Diverse Region-Based CNN for Tongue Squamous Cell Carcinoma Classification With Raman Spectroscopy

HAO YAN\textsuperscript{1}, MINGXIN YU\textsuperscript{1}, JIABIN XIA\textsuperscript{1,2}, LIANQING ZHU\textsuperscript{1,2}, (Member, IEEE), TAO ZHANG\textsuperscript{3}, ZHIHUI ZHU\textsuperscript{3}, AND GUANGKAI SUN\textsuperscript{1}

\textsuperscript{1}Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing 100192, China
\textsuperscript{2}School of Instrument Science and Opto-electronics Engineering, Hefei University of Technology, Hefei 230009, China
\textsuperscript{3}Department of Stomatology, Peking Union Medical College Hospital, Beijing 100730, China

Corresponding authors: Lianqing Zhu (lqzhu_bistu@sina.com) and Tao Zhang (drtzhang@126.com)

This work was supported in part by the Key projects of the National Natural Science Foundation of China under Grant 51535002, in part by the Natural Science Foundation of Beijing—Program for Original Innovation Joint Foundation of Haidian District under Grant L182066, and in part by The Programme of Introducing Talents of Discipline to Universities under Grant D17021.

\section*{ABSTRACT} Border discrimination is very important in the treatment of tongue squamous cell carcinoma (TSCC). This study proposes an ensemble convolutional neural network (CNN) framework based on fiber optic Raman spectroscopy and deep learning techniques to distinguish between TSCC and non-tumor tissue frameworks. First, the data used in the experiments was collected by a fiber optic Raman system. A total of 44 tissues of 22 patients were collected for Raman spectroscopy, with TSCC and adjacent normal tissues each accounting for half. The spectral data range used in the model from a full spectrum of 600-4000 cm\textsuperscript{-1}. Then, the ensemble CNN model was used in the experiment. By using two convolution kernels, the model is able to extract nonlinear feature representations from different spectral regions. It has two advantages, on the one hand, it reduces the generation of noise, on the other hand, it obtains a stronger distinguishing ability. Finally, a feature vector is formed by the fusion layer, and is sent to the fully connected layer for the TSCC classification task. The results showed that the sensitivity and specificity of the model were 99.2\% and 99.2\%, respectively. In addition, comparison with existing methods shows that our method achieves the highest accuracy of TSCC classification. By comparing the different channels, the results show that the spectral range of 1380-2250 cm\textsuperscript{-1} data has the greatest impact on the results. Therefore, Raman spectroscopy combined with the ensemble CNN model has great potential and can provide a useful technique for intraoperative evaluation of the margins of oral tongue squamous cell carcinoma.

\section*{INDEX TERMS} Ensemble convolutional neural networks, fiber optic Raman system, Raman spectroscopy, tongue squamous cell carcinoma.

\section*{I. INTRODUCTION} In 2018, the World Health Organization (WHO) reported 18.1 million new cancer cases and 9.6 million cancer deaths. Among them, oral cancer accounted for 2.0\% and 1.9\%, respectively [1]. Tumor resection is the main treatment, but investigation of oral cavity squamous cell carcinoma (OCSCC) found that in about 43\% of cases, marginal resection is incomplete [2]. This leaves a hidden danger to tumor recurrence. The 5-year survival rate is reported to be less than 50\% in oral tumor surgery [2]–[4]. Intraoperative testing is often performed by the surgeon’s visual inspection and palpation, but up to 87\% of the tumor-positive edges are located in the deeper soft tissue layer. Moreover, the use of frozen section method for pathological examination of tissue takes too long, so it is impossible to effectively remove oral tumor cells during surgery by doctor’s experience and pathological examination [5]–[9]. Therefore, the ultimate goal of many current studies is to obtain a non-invasive detection method to support the surgeon’s assessment of the margin of
resection. Optical technology has the advantages of robustness, accuracy, low cost, portability and operability, and can be effectively used in clinical applications.

A variety of optical technologies have been used in clinical applications such as optical coherence tomography (OCT), magnetic resonance imaging (MRI), auto-fluorescence, and Raman spectroscopy. Although both can improve the extent of tumor resection [10], compared to Raman spectroscopy, structure-based imaging techniques such as OCT and MRI lack the necessary tissue chemistry and can cause disruptions during surgery [11], [12]. The auto-fluorescence method has a high degree of recognition of TSCC and a low recognition rate for normal tissues [13]–[15]. Fluorescent agents are also needed as well as a dark room to detect weak fluorescence [12]. Raman spectroscopy is an ideal method for detecting in vivo, which has the advantages of being fast, quantitative and non-destructive. It can also be combined with optic fiber equipment for fast, real-time, high-precision tissue evaluation [16]. In terms of detection, compared with OCT and MRI methods, Raman spectroscopy is a molecular-based optical detection method, which is more sensitive to the tumor tissue remaining in the edge of the tissue than the image. On the other hand, the detection and collection of Raman spectroscopy requires only a hand-held fiber-optic probe. Compared with the large-scale equipment required by OCT and MRI methods, it has the advantages of portability and real-time. Compared with fluorescence spectroscopy, it can be seen that Raman spectroscopy does not require a fluoroscope in the implementation process, and can also acquire collected information under the interference of ambient light.

Raman spectroscopy can acquire the chemical traits and structural features of tissues from spectral features such as peak positions, peaks, and bands [17]. Unlike the image, it does not express features in the shape of the tissue. Krishna et al. [17] and Bergholt et al. [18]. first applied Raman spectroscopy to oral in vivo experiments. The results indicate that spectra can be used to characterize healthy oral cavity in the high wavenumber (1800-3000 cm\(^{-1}\)) region or in the 800-1800 cm\(^{-1}\) range. Singh et al. first proposed the application of Raman spectroscopy in the identification of oral lesions [19]. In traditional medical diagnosis, pathological examination is mainly used, which includes morphology and molecular biology techniques. Tissue section during surgery can provide rich tissue information, but it is not real-time, and it is not effective for judging the residual depth and range of tumor cells. These are the advantages of Raman spectroscopy. Raman spectroscopy has important advantages in the identification of oral tumor tissue: (1) The cancerous tissue cells change in the number of structures and constituent substances, and the peaks corresponding to proteins and nucleic acid substances change [20]. (2) It can be used to guide marginal resection during surgery. (3) Fiber optic Raman devices can reduce sampling problems and reduce evaluation time during surgery. (4) It has high specificity for tumor tissue recognition. (5) It can reduce the need for adjuvant therapy [21].

Raman spectroscopy has been extensively studied in cancer analysis. Although oral cancer is only a small part of the cancer category, there is a lot of research results. Krishna et al. found in oral experiments that Raman spectroscopy showed spectral sensitivity to molecular composition and morphological changes associated with malignant tissue. Oral squamous cell carcinoma (OSCC), oral leukoplakia (OLK), and oral sub-mucous fibrosis (OSMF) could be distinguished from normal tissues by a multivariate statistical algorithm [16]. Carvalho et al. studied the application of high-wavenumber (2800-3600 cm\(^{-1}\)) region of Raman spectroscopy in oral cancer, and identified the oral cells by principal component analysis (PCA) and feature discriminant analysis [22]. Krishna et al. studied ex vivo oral tissue and used Raman spectroscopy combined with PCA to compare protein changes in oral squamous cell carcinoma and normal tissues [17]. Venkata Krishna et al. used a fiber optic probe as a Raman spectroscopy collection tool, and also processed the spectrum using PCA, using Mahalanobis distance and residual squared sum as a standard to classify normal tissue and oral cancer tissue [23]. Malini et al. studied the potential of Raman spectroscopy in oral tumor recognition, using PCA and multi-parametric limit tests to identify normal, inflammatory, precancerous, and malignant tissues. The results of the study on the spectrum confirmed that the normal tissue mainly characterized the spectral characteristics of the lipid, and the diseased tissue showed the spectral characteristics of the protein [23]. Guze et al. used Raman spectroscopy to analyze the spectra collected from different oral sites. A linear discriminant analysis (LDA)-based algorithm was used to demonstrate the possibility of Raman spectroscopy for oral mucosal diseases and specificity for specific mucosal types of the oral cavity [17]. In the above research work, these traditional machine learning methods were used for the processing of Raman spectral data. To the best of our knowledge, there was very little work to extract spectral information from the fiber optic Raman system for OSCC classification.

Deep learning does not require manual intervention to extract data features compared to traditional machine learning methods. Therefore, deep learning is used in many medical tests, such as electrocardiogram (ECG) detection [42], [45]–[48], lung segmentation on x-ray images [43], and identification of bacterial strains [44]. In addition, it can extract more feature information from the tumor tissue Raman spectrum and can be more useful in the process. Therefore, the accuracy of tumor tissue recognition is improved. In our previous article [39], [40], the data we collected was only from 12 people in total, with a total of 24 tissues, and the spectral range was only 600-1800 cm\(^{-1}\). In this experiment, 22 patients were obtained, a total of 44 organizations, a significant increase compared to the previous data size. In this article, the study used a full spectrum (600-4000 cm\(^{-1}\)) in order to achieve better results with sufficient features. In the last experiment, the traditional CNN model was adopted. Because the data size is small, the depth
TABLE 1. Descriptive characteristics: patient cohort.

| Gender | Total |
|--------|-------|
|        | Male  | Female |       |
| n      | 13    | 9      | 22    |
| Age (years) mean | 74.6  | 71.7   | 69.1  |
| min    | 51    | 51     | 51    |
| max    | 85    | 84     | 85    |

is shallow. Although some results have been achieved, there still can be improvements.

In summary, the main contributions of this paper are as follows: (1) A new classification ensemble model based on convolutional neural network for normal tissue and tongue squamous cell carcinoma (TSCC) is proposed. (2) In the convolution process, two kinds of size convolution kernels are used. Large convolution kernel can suppress noise better, and small convolution kernel can extract more detailed features. (3) The use of joint representations of different regions in the proposed framework can be performed simultaneously. This model can be used as a measure to determine the marginal resection of oral cancer in the future.

II. MATERIALS AND METHODS

A. SAMPLE PREPARATION

The spectral acquisition experimental procedure in this study was performed at Peking Union Medical College Hospital (PUMC), and the informed consent of PUMC and patients was obtained. All tissue samples were obtained from the PUMC stomatology department, from surgically resected patient’s tumor tissue and the safety margin maintained was of the surrounding 1 cm physiological mouth tongue. The choice of the patient, regardless of gender and age, is determined by the doctor based on the patient’s condition.

In the Department of Stomatology at Peking Union Medical College Hospital, a total of 44 samples were obtained from 22 patients, including 22 cancerous tissues and 22 adjacent tissues, respectively from different parts of the tongue of the same patient. There were 13 men and 9 women. The average age of the patients was 69.1 years. The oldest patient is 85 years old and the youngest patient is 51 years old. (Table 1).

The study involved 22 patients, 44 tissues, half of which were tongue squamous cell carcinoma tissue and the other half were normal tissues surrounding cancerous tissue. The type of tissue obtained is subject to histopathological calibration by a medical professional. Fresh samples of tongue squamous cell carcinoma were subjected to ex vivo experiments. After excision of the samples, the temperature was taken to 25 °C (room temperature) within 3 hours. We have not dealt with the organization to maintain its original appearance. The goal of this process is to collect spectral samples from TSCC and non-tumor tissue and to build a data set for the model.

B. RAMAN SPECTROSCOPY

This study was carried out on an independently constructed fiber optic Raman spectroscopy system. Fiber optic Raman spectroscopy was recorded on a spectrometer (QE65PRO-RAMAN, Ocean Optics, USA), which provided a spectral resolution of 6 cm⁻¹. Software recording of the spectra was performed using high quantum efficiency (up to 90%) and low standard CCD (S7031-1006, Hamamatsu, Japan). The system uses a 785 nm diode laser (maximum output: 300 mW, FC-D-785 nm, China Changchun New Industrial Optoelectronics Technology Co., Ltd.) as the excitation source. The laser power in the experiment was 50 mW, the spatial resolution was 100 µm, and the spectrum was in the range of 600-4000 cm⁻¹. The excitation light was transmitted to the target tissue using a fiber optic probe (diameter: 1.0 mm, centered on an optical transmission fiber, around six acquisition fibers, core diameter 100 µm, numerical aperture, NA = 0.22). The distal end of the collection fiber has filters for suppressing Rayleigh scattering, fluorescence and silica Raman signals while maximizing signal collection. For the measurement of the sample, the integration time was 20 seconds, and three different positions of each sample were selected.

In this system, the laser emits laser light and transmits it to the target tissue through the center fiber of the fiber probe. Then, the light carrying the Raman spectrum information is returned and collected through six collection fibers around the center of the fiber probe. Passed through the filter to the CCD. Finally, the spectral data is stored in a computer.

The complete system includes a Raman spectrum collection part and a model part for spectral classification. At this stage, the experiment has not yet undergone clinical trials. At present, it is necessary to perfect the model and then deploy the model to a computer for simulation verification.

C. FRAMEWORK OF DIVERSE REGIONS-BASED CNN

Figure 1 illustrates the framework of our proposed diverse region-based CNN tongue squamous cell carcinoma classification method. This method consists of two stages, i.e., training stage and testing stage. In each stage, it contains two successive modules: 1) preprocessing module and 2) ensemble CNN module. The purpose of the preprocessing module is to remove noise signals. The ensemble CNN is taken as a TSCC classifier in this framework.

To be specific, in the training stage, the inputs for ensemble CNN include a feature matrix X and a class label vector Y. Each row of the feature matrix is composed of the preprocessed training TSCC data. The class label is represented to a TSCC or normal tissue corresponding to a feature vector. Then, X and Y are fed to the ensemble CNN module to train the diverse region-based CNN model to classify TSCC tissue. In the testing stage, the same pre-treatments of TSCC data are carried out for noise signals. Then, the preprocessed data...
pass through the trained ensemble CNN to output TSCC class labels.

D. RAMAN DATA PREPROCESSING

First, in order to fully maintain the integrity of the information contained in the Raman spectrum, and for statistical analysis, we considered the full spectrum information of the 600~4000 cm⁻¹ region. This contains important information about tumor tissue and normal tissue, such as information about its molecular structure and conformation. Then, because the acquired Raman spectrum contains fluorescent backgrounds of aromatic chromophores and other noises, the Savitzky-Golay [24] method is used to smooth the spectrum in order to reduce spectral noise. Fluorescent background noise is removed using asymmetric weighted penalty least squares [25], [26]. Finally, the data is normalized.

In this experiment, the data set we prepared contained a total of 2004 spectral data, with 1172 spectra of normal tissue and 832 spectra of tumor tissue. The data set is divided into random sampling methods, 80% is the training set, and 20% is divided into the test set. The training set contains 1604 spectra, of which 948 are normal tissues and 656 are tumor tissues. The test set contains 400 spectra, of which 224 are normal tissues and 176 are tumor tissues.

The model of this experiment is mainly based on Convolutional Neural Network (CNN). One of the advantages of CNN is to obtain high-performance models through a large amount of data, and to prevent over-fitting during training. Moreover, by increasing the training set, it solves the problem of insufficient data, improves the accuracy, and enhances the generalization ability and robustness. So we increase the amount of data through data augmentation. The offset, slope and multiplication are mainly used here to randomly simulate the baseline offset, slope difference and intensity difference. This method does not change the expression characteristics of the spectrum, but it is different from the original data. In this way, the training data reached more than 10,000.

E. ENSEMBLE CONVOLUTIONAL NEURAL NETWORKS

The model in this study mainly uses Convolutional Neural Networks (CNN), which are widely used in image classification tasks. In addition, the idea of an ensemble method is also adopted. The idea means that when there are differences of extracted features between the models, the prediction results tend to use better results [27]–[29]. The Raman spectral signal belongs to a one-dimensional signal and the model's characteristics are extracted by a one-dimensional convolution kernel.

As shown in Figure 2, the set model in this study includes an input layer, a convolution layer, a pooling layer, a global average pooling layer, a fully connected layer, and an output layer. Raman spectral data is used as input data in the form of \( N \times 1 \) and the entire input layer contains the entire spectrum (full sampled intensity at the same interval). By studying the bandwidth information of the feature, an appropriate one-dimensional convolution kernel is selected to achieve a good effect on feature extraction. First, for a complete spectrum, the convolution kernel is \( 7 \times 1 \) in order to reduce noise interference. In the main channel, the convolution kernel size is \( 1 \times 5 \), and the larger convolution kernel introduces less noise than the smaller convolution kernel. The process of convolution is a process of weighted averaging by convolution kernels. For noises with a mean of zero, when the size of the convolution kernel is \( n \), it is weakened to \( 1/n \). Therefore, a larger convolution kernel can suppress noise with a mean of zero. In the secondary channel, the convolution kernel size is \( 1 \times 3 \), and the smaller convolution kernel can have more detail than the large convolution kernel [41].

Feature extraction is done by the convolutional layer and the pooling layer. Among them, each channel has 5 layers of convolution and 6 layers of pooled layers. Convolutional layer weight sharing saves computational complexity and stability. The spectral characteristic information of the extracted frequency band is then sent to the fully connected layer through global average pooling. In the convolutional layer, Parametric Rectified Linear Unit (PReLU) [30] is used as the activation function. Effectively eliminates the problem of gradient elimination, which is defined as:

\[
f(x_i) = \begin{cases} 
  x_i, & \text{if } x_i > 0 \\
  a_i x_i, & \text{if } x_i \leq 0
\end{cases}
\]

where \( x_i \) is any input on the i-th channel, \( a_i \) is the negative slope, and its value is updated with the extracted features changes.

The convolutional layer is followed by a pooling layer. The features obtained by the convolutional layer form new features through pooling techniques. Three goals were achieved: (1) removing redundant information, (2) reducing parameters, and (3) preventing overfitting. The average pooling is used in the main channel to reduce the problem of increased variance of the estimated values and to retain more overall information about the data. Maximum pooling is used in the secondary channel to reduce the offset of the estimated mean.
caused in the convolutional layer and to preserve more texture information.

Batch normalization is used before the first pooling layer of each channel. It has three advantages: (1) It speeds up the training and speeds up the convergence process. (2) It increases the classification effect to prevent over-fitting. (3) It simplifies the tuning process, reduces the initialization requirements, and uses a large learning rate.

Prior to the output classification, the global average pooling (GAP) was used instead of the leveling layer, while reducing the number of parameters and avoiding overfitting. The features extracted from the five channels are then merged into a new one-dimensional array through the fusion layer. Finally, through the output layer as a classifier, the extracted features of the Raman spectral data are used to identify tumor tissue and normal tissue. The accuracy and loss curves during the model training process are shown in Figure 3. The convergence speed is fast and stable.

III. RESULTS

A. SPECTRAL DATA ANALYSIS

The Raman spectrum of the sample was obtained by a fiber optic Raman spectroscopy system, and the average spectra of TSCC and non-tumor tissues with wave numbers between 600 and 2100 cm$^{-1}$ were shown in Figures 4A and 4B, respectively. High wavenumber (2500-4000 cm$^{-1}$) spectra of TSCC tissue and normal tissue are shown in Figures 5(A) and 5(B), respectively. From Raman spectroscopy, it can be observed that there is a significant difference in peak, and band between tumor tissue and normal tissue. The reasons for these differences may be related to the composition and growth characteristics of the tissue.
FIGURE 3. The Loss and accuracy curve of model. A: accuracy curve; B: loss curve.

content of some substances, such as proteins, DNA, lipids and water [17],[23],[31]-[36].

In the range of 600-2100 cm\(^{-1}\), we can see the peak at 1770 cm\(^{-1}\) in Figure 4A, which is due to the existence of lamina propria and submucosa in normal tissue, the main components of this kind of tissue are reticular protein and collagen fiber and adipose tissue. In tumor tissue, lipids are not easily observed because of the loss of different layers of structure. On the contrary, as it contains more surface proteins, including receptor proteins, enzymes, antigens and antibodies, its Raman spectra more show the spectrum of proteins, such as the peak at 1650 cm\(^{-1}\) is expressed as amide I. In the spectral comparison, the difference position mainly appears in the range of 1200-1700 cm\(^{-1}\), where hemoglobin is mainly manifested, which means there is blood supply in the oral cavity. In normal tissues, the peak at 1299 and 1376 cm\(^{-1}\) was more obvious, which was mainly due to the action of lipids in the tissues [37]. In tumor tissue, the content of nucleic acid is higher, so the peak at 798 cm\(^{-1}\) is higher. The peak at 1236 and 1708 cm\(^{-1}\) is caused by the vibration of thymine amide III, while the peak at 1452 cm\(^{-1}\) is caused by the deformation of CH\(_3\)/CH\(_2\). There are also some different peaks and bands in the spectrum, such as tyrosine at 993 cm\(^{-1}\), phenylalanine at 1605 cm\(^{-1}\), and \(\beta\)-carotene at 1510 cm\(^{-1}\). In addition, many weak bands in the spectrum are produced because of the existence of a large number of immunoglobulin in the cells, such as 955, 1123, 1240, 1400 cm\(^{-1}\) and so on [38].

Figures. 5A and 5B show Raman spectra in the range of 2500-4000 cm\(^{-1}\). It is mainly characterized by the characteristic peak of lipid and protein in the range
of 2800-3000 cm$^{-1}$, and is characterized by residual environmental water, other-OH group, N-H expansion vibration and Amide A-band protein in the range of about 3100-3400 cm$^{-1}$ [35], [36]. The band at 2851 cm$^{-1}$ showed an increase in protein content in TSCC cells, which was related to the extension of Lipidic components, fatty acids and CH2 in phospholipids. However, 3180, 3210, 3285 and 3345 cm$^{-1}$ showed the characteristics of Amide A-water and bound water, which could be used to identify the characteristic peaks of normal tissues. Related to tumor tissue, the peak was at 3470 and 3550 cm$^{-1}$, which was related to unbound water.

Figure 6 shows the comparison of the feature map of the tumor tissue and normal tissue after the fifth layer of convolution in the global channel. From the figure, it can be seen that the difference points considered in the model are distributed within the range of the difference points of the two tissues. Because the difference in the spectrum is a range, the marked point is a significant peak in this range of
differences. So the difference between the two organizations at the molecular level is also used in program identification.

B. CLASSIFICATION RESULTS
In this paper, a total of 1604 spectral data points are obtained for training through the fiber optic Raman spectroscopy system, and each data can be regarded as independent data. The experiment used a 5-fold crossover model to assess the generalization ability of the model results. And to assess model performance, this paper used a confusion matrix and receiver operating characteristic curve (ROC). The ROC curve is a graphical method for displaying the true rate (TPR) and false positive rate (FPR) of a classifier.

Table 5 shows the confusion matrix for classifying the TSCCs to show the performance of the model. The visual effects of the test results for 176 TSCC tissues and 224 normal
tissues are shown in the table. The two types of organizations are balanced in quantity and can accurately represent model performance. Among the 176 TSCC tissues, 3 were misjudged as normal tissues, and 2 out of 224 normal tissues were misjudged as TSCC tissues. This means that in the existing ensemble CNN model, most TSCC organizations and normal organizations can be distinguished. In order to evaluate the model in more aspects, the accuracy, sensitivity and specificity are calculated using the statistical information obtained from the confusion matrix. Specificity refers to the specificity of the model to TSCC and indicates the probability of not being missed. The result is 99.10%. Conversely, sensitivity is a probability of not being misdiagnosed, which is 98.29%. The evaluation of the classification results is shown in Figure 7.

The sensitivity calculation method is as follows:

\[ \text{Sensitivity} = \frac{TP}{TP + FN} \]

| TABLE 4. Raman peaks assignment for the observed bands. |
|--------------------------------------------------------|
| Raman shift (cm\(^{-1}\)) | Peak assignment          |
|-----------------------------|--------------------------|
| 798                         | Phosphoric acid, nucleic acid |
| 993                         | Phenylalanine             |
| 1236                        | Amide III                |
| 1299                        | Fatty acid                |
| 1376                        | Phospholipid (CH\(_3\))  |
| 1452                        | C-H / CH\(_2\)           |
| 1510                        | \(\beta\)-carotene        |
| 1605                        | Tyrosine, Tryptophan      |
| 1650                        | Amide I, C-C stretch, keratin |
| 1708                        | Thymine                  |
| 1770                        | Cholesteryl linolate      |
| 2851                        | CH\(_2\) symmetric stretch of lipids |
| 2870                        | CH3 asymmetric stretch of lipids and proteins |
| 2890                        | CH\(_2\) deformation      |
| 2936                        | CH2 asymmetric stretch of proteins, |
| 3015                        | \(-\text{CH}, lipids, fatty acids, |
| 3180                        | OH bound                 |
| 3210                        | OH (tightly bound)       |
| 3285                        | NH2 asymmetric stretch (Amide A) |
| 3345                        | OH (partially bound)     |
| 3470                        | OH symmetric stretch (unbound) |
| 3550                        | OH stretching vibration  |

| TABLE 5. Confusion matrix for Raman spectra. |
|---------------------------------------------|
| Predicted                                   |
| Actual | TSCC tissue | Normal tissue |
|--------|-------------|---------------|
| TSCC tissue | 173          | 3             |
| Normal tissue | 2           | 222       |
TABLE 6. Classification performance of identifying tumor with different models.

| Metrics           | Ensemble CNN | CNN (New) | CNN (Past) |
|-------------------|--------------|-----------|------------|
| True positive     | 222          | 219       | 190        |
| False negative    | 2            | 5         | 34         |
| False positive    | 3            | 6         | 42         |
| True negative     | 173          | 170       | 134        |
| Sensitivity       | 99.10%       | 97.76%    | 84.82%     |
| Specificity       | 98.29%       | 86.59%    | 76.14%     |
| Precision         | 98.67%       | 97.33%    | 81.90%     |
| Accuracy          | 98.75%       | 97.25%    | 81.00%     |

The specificity calculation method is as follows:

\[ \text{Specificity} = \frac{TN}{TN + FP} \]

Among them, TP indicates that the TSCC organization is identified as TSCC organization as 173; TN indicates that the normal tissue is identified as normal organization as 222; FP indicates that the normal organization is incorrectly identified as TSCC organization as 3; FN indicates that the TSCC organization is incorrectly identified as normal tissue, it is 2.

Figure 8 shows the ROC’s performance evaluation of the model. ROC analysis is used to select the best classification model, abandon the suboptimal model, and set the optimal threshold in the same model. The area under the curve (AUC) is the area covered by the ROC curve. The larger the AUC, the better the classifier classification effect. In this model, the ROC average AUC is 0.97 ± 0.02.

C. SPECTRAL DATA ANALYSIS

In order to verify whether the proposed method is superior, a simple CNN model is selected and compared with our method under the same number of layers and parameters. In addition, the CNN model used in the previous article is also used, which has only five layers of neural networks to compare the model capabilities in the full spectrum [39]. Figure 9 shows the accuracy and loss curves of this model for this data set. The data used for testing included 224 normal tissues and 176 TSCC tissues. Both the ensemble CNN and CNN models use our spectral database. The performance indicators (sensitivity, specificity, and accuracy) of the two methods are shown in Table 6. In Table 6, new CNN refers to only retaining the full spectrum channel in the ensemble CNN, and past CNN refers to the CNN model in the previous paper [39]. The results of the CNN model show that in the case of normal tissue, 219 of the 224 spectra are correctly...
classified. In the case of proper tumor tissue, 170 of the 176 spectra were correctly classified. The accuracy of CNN is 97.25%, its sensitivity is 97.76%, and its specificity is 96.59%. The CNN model used in the previous paper is not applicable to the full-spectrum model with an accuracy of 62.4%, a specificity of 88.3%, and a sensitivity of 36.3%. For the ensemble CNN model used in this article, it achieved the highest accuracy (98.75%), sensitivity (99.10%) and specificity (98.29%).

In order to verify which features in the full spectrum make the classification of the model more efficient, one of the four secondary channels is selected to be combined.
with the primary channel, using the same spectral data set and the same parameter configuration for comparison. Figure 10 shows the performance specifications (accuracy, sensitivity, and specificity) for these four channels. In comparison, the spectral band range (about 1380-2250 cm\(^{-1}\)) contained in channel 2 has the best effect in the identification of spectral data. Channel 2 has the highest accuracy rate of 96.5%, and the highest sensitivity is 94.9%. Channel 3 has the highest specificity of 98.9%. In addition, the second good channel is channel 4 (about 3100-4000 cm\(^{-1}\)) with an accuracy of 96.1%, a sensitivity of 93.8%, and a specificity of 98.4%. 

**FIGURE 12.** (Continued.) Feature map of data after each layer. A: Input data; B: Average pooling 1; C: Average pooling 2; D: Average pooling 3; E: Average pooling 4; F: Convolution layer 5; G: Classification probability.
In this experiment, two kinds of tissue spectra were collected by fiber optic Raman equipment, under the condition of illumination and no light. Under the condition of illumination, the light source was mainly indoor light, and there was no natural light to keep the light source stable. It is then divided into four Raman spectral data sets with or without preprocessing. The main data used in this experiment is the pre-processed data set under no light conditions. The effects of the four data sets in this model are shown in Figure 11, respectively. What can be seen is whether there is illumination in comparison, the noise of the data in the absence of illumination is the smallest, and the Raman spectral signal
has the clearest performance. On the other hand, the preprocessed data set has better performance under the condition of preprocessing.

**IV. CONCLUSION**

This paper proposes a method for classifying spectral data of TSCC and normal tissues using fiber optic Raman spectroscopy and ensemble CNN model. In this study, the data consisted of Raman spectral information collected by the fiber optic Raman system. The identification process for the TSCC is done by the ensemble CNN model. The experimental results show that the model has high sensitivity and specificity in the whole spectrum, and the method has high accuracy in TSCC classification. The comparison with a single CNN method shows that the ensemble CNN model achieves the best performance on the spectral data. In addition, the difference features included in the frequency band of 1380-2250 cm⁻¹ are most effective for identification. The experimental results show that our work can be used to identify TSCC. We hope that the TSCC detection method combined with Raman spectroscopy and ensemble CNN model can be applied in future TSCC edge surgery.

**APPENDIX**

See Figure 12.

**REFERENCES**

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: A Cancer J. Clinicians, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[2] R. W. H. Smits, S. Koljenovic, J. A. Hardillo, I. ten Hove, C. A. Meeuwis, A. Sewnaik, E. A. C. Dronkers, T. C. Bakker Schut, T. P. M. Langeveld, J. Molenaar, V. N. Hegi, G. J. Puppels, and R. J. Baatenburg de Jong, “Resection margins in oral cancer surgery: Room for improvement,” Head Neck, vol. 38, no. S1, pp. E2197–E2203, Apr. 2016.

[3] Cancer Research UK. (2012). Oral Cancer Incidence Statistics. Accessed: Mar. 1, 2015. [Online]. Available: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/oral/ukoral-cancer-statistics

[4] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, “Global cancer statistics,” CA: A Cancer J. Clinicians, vol. 61, no. 2, pp. 69–90, Mar. 2011.

[5] E. M. Barroso, R. W. H. Smits, T. C. Bakker Schut, I. T. Hove, J. A. Hardillo, E. B. Wolvius, R. J. Baatenburg de Jong, S. Koljenovic, and G. J. Puppels, “Discrimination between oral cancer and healthy tissue based on water content determined by Raman spectroscopy,” Anal. Chem., vol. 87, no. 4, pp. 2419–2426, Feb. 2015.

[6] J. A. Woolgar and A. Triantafyllou, “A histopathological appraisal of surgical margins in oral and oropharyngeal cancer resection specimens,” Oral Oncol., vol. 41, no. 10, pp. 1034–1043, Nov. 2005.

[7] C. J. Kerawala and K. Thian Ong, “Relocating the site of frozen sections—Is there room for improvement?” Head & Neck: J. Sci. Specialties Head Neck, vol. 23, no. 3, pp. 230–232, 2001.

[8] L. J. DiNardo, J. Lin, L. S. Karageorge, and C. N. Powers, “Accuracy, utility, and cost of frozen section margins in head and neck cancer surgery,” Laryngoscope, vol. 110, no. 10, pp. 1773–1776, Oct. 2000.

[9] N. F. F. Ribeiro, D. R. P. Godden, G. E. Wilson, D. M. Butterworth, and R. T. M. Woodwards, “Do frozen sections help achieve adequate surgical margins in the resection of oral carcinoma?” Int. J. Oral Maxillofacial Surg., vol. 32, no. 2, pp. 152–158, Apr. 2003.

[10] C. Senft, A. Bink, K. Franz, H. Vatter, T. Gasser, and V. Seifert, “Intraoperative MRI guidance and extent of resection in glioma surgery: A randomised, controlled trial,” Lancet Oncol., vol. 12, no. 11, pp. 997–1003, Oct. 2011.

[11] I. P. Santos, E. M. Barroso, T. C. B. Schut, and P. J. Caspers, “Raman spectroscopy for cancer detection and cancer surgery guidance: Translation to the clinics,” Analyst, vol. 142, no. 17, pp. 3025–3047, 2017.

[12] S. J. Erickson-Bhatt, R. M. Nolan, N. D. Shemonski, S. G. Adie, J. Putney, D. Darga, D. T. McCormick, A. J. Cittadine, A. M. Zysk, Marjanovic, E. J. Chaney, G. L. Monroy, F. A. South, K. A. Cradock, Z. G. Liu, M. Sundaram, P. S. Ray, and S. A. Boppart, “Real-time imaging of the resection bed using a handheld probe to reduce incidence of microscopic positive margins in cancer surgery,” Cancer Res., vol. 75, no. 18, pp. 3706–3712, Sep. 2015.

[13] L. M. W. K. Song, “Autofluorescence imaging,” Gastrointestinal Endoscopy, vol. 73, no. 4, pp. 647–650, 2011.
M. S. Bergholt, W. Zheng, and Z. Huang, “Characterizing variability in H. Krishna, S. K. Majumder, and P. Chaturvedi, “Raman spectroscopy for the early detection of oral neoplasia: A pilot clinical study,” J. Biophoton., vol. 2, nos. 1–2, pp. 39–26, Feb. 2009.

H. Krishna, S. K. Majumder, and P. Chaturvedi, “In vivo Raman spectroscopy for detection of oral neoplasia: A pilot clinical study,” J. Biophoton., vol. 7, no. 9, pp. 690–702, 2014.

C. M. Krishna, G. D. Sockalingum, J. Kurien, L. Rao, L. Venteo, M. Phut, M. Manfait, and V. B. Kartha, “Micro-Raman spectroscopy for optical pathology of oral squamous cell carcinoma,” Appl. Spectrosc., vol. 58, no. 12, pp. 1128–1135, Dec. 2004.

M. S. Bergholt, W. Zheng, and Z. Huang, “Characterizing variability in vivo Raman spectroscopic properties of different anatomical sites of normal tissue in the oral cavity,” J. Raman Spectrosc., vol. 43, no. 2, pp. 255–262, Feb. 2012.

S. P. Singh, A. Deshmukh, and P. Chaturvedi, “In vivo Raman spectroscopic identification of premalignant lesions in oral buccal mucosa,” J. Biomed. Opt., vol. 17, no. 10, 2012, Art. no. 105002.

L.-P. Choo-Smith, H. G. M. Edwards, H. P. Endtz, J. S. Robinson, H. A. Bruining, and G. J. Puppels, “Medical applications of Raman spectroscopy: From proof of principle to clinical implementation,” Biopolymers, vol. 67, no. 1, pp. 1–9, 2002.

C. Kallaway, L. M. Almond, H. Barr, J. Wood, J. Hutchings, C. Kendall, and N. Stone, “Advances in the clinical application of Raman spectroscopy for cancer diagnosis,” Photodiagnosis Photodyn. Therapy, vol. 10, no. 3, pp. 207–219, Sep. 2013.

L. F. C. S. Carvalho, F. Bonnier, C. Tellez, L. dos Santos, K. O’Callaghan, J. O. Sullivan, L. E. Soares, S. Flint, A. A. Martin, F. M. Lyng, and H. J. Byrne, “Raman spectroscopic analysis of oral cells in the high wavenumber region,” Exp. Mol. Pathol., vol. 103, no. 3, pp. 255–262, Dec. 2017.

S. Venkatarkrishna, J. Kurien, and K. M. Pai, “Optical pathology of oral tissue: A Raman spectroscopy diagnostic method,” Current Sci.-Bangalore, vol. 80, no. 5, pp. 665–668, 2001.

A. Savitzky and M. J. E. Golay, “Smoothing and differentiation of data by simplified least squares procedures,” Anal. Chem., vol. 36, no. 8, pp. 1627–1639, 1964.

H. F. M. Boelens, P. H. C. Eilers, and T. Hankeemeier, “Sign constraints improve the detection of differences between complex spectral data sets: I.C.-IR as an example,” Anal. Chem., vol. 77, no. 24, pp. 7998–8007, 2005.

S.-J. Baek, A. Park, Y.-J. Ahn, and J. Choo, “Baseline correction using simplified least squares procedures,” Anal. Chem., vol. 77, no. 24, pp. 7998–8007, 2005.
JIABIN XIA received the M.D. degree from the School of Instrument Science and Opto-Electronics Engineering, Hefei University of Technology, Hefei, China, in 2019, where he is currently pursuing the Ph.D. degree in engineering.

He is mainly researched in oral cancer Raman spectroscopy, deep learning algorithm, and development of in vivo real-time monitoring system of Raman spectrum.

LIANQING ZHU (Member, IEEE) received the M.S. and Ph.D. degrees from the Hefei University of Technology and the Harbin Institute of Technology, in 1989 and 2013, respectively.

He is currently a Professor and a Ph.D. Supervisor with the Beijing Information Science and Technology University. His main research directions include soft robotics, sensing technology, and optical measurement.

TAO ZHANG received the M.D. and D.D.S. degrees from the Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China, in 2005. He held a Fellowship of Maxillofacial Oncology and Microvascular Reconstruction, University of Maryland Medical Center, Baltimore, USA, a Dr. Cheng Yu Tung Fellowship at the Department of Head Neck and Plastic Reconstruction Surgery, Queen Mary Hospital, and a Li Ka Shing Faculty of Medicine at The University of Hong Kong, Hong Kong. He was also as a Visiting Scholar of facial plastics and the Head and a Neck Surgery Surgery with the Harvard Medical School, Eye and Ear Infirmary, Boston, USA, and the Head and Neck Surgery and Plastic Reconstruction Surgery, MD Anderson Cancer Center, Houston, USA, supported by Chinese Government Scholarship. He is currently working with the Department of Stomatology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, and he is also the Vice Director of dentistry, a Professor, a Chief Physician, as well as a Doctoral Tutor. He is mainly engaged in clinical work such as head neck tumor, head neck plastic, and reconstruction and microsurgery.

He published 33 articles including SCI articles and core journals articles, received a Licensing. Meanwhile, his main research interests including oral cancer, head neck plastic and reconstruction, head neck tumor, microsurgery and fiber Raman spectroscopy, digital 3D printing, and augmented and mixed reality.

Dr. Zhang related work has obtained Hospital Medical Achievement Award, the China Science and Technology Award, and the China Science and Technology Award. In recent years, he has received financial support from the National Natural Science Foundation and the Beijing Natural Science Foundation.

ZHUIHUI ZHU received the M.D. degree from the Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China, in 2019. She is currently a Project Research Assistant with the Department of Stomatology, Peking Union Medical College Hospital. She is mainly engaged in oral cancer Raman spectroscopy, microsurgery, and application and promotion of a CAD/CAM cutting and positioning composite template in functional jaw reconstruction.

ZHIHUI ZHU received the M.D. degree from the Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China, in 2019. She is currently a Project Research Assistant with the Department of Stomatology, Peking Union Medical College Hospital. She is mainly engaged in oral cancer Raman spectroscopy, microsurgery, and application and promotion of a CAD/CAM cutting and positioning composite template in functional jaw reconstruction.

GUANGKAI SUN received the B.Sc. and M.Sc. degrees from the Hebei University of Science and Technology, in 2007 and 2010, respectively, and the Ph.D. degree from Beihang University, in 2015.

He is currently an Associate Professor with the School of Instruments Science and Opto-Electronic Engineering, Beihang Information Science and Technology University. His main research interests include novel optical sensing, visual measurement technologies, and soft robotics.