
Proceedings Paper:
Lin, Z., Matella, M., Furmidge, R. et al. (2026) Machine learning-based electrical impedance spectroscopy classification for oral cancer. In: 2025 International Workshop on Impedance Spectroscopy (IWIS). 18th International Workshop on Impedance Spectroscopy (IWIS 2025), 23-26 Sep 2025, Chemnitz, Germany. Institute of Electrical and Electronics Engineers, pp. 2-6. ISBN: 979833159323-0.

https://doi.org/10.1109/iwis69004.2025.11339390

Machine Learning-Based Electrical Impedance Spectroscopy Classification for Oral Cancer

Zhicheng Lin
School of Electrical and Electronic Engineering
University of Sheffield
Sheffield, UK
zhicheng.m.lin@sheffield.ac.uk

Malwina Matella
School of Computer Science
University of Sheffield
Sheffield, UK
m.matella@sheffield.ac.uk

Rachel Furmidge
School of Clinical Dentistry
University of Sheffield
Sheffield, UK
rachel.furmidge@sheffield.ac.uk

Dawn Walker
School of Computer Science
University of Sheffield
Sheffield, UK
d.c.walker@sheffield.ac.uk

Helen E Colley
School of Clinical Dentistry
University of Sheffield
Sheffield, UK
h.colley@sheffield.ac.uk

Craig Murdoch
School of Clinical Dentistry
University of Sheffield
Sheffield, UK
c.murdoch@sheffield.ac.uk

Zi-Qiang Lang*
School of Electrical and Electronic Engineering
University of Sheffield
Sheffield, UK
z.lang@sheffield.ac.uk

***Abstract*—Head and neck cancers, particularly oral squamous cell carcinoma (OSCC), are major global health concern. An early detection can significantly improve the survival rate. Traditional diagnostic methods, such as biopsy, are invasive and analysis subjective, posing challenges for routine screening. This study explores the application of electrical impedance spectroscopy (EIS) combined with machine learning (ML) to classify oral cancer and healthy tissues non-invasively. By utilizing tissue-engineered oral mucosal model-based finite element modeling (FE) and simulations, we generate synthetic EIS data to train ML classifiers and apply the classifiers to real tissue data to classify the tissue status, Among the ML models tested, Random Forest (RF) showed the best performance, achieving an AUC of 0.95 and good sensitivity and specificity. This approach overcomes data scarcity issues by leveraging synthetic data and provides a promising pathway toward reliable, non-invasive diagnostic tools for oral cancer. Further research will focus on enhancing model accuracy by integrating clinical prior knowledge and refining tissue engineering techniques.**

Keywords— oral cancer diagnosis, electrical impedance spectroscopy, machine learning, tissue engineering, finite element modelling

I. INTRODUCTION

Head and neck cancer is the seventh most common type of cancer worldwide, with oral squamous cell carcinoma (OSCC) being one of its most prevalent subtypes [1]. In 2020, more than 377,000 new cases of OSCC in 185 countries were reported, resulting in nearly 178,000 deaths [1]. Despite advances in treatment, the overall 5-year survival rate remains low at just 54%, largely due to late-stage diagnosis. In contrast, early detection can significantly improve survival rates—up to 90% in some cases [2]. At present, diagnosis typically relies on clinical examination followed by an invasive biopsy. This process is not only subjective, showing considerable variability between pathologists [3], but also ill-suited for routine screening or frequent monitoring. While machine learning (ML) methods have helped standardize histological analysis [4], these tools still depend on biopsy-derived tissue samples.

Electrical impedance spectroscopy (EIS) offers a non-invasive alternative for the assessment of the structural and compositional characteristics of biological tissues [5]. Previous studies have shown that EIS spectra can differentiate between healthy, potentially malignant, and cancerous tissues in the cervix [6] . Although EIS has found clinical use in cervical cancer detection [7], exploration of its application in oral diagnostics is limited [8], [9], [10]. Moreover, the epithelium of the cervix and oral cavity differ, meaning that EIS for oral applications requires a comprehensive further study. Nonetheless, our recent study has shown that ML has potential in extracting features from EIS data, improving its utility in distinguishing OSCC from oral potentially malignant disorders (OPMD).

A key barrier to deploying EIS-based ML models in clinical practice is the limited availability of high-quality, labelled human data. Oral tissue samples are not only scarce and difficult to obtain but also highly variable across different oral sites and between individuals. To address this limitation, we investigate the feasibility of applying ML to classify EIS data obtained from *in vitro* tissue-engineered (TE) healthy and diseased oral mucosa from which highly localized histology samples can be obtained. By combining tissue engineering with histology-informed finite element (FE) modeling, we aim to create a controlled framework for developing and validating ML models. This approach serves as a foundational step toward future applications in human diagnostics, where data availability and variability present greater challenges.

In this preliminary study, we first collected impedance spectra from tissue-engineered oral mucosal models, then developed FE models based on the histology images obtained from the TE constructs to generate synthetic EIS data for ML training.

By training ML-based classifiers on the FE simulated data and validating the classifiers on real TE measurements, we show that models trained on FE datasets can accurately classify tissue status. This provides a compelling proof-of-concept for non-invasive, ML-assisted diagnostics in contexts where clinical data is limited. Our approach effectively bridges the gap between data scarcity and model performance,

This study is funded by Yorkshire Cancer Research (YCRSPF/2023/172).

*Corresponding author

paving the way for further studies and potential future applications in patient diagnostics.

II. Materials and Methods

A. *Tissue-engineering and computational modelling of oral epithelium*

Three batches of TE constructs comprising 14 biological replicates of healthy oral mucosal models (consisting of FNB6 oral keratinocytes; Fig. 1) and two batches comprising eight OSCC models (comprised of H357 oral cancer cells) were seeded onto oral fibroblast-populated decellularized dermis (DED) scaffolds. The TE oral epithelium constructs were cultured for 10 days (H357) or 14 days (FNB6) at an air-to-liquid interface in standard cell culture medium.

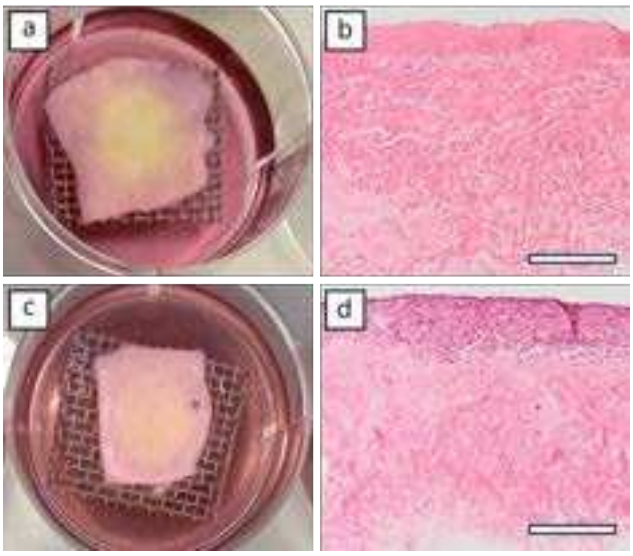

Fig. 1. Tissue-engineered models of oral mucosa. (a) Representative macroscopic image and (b) corresponding histological micrograph of tissue-engineered normal oral mucosa (TENOM). (c) Representative macroscopic image and (d) corresponding histological micrograph of a tissue-engineered squamous cell carcinoma (TESCC) model. Scale bar = 200 µm.

Fig. 2. The electrical impedance spectroscopy measuring probe (Zilico Ltd., UK).

On day 10, EIS measurements were conducted on each TE construct using a custom tetrapolar probe (Fig. 2). A small alternating current was applied to the tissue via a pair of driving electrodes, while the resulting potential difference was captured by a separate pair of passive electrodes. Impedance is then calculated using Ohm's law, as the ratio of voltage to current. The device records impedance at 14 frequencies, ranging from 76 Hz to 625 kHz. The EIS measurement of each TE construct was performed five times to ensure consistency and repeatability. The EIS at a particular frequency ω is a complex number represented by two components: the real part Z_{Re} and the imaginary part Z_{Im} as follows:

$$Z(\omega) = Z_{Re}(\omega) - jZ_{Im}(\omega) \quad (1)$$

The real part of impedance describes the resistive component of the tissue behavior, while the imaginary part is associated with the capacitive tissue properties. In studies on the electrical properties of tissues, the inductive properties are often neglected due to their small impact on the bulk electrical impedance when compared to resistive and conductive properties. Following the EIS measurements, the samples were fixed in formalin, paraffin wax-embedded and processed for histological analysis.

Virtual models of oral tissue constructs were developed by adapting multiscale FE modelling methodologies. The modelling framework includes two interconnected spatial scales. At the microscale (cellular level), transfer impedance simulations are used to derive effective electrical properties—namely, conductivity (σ) and relative permittivity (ε_r). These properties are then assigned to corresponding compartments in the macroscale (tissue-level) model of the oral epithelium. The macroscale level model additionally incorporates the tetrapolar electrode arrangement on its surface which is directly linked to the EIS device. This permits the comparison between the *in vitro* measured EIS data collected on the TE samples with the computed spectra.

This multiscale approach enables simulation of tissue-level impedance while avoiding the computational burden of explicitly modelling individual cells across the entire tissue volume. Cellular features, however, remain essential for capturing the passive electrical properties of biological tissue, particularly due to the dielectric behaviour of the cell plasma membrane. These membranes, approximately 8 nm thick, surround and isolate the cytoplasm and exhibit dielectric behaviour. In the kilohertz range, this leads to characteristic β-dispersion, manifested as a pronounced fall in the real component and a peak in the imaginary component of the impedance spectrum. All FE simulations were performed using the commercial software ANSYS Mechanical APDL, employing its quasi-static, time-harmonic electromagnetic solver.

Fig. 3 illustrates the geometries employed in both micro- and macroscale simulations. The microscale model consists of three compartments: cytoplasm, cell membrane, and extracellular space (ECS). Reflecting the layered architecture of the oral epithelium and the variation in cell morphology with tissue depth, three distinct epithelial regions were defined - basal, prickle (spinous) and superficial, with a dedicated microscale model constructed for each. Due to the flattened, cuboidal shape and organized orientation of epithelial cells, the resulting structures exhibit anisotropic electrical behaviours; that is, their conductivity and relative permittivity differ depending on the direction of current flow. Accordingly, transfer impedance simulations were conducted

in both the Z-direction (parallel to applied current) and the XY-plane (perpendicular to applied current).

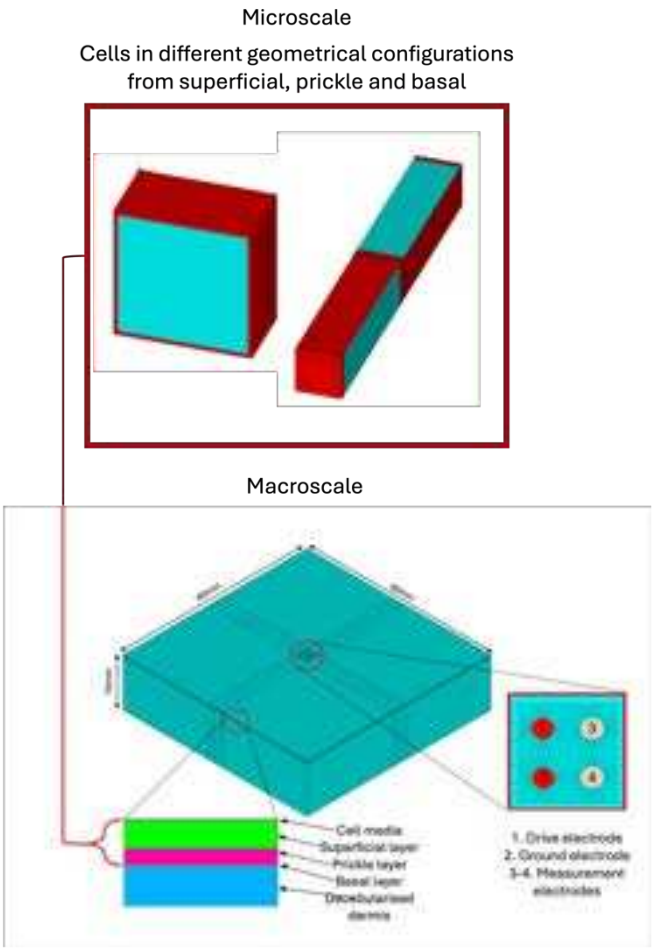

Fig. 3. Micro- and macroscale geometries of the multiscale finite element model.

At the macroscale, the TE model comprises a fibroblast-populated decellularized dermis scaffold, on top of which three epithelial layers reside; the basal, prickle (spinous) and superficial strata, followed by a thin cell media layer (mimicking the saliva in the real oral cavity environment). The mesh independence study was conducted for both cell and tissue scale models. The density of the mesh gradually increased until the difference between the numerical and analytical calculation for the transfer impedance was smaller than 1% after assigning uniform material properties across the whole domain.

The histology images of the TE samples provide the morphological information on cell sizes and thickness of the epithelium layers for each TE construct. This information is used as geometrical parameters assigned to the micro- and macroscale FE models. The electrical material properties of cytoplasm, ECS, cell membrane are obtained from the literature, while the DED and cell media properties were derived from the EIS measurements of these materials only.

The constructed multiscale FE model was utilized to compute 200 virtual EIS spectra of both healthy (FNB6, n=100) and cancer (H357, n=100) which were used to train the ML model. The ML model was then tested using 400 TE data samples (readings) from both healthy (200) and cancerous (200) tissue-engineered constructs. The FE data was computed using histology images from the first batch of healthy and cancerous oral epithelium. The EIS measurements of the remaining TE data were used in the ML model testing. Written informed consent was obtained from volunteers before the collection of cutaneous tissue to produce DED. This was ethically approved by the University of Sheffield Research Ethics Committee (REC reference 21/NE/0115). All measurements were conducted on tissue-engineered models that were chemically fixed immediately after data acquisition. The applied alternating current (<12 μA p–p) at 76 Hz to 625 kHz was low enough to avoid any electrochemical or thermal damage to the constructs during measurement.

B. Machine learning models

In this study, we employed four of the most widely used and well-established machine learning algorithms from the literature to develop our prediction and classification models. A brief overview of each algorithm is provided below.

a) Decision tree (DT) is a popular supervised machine learning algorithm used for both classification and regression tasks. It works by recursively splitting the dataset into subsets based on the value of input features, creating a tree-like structure of decision nodes and leaf nodes. Each internal node represents a test on a feature, each branch corresponds to an outcome of the test, and each leaf node represents a final decision or prediction. Decision trees are easy to understand, interpret, and visualize, making them especially useful for explaining model decisions.

b) Random Forest (RF) is an ensemble learning model that builds multiple decision trees using the bagging technique. Each tree is trained on a bootstrap sample of the data, and at each split, a random subset of features is considered to determine the best split. This randomness ensures that each tree is diverse, reducing the risk of overfitting. The final prediction is made by averaging the output (for regression) or taking a majority vote (for classification) from all trees, improving overall accuracy and robustness.

c) Support Vector Machine (SVM) is a generalized linear classifier for binary classification of data according to supervised learning, and its decision boundary is the maximum margin hyperplane for solving the learning samples. It works by finding a hyperplane that best separates data points into different classes while maximizing the margin between the classes.

d) K-Nearest Neighbors (KNN) is a simple, non-parametric, and instance-based learning algorithm used for classification and regression. It works by identifying the 'k' closest data points to a given input based on a distance metric, typically Euclidean distance. The prediction is then made based on the majority class (for classification) or average value (for regression) of these neighbors. KNN requires no model training, making it easy to implement, but its performance can be sensitive to the choice of 'k' and the scale of features.

All selected models were implemented in MATLAB (2022a). For each model, grid search was employed to systematically explore a range of hyperparameter combinations. The optimal parameters were selected based on sensitivity analysis, using the Youden index (Youden = sensitivity + specificity – 1) evaluated on the test set. The final hyperparameter settings for each model were as follows: DT: MinParentSize = 5, MinLeafSize = 5; RF: NumTrees = 1, MinLeafSize = 5; KNN: k = 1; SVM: Kernel = linear, BoxConstraint = 0.3).

C. Evaluation metrics

We calculated area under curve (AUC), sensitivity, specificity for test dataset based on the confusion matrix of each model and derived them as the following:

AUC: AUC assesses a model's ability to distinguish between positive and negative classes by quantifying its capacity to rank instances correctly. A higher AUC value indicates better class discrimination. AUC evaluates performance across all possible classification thresholds, making it particularly suitable for imbalanced datasets.

$$Specificity = TN/(TN + FP) \quad (2)$$

$$Sensitivity = TP/(TP + FN) \quad (3)$$

where TP: Number of true positive cases; TN: Number of true negative cases; FP: Number of false positive cases; FN: Number of false negative cases.

The difference between the performances was evaluated statistically using Wilcoxon signed-rank test with the significance level $\alpha = 0.05$.

III. RESULTS

Figure 4 presents the real part of the EIS signal obtained from both TE normal tissues (TENOM) and TE oral squamous cell carcinoma tissues (TESCC) across a range of frequencies. A clear distinction can be observed between the spectral profiles of TENOM and TESCC. For TENOM samples (left panel), the real part of the impedance decreases steadily with increasing frequency. The two experimental datasets (TE 241118 TENOM and TE 241220 TENOM) exhibit similar trends, and the finite element (FE) modeling data closely follow the experimental curves. This indicates that the FE model successfully replicates the real EIS behavior in normal tissues. In contrast, the TESCC data (right panel) show a markedly different spectral shape. Obvious difference can be observed that disease case with small amplitude and normal case with high amplitude. This contrast between TENOM and TESCC spectra suggests that EIS can effectively differentiate between normal and cancerous tissues based on their electrical properties.

In binary classification tasks, the output of a ML model is typically a probability value, with a threshold of 0.5 commonly used for decision-making which lacks strong explanation. However, in this study, a more reasonable threshold selection method was applied. The optimal threshold was determined by evaluating the Youden's index (J) across different threshold values in the ROC curve of the training set, and the threshold that provided the best overall performance on the training set was selected. This threshold was then applied to the test set to generate the final classification results.

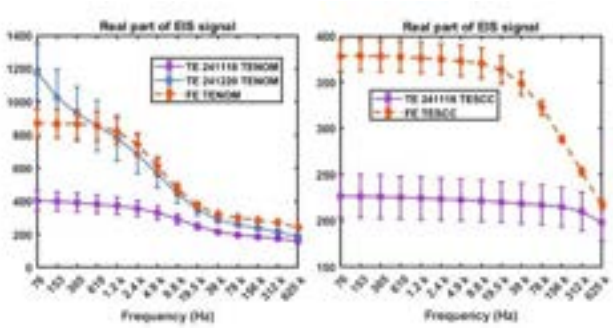

Fig. 4. Visualization of the EIS spectra of the collected data from TE/FE including normal (left) and cancer (right), showing the average values of all models and readings within the same batch. The left plot represents data from healthy samples (TE/FE), while the right plot represents data from cancerous samples (TE/FE).

TABLE I. MODEL PERFORMANCE ON TESTING DATASET.

Model	AUC	Specificity	Sensitivity
DT	0.913 ± 0.001	0.999 ± 0.001	0.825 ± 0.001
RF	0.864 ± 0.221	0.828 ± 0.325	0.890 ± 0.203
KNN	0.887 ± 0.001	0.840 ± 0.001	0.830 ± 0.001
SVM	0.857 ± 0.001	0.800 ± 0.001	0.795 ± 0.001

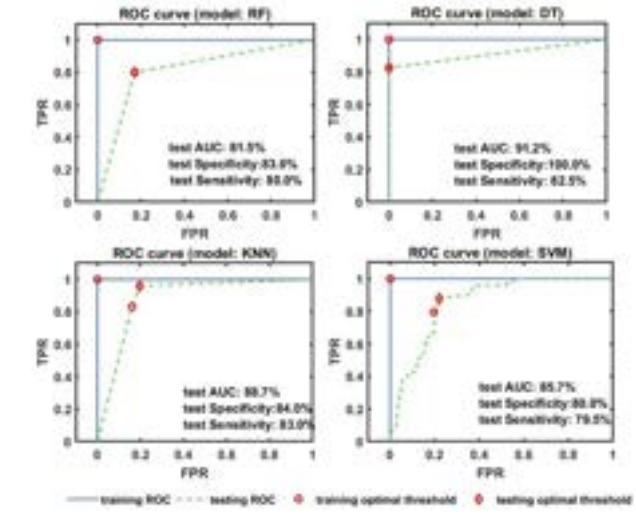

Fig. 5. Visualization of the ROC curve of each ML model on training and testing dataset. The red circles represent the optimal threshold points found on the training dataset using the Yuden index, and the red dimond represents the optimal threshold points on the testing dataset. The closer the two points are, the more consistent the model performs on the two datasets.

Among all models, the Decision Tree (DT) classifier achieved the best comprehensive performance on the test dataset (AUC: 0.91, Specificity: 0.99, Sensitivity: 0.83, Table I, $p<0.05$), indicating strong classification ability for both normal and cancerous samples. However, in terms of sensitivity alone, the Random Forest (RF) model achieved the best performance, (Sensitivity: 0.89, $p<0.05$). As shown in Fig. 5, the optimal threshold point on the training set ROC curve is very close to the optimal threshold point on the test set ROC curve. This suggests that the model generalizes well and does not exhibit significant overfitting or underfitting.

IV. CONCLUSION

This study presents a preliminary framework that integrates tissue engineering, multiscale finite element modeling, and machine learning to enable non-invasive classification of healthy and cancerous oral epithelial tissues using electrical impedance spectroscopy. By developing tissue-engineered constructs, finite element modelling, simulating their electrical behavior across micro- and macroscale geometries, we generated a robust dataset to train and validate ML classifiers in a controlled environment. Among the tested algorithms, the Random Forest model achieved the highest classification performance, demonstrating strong generalization to unseen data. This approach addresses key limitations in current EIS-based diagnostics, particularly the limitation of high-quality, annotated human samples. The results provide a compelling proof of concept for using synthetic data to train clinically relevant ML models, offering a promising step toward reliable, non-invasive screening tools for oral cancer and other epithelial disorders.

In future work, we plan to improve the existing TE and computational methodologies working towards reliable TE models and the FE model calibration and validation. Additionally, we also aim to explore the integration of prior knowledge to enhance ML model predictions. Purely data-driven predictions can be difficult for clinicians and patients to trust, especially when they lack interpretability. However, prior knowledge, whether embedded in a clinician's experience or derived from large-scale statistical patterns, is often difficult to formalize into mathematical equation. Incorporating such contextual information, such as a patient's smoking and alcohol consumption historical record in oral cancer diagnosis, may improve both the accuracy and credibility of ML-based systems. This represents a promising direction for our ongoing research.

ACKNOWLEDGMENT

This work was funded by a UK Engineering and Physical Sciences Research Council Knowledge Exchange Award to ZQL and CM. ZQL, CM and DCW are also recipients of a Yorkshire Cancer Research Sheffield Pioneer Award (YCRSPF/2023/172) for the translation of EIS-based machine learning for oral cancer diagnosis. The authors would like to acknowledge Zilico Limited for the use of the oral electrical impedance device in this study.

REFERENCES

[1] H. Sung *et al.*, 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] Muller P, Belot A, Morris M, and Rachet B, 'Oral Health Foundation: The State of Mouth Cancer UK Report 2020/21.', Cancer Research UK Cancer Survival Group, London School of Hygiene and Tropical Medicine, 2022. [Online]. Available: Available from http://csg.lshtm.ac.uk/rare-cancers/.Accessed July 2016. www.mouthcancer.org

[3] K.A. Shahul Hameed, K.A. Shaheer Abubacker, A. Banumathi, and G. Ulaganathan, 'Immunohistochemical analysis of oral cancer tissue images using support vector machine', *Measurement*, vol. 173, no. 108476, 2021, doi: https://doi.org/10.1016/j.measurement.2020.108476.

[4] S. Camalan *et al.*, 'Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results', *Cancers*, vol. 13, no. 6, p. 1291, Mar. 2021, doi: 10.3390/cancers13061291.

[5] M. Grossi and B. Riccò, 'Electrical impedance spectroscopy (EIS) for biological analysis and food characterization: a review', *J. Sens. Sens. Syst.*, vol. 6, no. 2, pp. 303–325, Aug. 2017, doi: 10.5194/jsss-6-303-2017.

[6] P. Li, P. E. Highfield, Z.-Q. Lang, and D. Kell, 'Cervical cancer prognosis and diagnosis using electrical impedance spectroscopy', *J. Electr. Bioimpedance*, vol. 12, no. 1, pp. 153–162, Dec. 2021, doi: 10.2478/joeb-2021-0018.

[7] J. A. Tidy and B. H. Brown, 'Increased detection of high grade CIN, when using electrical impedance spectroscopy as an adjunct to routine colposcopy, is maintained when used across international boundaries: Prospective data from nine European countries', *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 275, pp. 41–45, Aug. 2022, doi: 10.1016/j.ejogrb.2022.05.025.

[8] C. Murdoch *et al.*, 'Use of electrical impedance spectroscopy to detect malignant and potentially malignant oral lesions', *Int. J. Nanomedicine*, p. 4521, Sep. 2014, doi: 10.2147/IJN.S64087.

[9] Z. Lin *et al.*, 'Deep learning-based electrical impedance spectroscopy analysis for malignant and potentially malignant oral disorder detection', *Sci. Rep.*, vol. 15, no. 1, p. 19458, Jun. 2025, doi: 10.1038/s41598-025-05116-8.

[10] S. A. Lloyd *et al.*, 'In Vivo Classification of Oral Lesions Using Electrical Impedance Spectroscopy', *IEEE Trans. Biomed. Eng.*, pp. 1–10, 2025, doi: 10.1109/TBME.2025.3581465.