. Each pixel in the image contains 50 features including intensity, modulation, and phase features. While these features are difficult to visualize in a meaningful way, we can see some information by plotting the summed intensity maps across all excitation-emission bands (see Fig. 1).

Although visually inspecting the maFLIM intensity images leads to some intuition that the images retain tissue type information, we want to study the differences quantitatively and understand if tissue information is retained in all maFLIM image features. To remove the effect of changes due to the disease process, we first study only the normal images from patients coming from lip, cheek, and tongue locations. We focus on lip, cheek, and tongue images to avoid complications from the gingiva and palate artifacts. In gingiva images, teeth often take up a large, but variable, portion of the image. Gingiva images frequently have teeth in the images as well as two types of tissue: attached gingiva and free gingiva. The palate is more difficult to access and thus may be out of focus, leading to complications in the analysis. In addition, since the difference in tissue type may be smaller than the difference between patients, we compare each patient's lip and cheek images to their tongue image. Since the inside of the lip and buccal mucosa are very similar, we expect the maFLIM lip and cheek images to be similar, but different from the tongue images for each patient. Figure 2 shows the percentage of patients where the feature difference between lip and cheek images is smaller than the average of the cheek-to-tongue and lip-to-tongue differences. We conduct this analysis for each feature. We highlight the 50% corresponding to a random level with a red line. If the bars are near this value, we can conclude that the features

are not dependent on their location. Figure 2 shows that a majority of the bars are above the random level percentage, indicating that the features of the cheek and lip are similar but different from the features of the tongue. Normal images contain tissue-type information. However, Figure 2 indicates that some features are less descriptive of the tissue type than others. Given that the previous analysis only considered normal images, we raise the following questions:

- Is tissue type information degraded in lesion images?
- Are the features that separate diseased tissue from normal tissue different from the features that indicate the oral location? i.e., can we remove oral location information by picking the right subset of features?

We explore these questions in the following section.

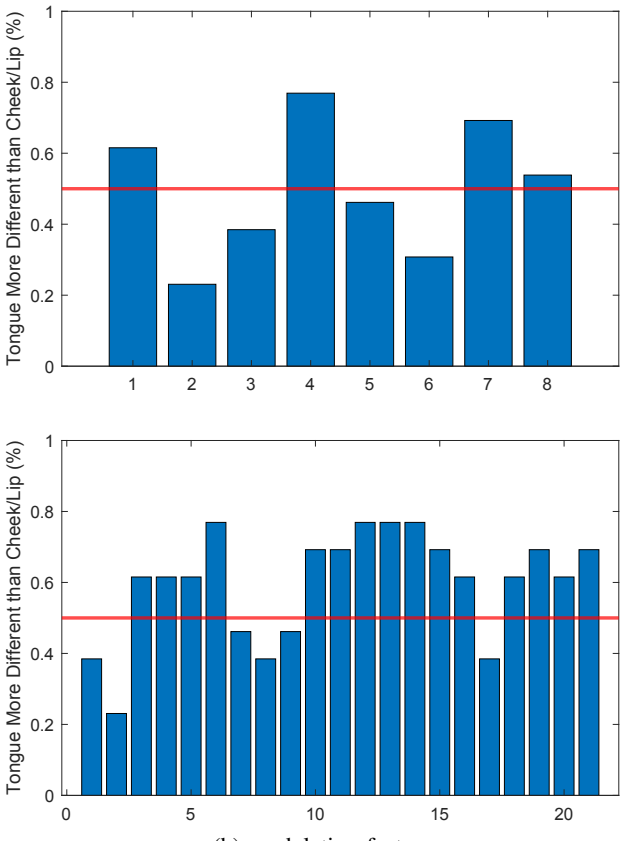
IV. ANALYSIS OF TISSUE-LEVEL DIFFERENCES

A. Understanding Tissue-Level Differences: SVM Models

To address the questions mentioned in Section III, we conduct discriminative analysis using *support vector machine* (SVM). We use leave-one-patient-out cross-validation. In each fold, we train the models with the data from all patients except one, testing the model on the data from this patient. We use the L1 regularization such that the less-relevant features for the specific task are set to zero. We set the regularization parameter such that the level of sparsity is similar for each task. Specifically, we set C = 1e-5 for the tissue-type task and C = 5e-5 for the disease-related task.

Based on the plots from normal images in Section III, we expect that a model will be able to predict the tissue type in maFLIM images from normal tissue. However, the disease process in lesion images may alter the characteristics of the tissue such that tissue type information is degraded. To understand how the disease process may affect the tissue type characteristics, we train a *support vector machine* (SVM) as a binary classifier to predict if an image is from the cheek/lip area or from the tongue area. We first train the model using only normal images. As shown in Table I, the testing accuracy on normal images is 76.47%. Next, we train the models using both normal and lesion images, and test on the normal images. If the tissue type information is similar in both normal and lesion images, we expect that the larger amount of training data will increase the test performance. However, Table I, column 2, shows that adding the lesion images to the training set decreases test performance on the normal images in all metrics. The maximum decrease in performance is a 8.82% decrease in sensitivity, indicating that the tissue type information in normal images is degraded in the lesion images.

We repeat the experiment by testing on lesion images instead of normal images. The results are shown in Table II. Training the model on normal images and testing on lesion images shows near-random performance, with an accuracy of 57.78%, further confirming that the tissue-specific information in lesion images is degraded. However, when the lesion training data is added to the normal training data, the lesion test accuracy increases by 6.66%. The increase in lesion



(c) phase features

Fig. 2: Percentage of patients where the feature difference between lip and cheek images is smaller than the average of the cheek-to-tongue and lip-to-tongue differences.

test performance after adding lesion data to the training set indicates that some tissue-type information is retained in lesion images. However, this information is not the same as that contained in the normal images. Since different pathologies have a greater likelihood of development in different areas of the mouth [13], these results may be reflective of a disease process that occurs more frequently in a specific tissue type, rather than information actually related to the properties of that tissue type. For example, the classifier may be finding a relation from the data that lichen planus often occurs bilaterally in the buccal mucosa, while dysplasia and squamous cell carcinoma are more likely

TABLE I: Lesion training data degrades tissue type performance in normal images. Center column: normal image test performance on tissue type identification task using model trained on normal images only. Right column: normal image test performance on tissue type identification task using model trained on normal and lesion images.

Training Data	Normal Only	Normal + Lesion
Sensitivity	82.35	73.53
Specificity	72.55	70.59
Average	77.45	72.06
Precision	66.67	62.50
Accuracy	76.47	71.76
F1 Score	73.68	67.57

TABLE II: Tissue type information is degraded in lesion images. Center column: lesion image test performance on tissue type identification task using model trained on normal images only. Right column: lesion image test performance on tissue type identification task using model trained on normal and lesion images.

Training Data	Normal Only	Normal + Lesion
Sensitivity	95.24	90.48
Specificity	25.00	41.67
Average	60.12	66.07
Precision	52.63	57.58
Accuracy	57.78	64.44
F1 Score	67.80	70.37

to occur in the tongue.

The results from Tables I and II lead to the question - are the features that distinguish tissue type in normal images disentangled from the features that are relevant for distinguishing abnormal tissue from normal tissue? Basically, can we simply not use certain features that are related to tissue type when building models focused on aspects of the disease process? To explore these questions, we train two different SVM models: one that is trained to distinguish the tissue type and another that is trained to classify between lesion and normal images. We use the normal-lesion task as a proxy task to find features likely informative for disease-related tasks such as cancer diagnosis and margin delineation. The tissue-type type distinguishes the different oral locations. For each task, we identify the most discriminative features by selecting the top non-zero features after the L1-regularization. These features are the ones that are non-zero for at least 50% of the leave-one-patient out cross validation runs. Then, we train models by swapping the most discriminative features across both tasks (i.e., training the disease-related task with the tissue-type task features and vice versa). If the features are disentangled, we expect that training on the swapped features will reduce performance to near-random levels. Table III shows the results. For the tissue-type task, training only on the features most relevant for the disease-related task drastically decreases performance, with accuracy dropping more than 8%. The metrics are still well over random performance. For the disease-related task, accuracy drops by less than 1% when training on the swapped features. Combined with the fact that five of the fifty features are selected in both tasks, our results indicate that the features

TABLE III: Disease-related and tissue-type features are not entirely disentangled. NL: normal vs. lesion task. NL Swap: normal vs. lesion task with the model trained using top features from tissue-type task. TT: tissue-type identification task. TT Swap: tissue type identification task with the model trained using top features from the disease-related task.

Model	NL	NL Swap	TT	TT Swap
Sensitivity	75.56	82.22	82.35	67.65
Specificity	80.00	75.29	72.55	68.63
Average	77.78	78.76	77.45	68.14
Precision	66.67	63.79	66.67	58.97
Accuracy	78.46	77.69	76.47	68.24
F1 Score	70.83	71.84	73.68	63.01

TABLE IV: Model performance increases with tissue-specific fine-tuning.

Model Type	Baseline	Fine-Tuned
Sensitivity	68.89	68.89
Specificity	84.71	88.24
Average	76.80	78.56
Precision	70.45	75.61
Accuracy	79.23	81.54
F1 Score	69.66	72.09

are not disentangled. Therefore, we cannot ensure fair models simply by excluding features that contain information about the tissue type.

B. Proposed Tissue-Specific Fine-Tuning

If a very large training database existed, training a separate model for each oral location might be a good strategy. With the moderate number of patients in our data set, a trade-off exists: we may increase performance by training a model only on a specific location due to lower off-target variability. However, we also have a higher risk of overfitting due to an even smaller training set. As a proof-of-concept for how we can balance the trade-off, we implement a tissue-specific fine-tuning model. In the tissue-specific fine-tuning model, we can take advantage of the full training data by training a small neural network base model to identify lesion images, and then fine-tune the model for each specific oral location with the most relevant data. The fine-tuning step can slightly adjust the model specifically for the given oral location while reducing the likelihood of overfitting to the smaller training set by lowering the learning rate or freezing some layers. The results are shown in Table IV. The fine-tuned model performance meets or exceeds baseline performance on all metrics. The increases in performance range from 1.76% to 5.16%. Fairness also increases, with the lip/cheek model accuracy staying constant at 86.67%. The tongue model accuracy increases from 69.09 to 74.55%.

V. CONCLUSIONS

Our analysis indicates that information regarding the tissue type is contained in normal maFLIM images, but that this information is degraded in the lesion images. Furthermore, we showed that the descriptive features for tissue type identification are not disentangled from the features that are descriptive for tasks related to disease pathology. Therefore,

we cannot simply pick the right features and ensure that cancer diagnosis or margin delineation models perform equally well in all oral locations. We can, however, build models specific for an oral location and avoid overfitting by fine-tuning a tissue-specific model. We use the normal versus lesion task as proof-of-concept, with the fine-tuned model performance increasing or meeting baseline performance on all metrics. In addition, the model decreases the fairness gap by over 5%.

In future work, we are interested in using the degradation of tissue-type information in the lesion images as a way to select which pixels most likely correspond to lesion tissue in images of small lesions that do not take up the entire image field of view. For example, if certain pixels in a lesion image from the tongue are miss-classified as cheek, the miss-classified pixels may used to select the most likely lesion pixels in the image. These approaches may help ensure that models work equally well in all tissue types, as well as for lesions of varying sizes.

REFERENCES

[1] S. Wehrli, C. Hertweck, M. Amirian, S. Glüge, and T. Stadelmann, "Bias, awareness, and ignorance in deep-learning-based face recognition," *AI and Ethics*, vol. 2, pp. 509–522, August 2022.

[2] R. Navigli, S. Conia, and B. Ross, "Biases in large language models: Origins, inventory and discussion," *ACM Journal of Data and Information Quality*, vol. 15, no. 2, pp. 1–21, June 2023.

[3] B. Çaylak, "Issues that may arise from usage of AI technologies in criminal justice and law enforcement," in *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence*, M. Kılıç and S. Bozkuş Kahyaoğlu, Eds. Singapore: Springer, October 2023, pp. 119–132.

[4] J. Gichoya, K. Thomas, L. A. Celi, N. Safdar, I. Banerjee, J. Banja, L. Seyyed-Kalantari, H. Trivedi, and S. Purkayastha, "AI pitfalls and what not to do: mitigating bias in AI," *The British Journal of Radiology*, vol. 96, no. 1150, p. 20230023, October 2023.

[5] A. Larrazabal, N. Nieto, V. Peterson, D. Milone, and E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," *Proceedings of the National Academy of Sciences*, vol. 117, no. 23, pp. 12 592–12 594, May 2020.

[6] M. Lin, T. Li, Y. Yang, G. Holste, Y. Ding, S. Van Tassel, K. Kovacs, G. Shih, Z. Wang, Z. Lu, F. Wang, and Y. Peng, "Improving model fairness in image-based computer-aided diagnosis," *Nature Communications*, vol. 14, p. 6261, October 2023.

[7] K. Drukker, W. Chen, J. Gichoya, N. Gruszauskas, J. Kalpathy-Cramer, S. Koyejo, K. Myers, R. Sá, B. Sahiner, H. Whitney, Z. Zhang, and M. Giger, "Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment," *Journal of Medical Imaging*, vol. 10, no. 6, p. 061104, April 2023.

[8] E. Duran-Sierra, S. Cheng, R. Cuenca, B. Ahmed, J. Ji, V. Yakovlev, M. Martinez, M. Al-Khalil, H. Al-Enazi, Y. Cheng, J. Wright, C. Busso, and J. Jo, "Machine-learning assisted discrimination of precancerous and cancerous from healthy oral tissue based on multispectral autofluorescence lifetime imaging endoscopy," *Cancers*, vol. 13, no. 19, pp. 1–16, September 2021.

[9] K. Caughlin, E. Duran-Sierra, S. Cheng, R. Cuenca, B. Ahmed, J. Ji, V. Yakovlev, M. Martinez, M. Al-Khalil, H. Al-Enazi, J. A. Jo, and C. Busso, "End-to-end neural network for feature extraction and cancer diagnosis of in vivo fluorescence lifetime images of oral lesions," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2021)*, Guadalajara, Mexico, November 2021, pp. 3894–3897.

[10] K. Caughlin, E. Duran-Sierra, S. Cheng, R. Cuenca, B. Ahmed, J. Ji, M. Martinez, M. Al-Khalil, H. Al-Enazi, Y.-S. L. Cheng, J. Wright, J. Jo, and C. Busso, "Aligning small datasets using domain adversarial learning: Applications in automated in vivo oral cancer diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, pp. 457–468, January 2023.

[11] K. Caughlin, R. Cuenca Martinez, G. P. Tortorelli, Y.-S. L. Cheng, R. Hegde, C. Abraham, J. M. Plemons, Y. S. Wang, V. Woo, J. A. Jo, and C. Busso, "Beyond dysplasia: Uncovering structure in oral potentially malignant diseases with unsupervised contrastive learning," in *IEEE Engineering in Medicine and Biology Society (EMBS 2024)*, Orlando, FL, USA, July 2024.

[12] M. Marsden, B. Weyers, J. Bec, T. Sun, R. Gandour-Edwards, A. Birkeland, M. Abouyared, A. Bewley, D. Farwell, and L. Marcu, "Intraoperative margin assessment in oral and oropharyngeal cancer using label-free fluorescence lifetime imaging and machine learning," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 3, pp. 857–868, March 2021.

[13] L. Thompson and J. Bishop, *Head and neck pathology E-book: a volume in the series: foundations in diagnostic pathology*. Elsevier, December 2017.

[14] M. Serafino, B. Applegate, and J. Jo, "Direct frequency domain fluorescence lifetime imaging using field programmable gate arrays for real time processing," *Review of Scientific Instruments*, vol. 91, no. 3, p. 033708, March 2020.

[15] R. Cuenca, M. J. Serafino, G. P. Tortorelli, R. C. Faram, K. Higgins, S. Khajotia, Y.-S. L. Cheng, J. M. Plemons, V. Woo, C. Busso, K. Caughlin, and J. Jo, "Dual-excitation multispectral autofluorescence lifetime endoscopy for clinical label-free metabolic imaging of oral lesions," in *Imaging, Therapeutics, and Advanced Technology in Head and Neck Surgery and Otolaryngology 2023*, vol. SPIE PC12354, San Francisco, CA, USA, January-February 2023, pp. 12 354–8.

[16] P. Vasanthakumari, R. Romano, R. Rosa, A. Salvio, V. Yakovlev, C. Kurachi, and J. Jo, "Classification of skin-cancer lesions based on fluorescence lifetime imaging," in *SPIE Medical Imaging, 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317, Houston, TX, USA, February 2020.

[17] I. Pavlova, M. Williams, A. El-Naggar, R. Richards-Kortum, and A. Gillenwater, "Understanding the biological basis of autofluorescence imaging for oral cancer detection: high-resolution fluorescence microscopy in viable tissue," *Clinical Cancer Research*, vol. 14, no. 8, pp. 2396–2404, April 2008.

[18] C. Gulledge and M. Dewhirst, "Tumor oxygenation: a matter of supply and demand," *Anticancer research*, vol. 16, no. 2, pp. 741–749, March-April 1996.

[19] R. Drezek, C. Brookner, I. Pavlova, I. Boiko, A. Malpica, R. Lotan, M. Follen, and R. Richards-Kortum, "Autofluorescence microscopy of fresh cervical-tissue sections reveals alterations in tissue biochemistry with dysplasia," *Photochemistry and photobiology*, vol. 73, no. 6, pp. 636–641, June 2001.

[20] N. Ramanujam, R. Richards-Kortum, S. Thomsen, A. Mahadevan-Jansen, M. Follen, and B. Chance, "Low temperature fluorescence imaging of freeze-trapped human cervical tissues," *Optics Express*, vol. 8, no. 6, pp. 335–343, March 2001.

[21] B. Chance, B. Schoener, R. Oshino, and F. I. abd Y. Nakase, "Oxidation-reduction ratio studies of mitochondria in freeze-trapped samples. NADH and flavoprotein fluorescence signals," *Journal of Biological Chemistry*, vol. 254, no. 11, pp. 4764–4771, June 1979.

[22] Z. Zhang, D. Blessington, H. Li, T. Busch, J. Glickson, Q. Luo, B. Chance, and G. Zheng, "Redox ratio of mitochondria as an indicator for the response of photodynamic therapy," *Journal of Biomedical Optics*, vol. 9, no. 4, pp. 772–778, July/August 2004.

[23] M. Müller, T. Valdez, I. Georgakoudi, V. Backman, C. Fuentes, S. Kabani, N. Laver, Z. Wang, C. Boone, R. Dasar, S. Shapshay, and M. Feld, "Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 97, no. 7, pp. 1681–1692, April 2003.

[24] A. Shah, M. Beckler, A. Walsh, W. Jones, P. Pohlmann, and M. Skala, "Optical metabolic imaging of treatment response in human head and neck squamous cell carcinoma," *Plos One*, vol. 9, no. 3, p. e90746, March 2014.

[25] J. Jo, B. . Applegate, J. Park, S. Shrestha, P. Pande, I. Gimenez-Conti, and J. Brandon, "In vivo simultaneous morphological and biochemical optical imaging of oral epithelial cancer," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 10, pp. 2596–2599, October 2010.