Journal of Biophotonics

JOURNAL OF BIOPHOTONICS

RESEARCH ARTICLE

Spatially Resolved Fibre-Optic Probe for Cervical Precancer Detection Using Fluorescence Spectroscopy and PCA-ANN-Based Classification Algorithm: An In Vitro Study

Shivam Shukla[1] | Bhaswati Singha Deo[1] | Nemichand[2] | Pankaj Singh[3] | Prabodh Kumar Pandey[4] | Asima Pradhan[1,2,5]

[1]Center for Lasers and Photonics, IIT Kanpur, Kanpur, India | [2]Department of Physics, IIT Kanpur, Kanpur, India | [3]Department of Physics, Government PG College, Unchahar, India | [4]Department of Radiological Sciences, University of California, Irvine, California, USA | [5]PhotoSpIMeDx Pvt. Ltd., SIIC IIT Kanpur, Kanpur, India

Correspondence: Asima Pradhan (pradhanasima8@gmail.com)

Received: 24 June 2024 | **Revised:** 17 September 2024 | **Accepted:** 19 September 2024

Funding: The spatially resolved fibre-optic probe used in this study was designed and fabricated under a project funded by Board of Research in Nuclear Sciences (BRNS) of Department of Atomic Energy, India (project number 2013/36/16-BRNS).

Keywords: artificial neural network | cervical cancer | epithelial cancer | fibre-optic probe | spatially resolved fluorescence spectroscopy

ABSTRACT

Cervical cancer can be detected at an early stage through the changes occurring in biochemical and morphological properties of epithelium layer. Fluorescence spectroscopy has the ability to identify these subtle changes non-invasively and in real time with good accuracy in comparison with conventional techniques. In this paper, we report the usage of a custom designed spatially resolved fibre-optic probe (SRFOP), which consists of 77 fibres in two concentric rings, for the detection of cervical cancer using fluorescence spectroscopy technique. The aim of this study is to classify different grades of cervical precancer on the basis of their fluorescence spectra followed by a robust classification algorithm. Fluorescence spectra of 28 cervical tissue samples of different categories have been recorded using six detector fibres of FOP at different spatial locations with the source fibre (SF). A 405 nm laser diode source has been utilised to excite the samples and a USB 4000 Ocean Optics spectrometer to collect the output spectra in the wavelength range 400–700 nm. Principal component analysis (PCA) was applied to the collected spectra to reduce the dimensionality of the data while preserving the most significant features for classification. The first 10 principal components, which captured the majority of the variance in the spectra, were selected as input features for the classification model. Classification was then performed using an artificial neural network (ANN) with a specific architecture, including an input layer, hidden layers, and a softmax activation function in the output layer. Experimental and classification results both demonstrate that proximal fibres (PFs) perform better than distal fibres (DFs) in capturing the discriminatory features present in the epithelium layer of cervical tissue samples as PF collect most of the signal from the epithelium layer. The combined approach of spatially resolved fluorescence spectroscopy and PCA-ANN classification techniques is able to discriminate different grades of cervical precancer and normal with an average sensitivity, specificity and accuracy of 93.33%, 96.67% and 95.57%, respectively.

1 | Introduction

Early detection of cervical cancer is crucial to reduce the morbidity and mortality rate. In this regard, many optical techniques, such as fluorescence spectroscopy, diffuse reflectance spectroscopy, Mueller matrix imaging, Raman spectroscopy, optical tomography and so on [1–6], have been explored in last several decades among which fluorescence spectroscopy-based methods

have shown promising results in detecting the abnormality at an early stage as it has the ability to identify subtle biochemical and morphological changes that occur in cervical tissue [7–10]. A typical fluorescence spectroscopy set-up consists of an excitation source, a module for light illumination and collection from the sample and a detection system which can collect output intensity as a function of wavelength [11]. Cervical tissue is a layered structure, the thick (≈1 cm) bottom layer is called stroma and the thin (≈300 μm) superficial layer is called epithelium. Both these layers have some dominant endogenous fluorophores like NADH (nicotinamide adenine dinucleotide), FAD (flavin adenine dinucleotide), collagen, porphyrin and so on. The spectral response of these fluorophores changes with the disease progression such as increase in NADH fluorescence and decrease in FAD and collagen fluorescence [12–14]. Fluorescence spectroscopy has been involved in discrimination of various grades of cervical cancer: CIN I, CIN II, and CIN III (cervical intraepithelial neoplasia) which are categorised on the basis of severity of the disease [15]. In 1989, Lohmann et al. conducted a study for the collection of native fluorescence of cervix to discriminate dysplasia and invasive carcinoma on the basis of intensity of NADH fluorescence [16]. Fluorescence spectra of cervical biopsy samples were analysed by Mahadevan et al. using 330 nm excitation wavelength and the results were compared with that of the colposcopy results [17]. Later, many researchers started using fibre optic probes for providing systems with fixed illumination and collection geometries. Since, the fluorophores present in different layers of cervical tissue are depth dependent, it is important to choose an optimised design of fibre-optic system for the investigation of abnormality [18]. Pfefer et al. suggested two fibre-optic probe designs: single fibre probe and multiple fibre probe and collected fluorescence from a two layered phantom by varying various parameters of the two probe designs [19]. In another simulation study, two different approaches: variable aperture and multi-distance, were proposed and compared for the depth resolved fluorescence measurements from the cervical tissue [20]. Later, the same group developed an angled fibre-optic probe and tested it on a two layered phantom with two different fluorophores in each layer to proof its importance in depth sensitive fluorescence spectroscopy [21]. Weber et al. implemented a model based on combination of fluorescence and reflectance spectroscopy for in vivo detection of cervical cancer and dysplasia using a fibre-optic probe [22]. In 2002, Nair et al. developed a spatially resolved fluorescence spectroscopy (SRFS) technique to extract various optical parameters for characterisation of breast tissue samples by using a linear fibre array [23]. Stasic et al. applied SRFS to estimate the fluorophore concentration and thickness of the superficial layer for a two-layer tissue mimicking phantom [24]. In recent years, many fibre-optic probes have been developed and tested on cervical tissue samples but mostly they are based on spatially resolved reflectance spectroscopy (SRRS) measurements. SRRS and SRFS techniques are highly efficient for providing information about the changes in scattering and absorption properties of the medium and molecular environment and fluorophore structure of the medium, respectively. In this regard, our group also developed a bevel angled spatially resolved fibre-optic probe (SRFOP), having 77 fibres in 1.16 mm diameter. The fibre probe was initially tested on a two-layer cervical tissue mimicking phantom to calculate optical properties using spatially resolved reflectance technique [25]. In another study, the same probe was utilised to extract superficial layer thickness and fluorophore concentration using SRFS technique [26]. Details about the ability of this SRFOP for sensing the changes in superficial layer were also explained. In this paper, that bevelled angle SRFOP have been utilised for the collection of fluorescence spectra from cervical tissue samples in vitro.

For classification purpose, machine learning based algorithms are very popular these days especially for processing and analysing raw medical data. These algorithms have the potential to process a large amount of data with good accuracy which eventually helps in disease diagnosis at an early stage. In a fluorescence-based study due to large variations in spectral parameters it is challenging to classify different grades of abnormality just by analysing the raw fluorescence spectra. Classification algorithms such as artificial neural network (ANN), random forest, support vector machine (SVM), convolutional neural network (CNN), XGBoost and ensemble learning and so on are generally used for classification purpose. All these algorithms are well established and have been explored for the classification of cervical cancer, breast cancer, oral cancer and other diseases [27–37]. Majumdar et al. utilised a SVM based approach for feature extraction and classification by combining it with recursive feature elimination (RFE) method for optical diagnosis of oral cancer [38]. Sahoo et al. used the combination of principal component analysis (PCA) and SVM on the fluorescence decay profile of fluorescence lifetime images of cervical tissue samples to classify different grades of cervical cancer [39, 40]. Three different SVM-based approaches (SVM, SVM-recursive feature elimination and SVM-PCA) were utilised by Wu et al. for the diagnosis of cervical cancer with good accuracy [41]. Amrane et al. compared the performance of naive bayes (NB) classifier and *k*-nearest neighbour (KNN) for breast cancer

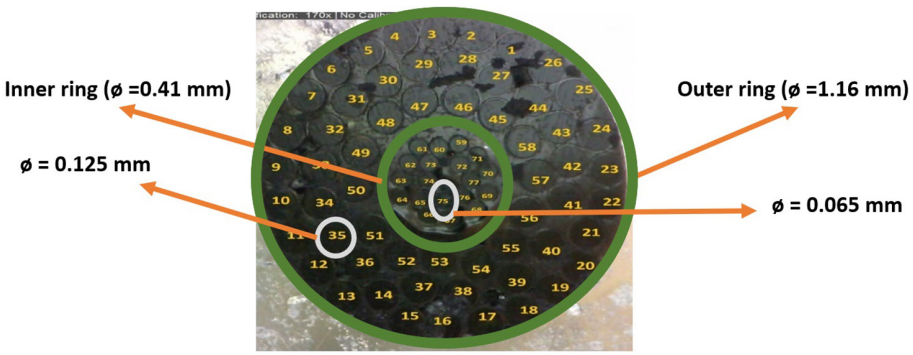


FIGURE 1 | Cross-sectional view of the front face of the (SRFOP).

classification and showed that KNN is more accurate than NB [42]. Hyperspectral images of oral samples were analysed and classified using a deep learning algorithm based on CNN by Jeyaraj et al. for early diagnosis of oral cancer [43]. Nayak et al. used PCA and SVM for classification of normal, pre-malignant and malignant oral tissues based on their collected fluorescence spectral data [44]. Normal, abnormal and cancerous cells in the cervix region were discriminated by Devi et al. using different architectures of ANN [45]. Barik et al. collected laser induced fluorescence from normal, cervicitis and cancerous cervix samples and applied PCA and SVM based model for classification purpose [46]. In recent studies, the challenge of classifying different grades of abnormality due to large variations in spectral parameters has not been fully addressed, leading to limitations in early disease detection. Moreover, existing fibre optic probes are limited in their ability to detect subtle changes occurring in the epithelial layer, which is critical for early diagnosis. This study aims to bridge the gaps in early disease detection by employing SRFOP to collect fluorescence spectra from a variety of cervix samples. Six fibres of SRFOP have been utilised for data collection from tissue samples. The most important features of the collected spectral data are selected using the PCA technique, and an ANN-based classifier is then applied to the reduced data for early detection of cervical precancer. By combining spatially resolved fluorescence spectroscopy with advanced classification

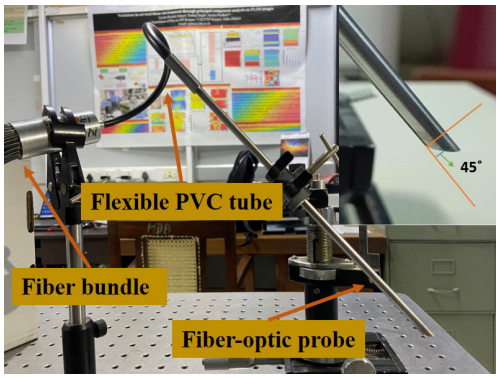


FIGURE 2 | Photograph of the (SRFOP) with angular geometry of its proximal end (inset).

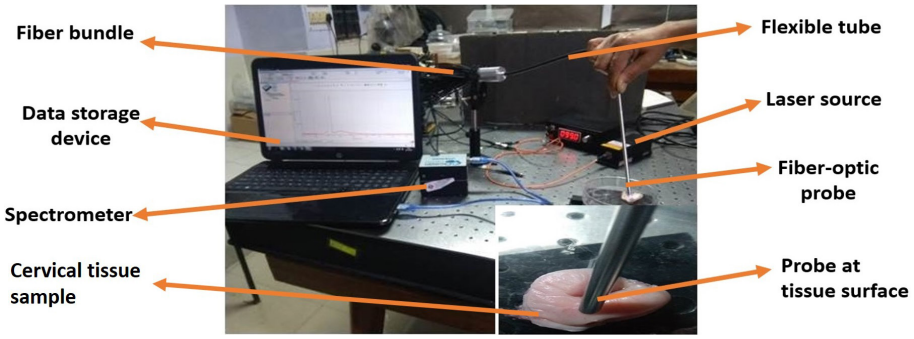


FIGURE 3 | Experimental set-up of the SRFS based system.

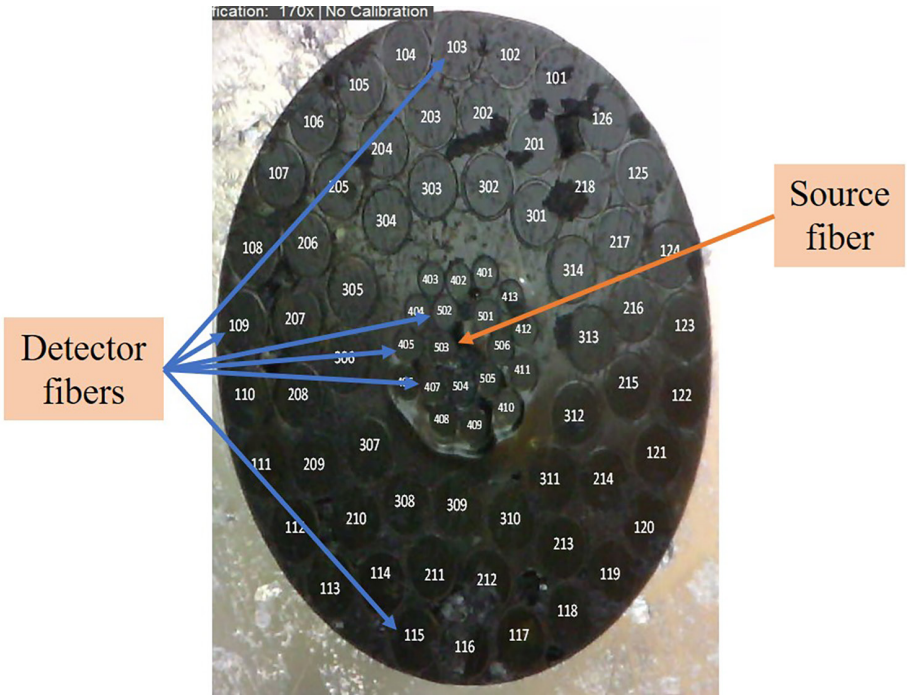


FIGURE 4 | Spatial locations of source and detector fibres in SRFOP (source fibre: 503, detector fibres: 103, 109, 115, 405, 407 and 502).

techniques, this approach enhances the accuracy and reliability of cervical precancer detection.

2 | Materials and Methods

2.1 | Design and Details of the Spatially Resolved Fibre-Optic Probe (SRFOP)

The proposed custom designed and developed SRFOP consists of 77 fibres in two concentric rings as shown in Figure 1. First ring of diameter 1.16 mm contains 58 fibres from # 1–58 with core/cladding diameter 100/125 μm and second smaller ring of diameter 410 μm has 19 fibres from # 59–77 with core/cladding diameter 50/65 μm. These all fibres have been covered in a stainless-steel cylindrical tube of diameter 5 mm and overall length 15 cm. At the proximal end, all fibres have been polished at 45° with respect to the axis of the cylindrical tube as shown in Figure 2. All fibres are decoupled at distal end from the tube and can be used as a source or detector fibre individually. Detailed description about the design and other aspects about SRFOP can be found out in our previously published paper based on spatially resolved reflectance technique which also prove its ability to provide high sensitivity for changes in epithelium layer which is crucial for early cancer diagnosis [25].

2.2 | Experimental Set-Up

The experimental set-up of the SRFS based system has been showcased in Figure 3. The whole system mainly consists of a 405 nm laser diode source (FC405, 500 mW, Shanghai, Optical System Co. Ltd.) to illuminate the cervix sample, a USB 4000 Ocean Optics spectrometer to collect the output fluorescence spectra, a data storage device (Laptop) to display the output and save it for further analysis, and a SRFOP to deliver the source light to the sample and collect the emitted back-scattered fluorescence light from the sample. The proximal end of the probe remains in direct contact with the sample to capture maximum fluorescence and to avoid any kind of noise from the ambience. One of the fibre # 503 near the centre of the probe is used as source fibre and connected to the light source through a FC-SMA fibre connector. Six other fibres from different spatial positions (demonstrated in Figure 4) of the probe are used as detector fibres and connected individually to the input port of the spectrometer. All the larger diameter distal fibres (DFs) # 103, 115 are 600 μm away from the SF # 503 and # 109 is at 475 μm. The smaller diameter proximal fibres (PFs) # 405, 407, and 502 are just 65 μm away from the SF. All the fibres chosen for experimental measurements are in triangular geometry from the SF. This way one gets the information from all directions which eventually enhances the diagnostic ability of the probe. Illumination power is kept approximately at 2 mW and integration time of the spectrometer is fixed at 1 s, for every set of measurements.

TABLE 1 | List of patients and sites from different categories used in this study.

Sr. no.	Category	No. of patients	No. of sites
1	Normal	9	38
2	CINI	11	50
3	CINII	8	33
	Total	28	121

2.3 | Samples for Testing and Validation

The fibre-optic probe has been tested on a variety of two-layered cervical tissue mimicking phantoms: (a) phantoms with different superficial layer thicknesses, (b) phantoms with different fluorophore concentrations in superficial layer and (c) phantoms with varying superficial layer concentrations and thicknesses. FAD and PpIX (Protoporphyrin IX) have been used as fluorophores in upper and bottom layer respectively. The results, presented in one of our previous works, have proven that the proposed system has the potential to identify the subtle changes occurring in the upper layer in the form of thickness and fluorophore concentration with good accuracy (not shown here) [26]. In this study, the same system has been tested on in vitro cervical tissue samples to examine its efficacy for the biological samples. Fluorescence spectra of 28 cervix samples (categorical details are shown in Table 1) have been recorded with this SRFS based system. The freshly resected cervix samples after surgery or biopsy are first stored in ice and brought to lab from the hospital. All measurements are done within 3–4 h of surgery and then again sent back to the hospital, after

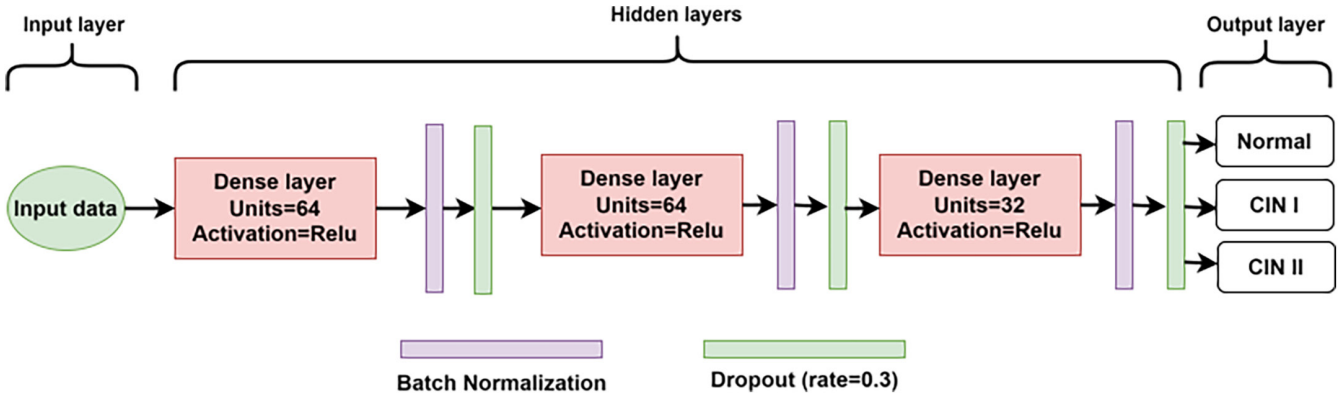


FIGURE 5 | Architecture of the proposed ANN classification network.

TABLE 2 | Architecture: Layer shapes and parameter specifications.

Layer name	Output shape	No. of parameters
Dense	(None, 64)	704
Batch normalisation	(None, 64)	256
Dropout	(None, 64)	0
Dense	(None, 64)	4160
Batch normalisation	(None, 64)	256
Dropout	(None, 64)	0
Dense	(None, 32)	2080
Batch normalisation	(None, 32)	128
Dropout	(None, 32)	0
Dense	(None, 3)	99
Total params	7683	
Trainable params	7363	
Non-trainable params	320	

preserving in a formaldehyde solution, for histopathology procedure. The whole study has been reviewed and approved by the ethics committees of both medical college and parent institute. All relevant guidelines and regulations are being followed while performing the experiments and informed consent forms are taken from all the volunteering patients and/or their local guardians(s) by the doctors.

2.4 | Classification Algorithm

A combination of PCA and ANN based algorithm has been utilised for dimensional reduction, feature selection and classification of different cervix samples using intensity values at different wavelengths in the range 470–700 nm as features. The original data was first oversampled using synthetic minority oversampling technique (SMOTE) to address class imbalance [47]. SMOTE not only helps in balancing the dataset but also prevents overfitting and improves the ability of the classifier to generalise by creating synthetic samples rather than duplicating existing ones. After applying SMOTE, the data was normalised using the standard scaler. This step standardises features by removing the mean and scaling to unit variance, ensuring equal

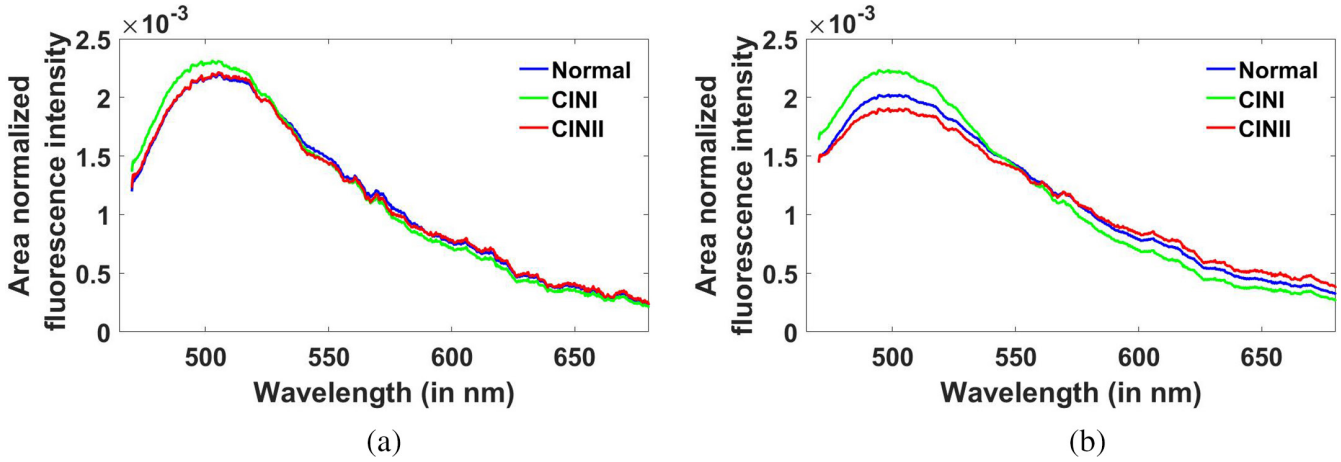


FIGURE 6 | Average area normalised fluorescence spectra for different classes of cervical tissue samples recorded through (a) distal fibres (DFs) and (b) proximal fibres (PFs) of SRFOP.

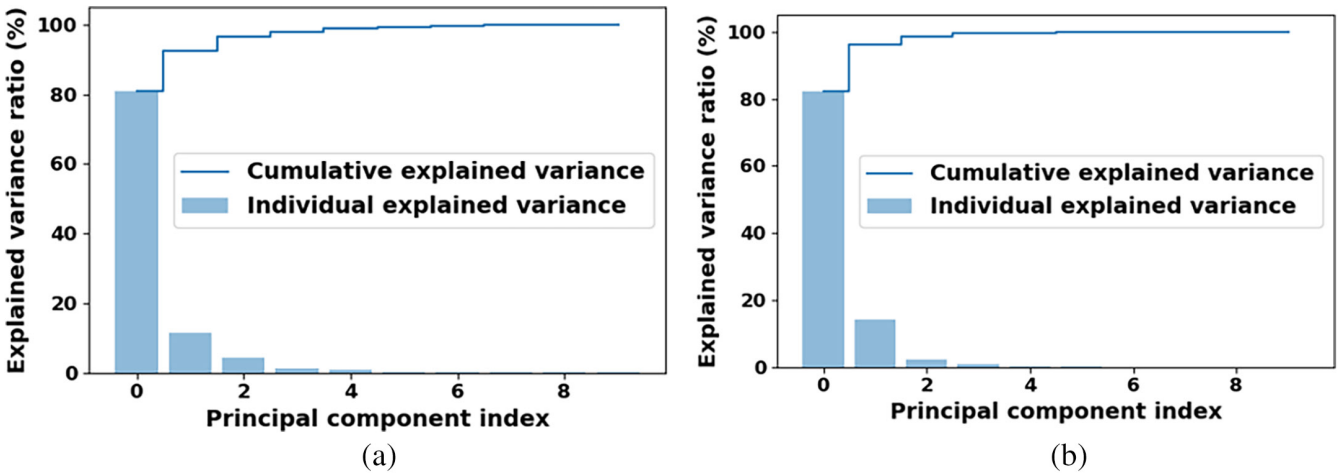


FIGURE 7 | Variance plot of different classes for (a) distal fibres (DFs) and (b) proximal fibres (PFs) of SRFOP.

contribution from each feature. Normalisation is vital for PCA as it relies on data variance to identify principal components, enhancing the effectiveness of dimensionality reduction. Since there are 933 features (different wavelength values), it is crucial to select the important features which carry more discriminatory information about the data sets than the others. PCA serves this purpose by eliminating the unnecessary features of the fluorescence spectra and choosing a certain number of features in the form of principle components (PC) which contain 95%–99% variance of the data set [48, 49]. These PC scores are further used as input to train the classification model for variety of data sets belonging to different classes of cervix samples. ANN is a machine learning algorithm which mimics human brain in its architecture and operation. It consists of three layers: Input layer, Hidden layer and Output layer. These layers are connected to each other through synapses [50, 51]. Each synapse has certain weight and that plays a crucial role in training the ANN model. Selected PC scores are stored in the input layer as features which are further sent to hidden layers. Weighted sum of the synapses and input values is performed in hidden layer and after that an activation function decides whether to pass this to next layer or not on the basis of the error between the target and predicted value of the output [52]. If the estimated error between these two is not acceptable, the weights of different neurons are adjusted for optimisation and then the final output is predicted for test dataset. The block diagram of the proposed ANN classification algorithm is shown in Figure 5. This ANN model is designed for classifying input spectral data into three categories: normal, CIN I and CIN II. Table 2 presents a detailed overview of the classification network architecture used in our study, including the shape and parameter count for each layer. The model begins with an input layer that matches the dimension of the training data. The first hidden layer consists of 64 neurons with ReLU activation, followed by batch normalisation and a dropout layer with a 30% rate to prevent overfitting. This structure is mirrored in the second hidden layer. The third hidden layer has 32 neurons with similar activation functions, also followed by batch normalisation and dropout. The output layer comprises three neurons with softmax activation, allowing the model to output probabilities for each of the three classes. This combination of dense layers, batch normalisation, and dropout layers enables the model to learn complex patterns while stabilising the training process and reducing overfitting risks. After training the ANN based model, several parameters, such as precision (P), sensitivity (SN) or recall (R), F1-score (F), specificity (SP) and accuracy (A) described in Equations (1–5) are calculated for individual classes to evaluate classification performance of the developed algorithm.

TABLE 3 | Individual variance of first 10 PCs for PF and DF.

Principle components	Variance for PF (%)	Variance for DF (%)
1	82.17	80.78
2	14.15	11.59
3	2.17	4.19
4	0.95	1.38
5	0.24	0.98
6	0.18	0.40
7	0.06	0.24
8	0.03	0.20
9	0.02	0.15
10	0.02	0.09

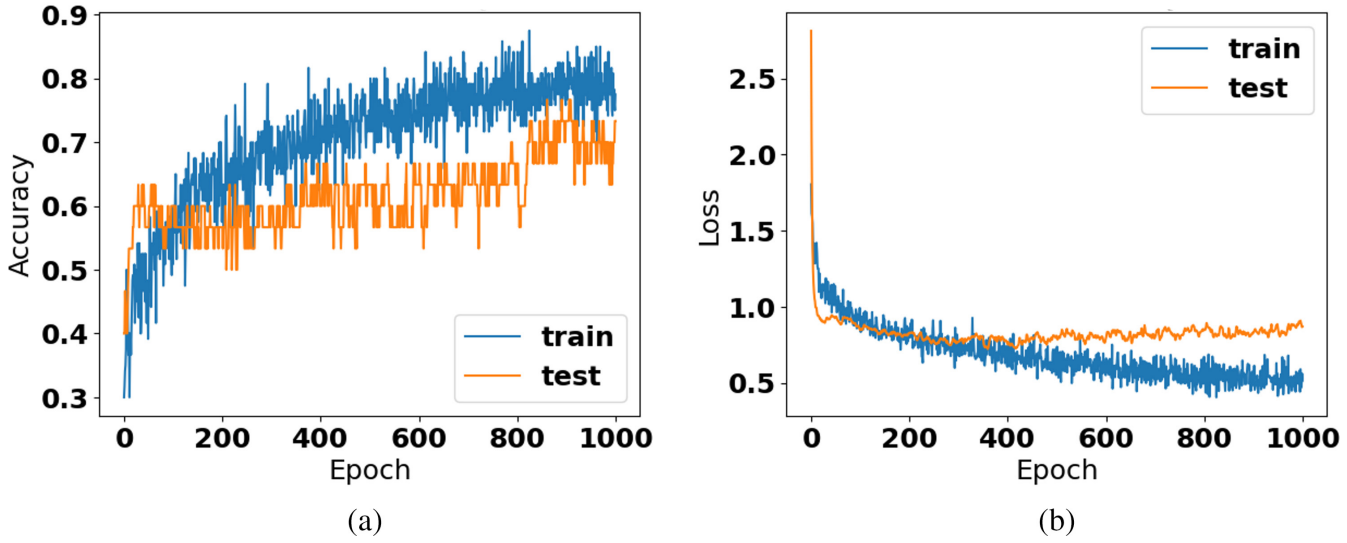
$$\text{Precision } (P) = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Sensitivity (SN) or Recall } (R) = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1} - \text{score } (F) = \frac{2 \times P \times R}{P + R} \tag{3}$$

$$\text{Specificity (SP)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

$$\text{Accuracy } (A) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$



(a) (b)

FIGURE 8 | Distal fibres: (a) accuracy with respect to epochs and (b) loss with respect to epochs.

where TP = true positive rate, FN = false negative rate, TN = true negative rate, and FP = false positive rate.

3 | Results and Discussion

Fluorescence spectra from 121 sites of 28 cervix samples have been measured using SRFOP. Spectral response from a bunch of distal and proximal fibres (DFs and PFs) have been recorded for different cervix samples. All these spectra are further analysed using MATLAB tool in the wavelength range 470–700 nm and a data matrix of dimension (121, 933) is generated. Average area normalised fluorescence spectra of different categories for DF and PF have been shown in Figure 6a,b. From these figures one can deduce that fluorescence spectra collected using PF are better discriminated than those collected using DF. Spectra, from PF, of all the three classes are well separated, especially, near the peak position which is attributed to peak of FAD fluorophore around 505 nm while for DF one can not see the difference for the entire wavelength range. Some other spectral changes can also be observed from spectra of PF and DF such as a blue shift of 7–10 nm in FAD peak positions between these two due to the fact that PF receives most of the photons from epithelium layer. It may be noted that NADH and FAD are the dominant fluorophore in the epithelium layer and collagen in the stromal layer. Hence, fluorescence spectra of PF will also have contribution from NADH fluorophore along with FAD. It implies the better epithelium sensitivity of PF of probe than the DF for abnormal changes in the cervix at early stages. Absorption dips of blood near 540 and 588 nm are clearly visible in fluorescence spectra for all the cases. This shows the ability of the probe for capturing changes occurring in spectral behaviour of fluorophores.

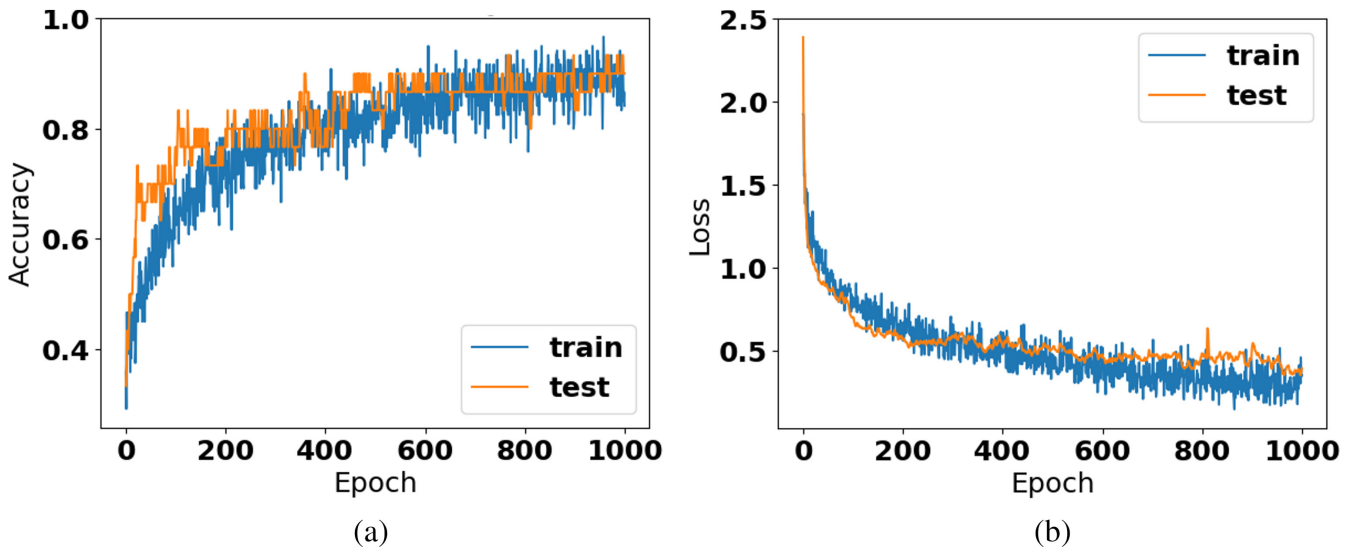


FIGURE 9 | Proximal fibres: (a) accuracy with respect to epochs and (b) loss with respect to epochs.

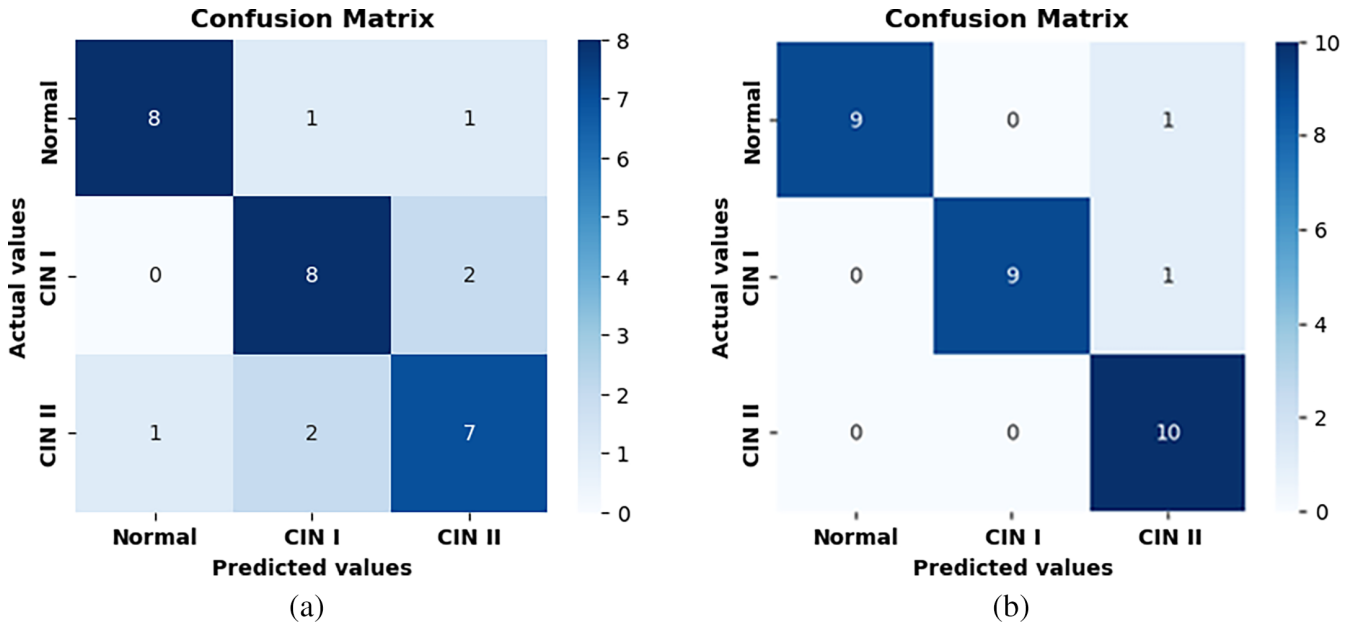


FIGURE 10 | Confusion matrices (CMs) of different classes for (a) distal fibres (DFs) and (b) proximal fibres (PFs) of SRFOP.

To address the class imbalance, the data matrix was oversampled using SMOTE, resulting in a new data matrix with dimensions (150, 933), with 50 sites for each category. All these collected fluorescence spectra are then used as inputs for training and testing a classification algorithm. The classification study, which utilises fluorescence-based spectroscopic data from various cervix classes as the input dataset, was conducted on a laptop featuring an AMD Ryzen 9 5900HX processor with Radeon Graphics. The modelling process was executed using Python 3.9.13, TensorFlow 2.7.0 and Keras 2.7.0. Additionally, the computational system was equipped with 16.0 GB of RAM. PCA is applied on the normalised data matrix of dimension (150, 933) and PC scores are calculated. Individual and cumulative variance of first 10 PCs have been plotted in Figure 7a,b for DF and PF respectively. Variance plots and individual variance values shown in Table 3 indicate that first three PCs capture almost 99% information about the original data matrix for both the cases. The normalised data matrix of dimension (150, 933) is now reduced to (150, 10) for building ANN based classification model. The input data for the ANN model was divided into training and testing sets using a stratified split, maintaining an 80:20 ratio. This method ensured that all data categories were equally represented in both sets. The Adam optimiser was used for model optimisation, with training performed over 1000 epochs at a learning rate of 0.001. Categorical Crossentropy served as the loss function for classification, and the batch size was set to 10. The training and test accuracy and loss curves for DF and PF fibres over 1000 epochs are shown in Figures 8 and 9, respectively. It shows that the model performs better for PF compared to DF. The confusion matrices (CMs) of the model using DF and PF fibres are shown in Figure 10a,b to depict the overall discrimination efficacy of the developed algorithm for different classes. Positive and negative predicted values obtained from CMs for DF and PF are listed in Tables 4 and 5. Calculated parameters: P, R or SN, F, SP and A, using equations given in Section 2.4, have been listed in Tables 6 and 7. It is evident from the Table 6 that the model achieves macro average values of precision, sensitivity, F1-score, specificity, and accuracy values as 77.20%, 76.67%, 76.80%, 88.33% and 84.44%, respectively, for DF fibres. Table 7 shows the macro average values of precision, sensitivity, F1-score, specificity and accuracy values as 94.44%, 93.33%, 93.46%, 96.67% and 95.57%, respectively, for PF fibres. Macro average values of SN, SP and A along with other evaluation parameters show that PF have better efficiency in discriminating cervix samples on the basis of the epithelial changes occurring with disease progression. Reason behind the higher values of evaluation metrics for PFs than that of DFs is due to the fact that PFs receive major signal from epithelium layer of cervical tissue because of bevelled angle nature of the spatially resolved fibre optic probe. Moreover, the fluorescence spectra of different classes of cervix samples from PF are well separated from each other, especially, near peak position (Figure 6) which eventually helps in feature extraction through PCA. From variance plots shown in Figure 7, it can be deduced that the first three PCs of PF carry 99% (approximately) variance of the input data while for DF it is 97% (approximately). It indicates that the fluorescence spectra collected using PF have more discriminatory signatures resulting in higher values of precision, sensitivity and accuracy after applying ANN.

TABLE 4 | Values of positive and negative predicted values for different classes of cervix samples for DF.

Class	TP	FN	FP	TN
Normal	8	2	1	19
CIN I	8	2	3	17
CIN II	7	3	3	17

TABLE 5 | Values of positive and negative predicted values for different classes of cervix samples for PF.

Class	TP	FN	FP	TN
Normal	9	1	0	20
CIN I	9	1	0	20
CIN II	10	0	2	18

TABLE 6 | Performance evaluation parameters of ANN-based classification algorithm for DF.

Class	Precision	Sensitivity	F1-score	Specificity	Accuracy
Normal	88.89	80.00	84.21	95.00	90.00
CIN I	72.72	80.00	76.19	85.00	83.33
CIN II	70.00	70.00	70.00	85.00	80.00
Average	77.20	76.67	76.80	88.33	84.44

TABLE 7 | Performance evaluation parameters of ANN-based classification algorithm for PF.

Class	Precision	Sensitivity	F1-score	Specificity	Accuracy
Normal	100.00	90.00	94.74	100.00	96.67
CIN I	100.00	90.00	94.74	100.00	96.67
CIN II	83.33	100.00	90.91	90.00	93.33
Average	94.44	93.33	93.46	96.67	95.57

4 | Conclusion

In this study, a SRFOP, specifically designed for capturing morphological and biochemical changes occurring in the epithelium layer of cervical tissue, has been utilised to collect fluorescence spectra from various classes of cervix samples. Six fibres of SRFOP (three distal fibres and three proximal fibres) out of 77 fibres have been used for the collection of fluorescence spectra from the cervix samples. It is evident from the collected spectra that the major peak attributed to FAD fluorophore shows slight variation from normal to CINI and CINII classes, especially for PF. A blue shift of 7–10 nm in peak positions of PF and DF depicts the presence of NADH fluorophore along with FAD which can play a vital role in identifying the discriminatory spectral features among different classes of cervix samples. A combination of PCA and ANN based algorithm is able to differentiate all the three classes with sensitivity, specificity and accuracy values of 76.67%, 88.33% and 84.44% for fluorescence spectra collected using DF and 93.33%, 96.67% and 95.57% for PF. These experimental and statistical results prove that the proposed SRFOP with fluorescence spectroscopy based technique has the potential to be used for in vivo testing on cervical cancer patients.

Acknowledgments

Authors would like to acknowledge Dr. Kiran Pandey, GSVM Medical College Kanpur, for providing cervix samples and Dr. Asha Agarwal, Pathway Diagnostic Labs Kanpur, for providing histopathology reports of samples. Authors would also like to thank Dr. Gyana Ranjan Sahoo, Cornell University, for his help in experimental measurements.

Disclosure

The authors have nothing to report.

Ethics Statement

The methodology and optical system used for this study was approved by the Human Research Ethics committee of GSVM Medical College and IIT Kanpur (IITK/IEC/2019-20/I/24).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

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

References

1. R. Alfano, G. Tang, A. Pradhan, W. Lam, D. Choy, and E. Opher, "Fluorescence Spectra From Cancerous and Normal Human Breast and Lung Tissues," *IEEE Journal of Quantum Electronics* 23, no. 10 (1987): 1806–1811.

2. A. Pradhan, P. Pal, G. Durocher, et al., "Steady State and Time-Resolved Fluorescence Properties of Metastatic and Non-Metastatic Malignant Cells From Different Species," *Journal of Photochemistry and Photobiology B: Biology* 31, no. 3 (1995): 101–112.

3. P. Shukla and A. Pradhan, "Mueller Decomposition Images for Cervical Tissue: Potential for Discriminating Normal and Dysplastic States," *Optics Express* 17, no. 3 (2009): 1600–1609.

4. M. Zaffar and A. Pradhan, "Assessment of Anisotropy of Collagen Structures Through Spatial Frequencies of Mueller Matrix Images for Cervical Pre-Cancer Detection," *Applied Optics* 59, no. 4 (2020): 1237–1248.

5. R. Shaikh, V. G. Prabitha, T. K. Dora, et al., "A Comparative Evaluation of Diffuse Reflectance and Raman Spectroscopy in the Detection of Cervical Cancer," *Journal of Biophotonics* 10, no. 2 (2017): 242–252.

6. I. Georgakoudi, E. E. Sheets, M. G. Müller, et al., "Trimodal Spectroscopy for the Detection and Characterization of Cervical Precancers In Vivo," *American Journal of Obstetrics and Gynecology* 186, no. 3 (2002): 374–382.

7. P. Kumar, S. K. Kanaujia, A. Singh, and A. Pradhan, "In Vivo Detection of Oral Precancer Using a Fluorescence-Based, In-House-Fabricated Device: A Mahalanobis Distance-Based Classification," *Lasers in Medical Science* 34 (2019): 1243–1251.

8. A. N. Sah, P. Kumar, and A. Pradhan, "In-Vivo Testing of Oral Mucosal Lesions With an In-House Developed Portable Imaging Device and Comparison With Spectroscopy Results," *Journal of Fluorescence* 1–9 (2023): 1375–1383.

9. S. Shukla, A. N. Sah, D. Hatiboruah, S. Ahirwar, P. Nath, and A. Pradhan, "Design, Fabrication and Testing of 3D Printed Smartphone-Based Device for Collection of Intrinsic Fluorescence From Human Cervix," *Scientific Reports* 12, no. 1 (2022): 11192.

10. S. Shukla, C. Vishwakarma, A. N. Sah, S. Ahirwar, K. Pandey, and A. Pradhan, "Smartphone-Based Fluorescence Spectroscopic Device for Cervical Precancer Diagnosis: A Random Forest Classification of In Vitro Data," *Applied Optics* 62, no. 25 (2023): 6826–6834.

11. N. Ramanujam, "Fluorescence Spectroscopy of Neoplastic and Non-Neoplastic Tissues," *Neoplasia* 2, no. 1–2 (2000): 89–117.

12. S. Devi, P. K. Panigrahi, and A. Pradhan, "Detecting Cervical Cancer Progression Through Extracted Intrinsic Fluorescence and Principal Component Analysis," *Journal of Biomedical Optics* 19, no. 12 (2014): 127003.

13. B. L. Meena, P. Singh, A. N. Sah, et al., "Intrinsic Fluorescence for Cervical Precancer Detection Using Polarized Light Based In-House Fabricated Portable Device," *Journal of Biomedical Optics* 23, no. 1 (2018): 15005.

14. B. S. Deo, A. N. Sah, S. Shukla, et al., "Cervical Pre-Cancer Classification Using Entropic Features and Cnn: In Vivo Validation With a Handheld Fluorescence Probe," *Journal of Biophotonics* e202300363 (2023).

15. I. Georgakoudi, B. C. Jacobson, M. G. Muller, et al., "NAD(P)H and Collagen as In Vivo Quantitative Fluorescent Biomarkers of Epithelial Precancerous Changes," *Cancer Research* 62, no. 3 (2002): 682–687.

16. W. Lohmann, J. Muβmann, C. Lohmann, et al., "Native Fluorescence of the Cervix Uteri as a Marker for Dysplasia and Invasive Carcinoma," *European Journal of Obstetrics & Gynecology and Reproductive Biology* 31, no. 3 (1989): 249–253.

17. A. Mahadevan, M. F. Mitchell, E. Silva, S. Thomsen, and R. R. Richards-Kortum, "Study of the Fluorescence Properties of Normal and Neoplastic Human Cervical Tissue," *Lasers in Surgery and Medicine* 13, no. 6 (1993): 647–655.

18. T. J. Pfefer, K. T. Schomacker, M. N. Ediger, and N. S. Nishioka, "Multiple-Fiber Probe Design for Fluorescence Spectroscopy in Tissue," *Applied Optics* 41, no. 22 (2002): 4712–4721.

19. T. J. Pfefer, L. S. Matchette, A. M. Ross, and M. N. Ediger, "Selective Detection of Fluorophore Layers in Turbid Media: The Role of Fiber-Optic Probe Design," *Optics Letters* 28, no. 2 (2003): 120–122.

20. C. Zhu, Q. Liu, and N. Ramanujam, "Effect of Fiber Optic Probe Geometry on Depth-Resolved Fluorescence Measurements From Epithelial Tissues: A Monte Carlo Simulation," *Journal of Biomedical Optics* 8, no. 2 (2003): 237–247.

21. Q. Liu and N. Ramanujam, "Experimental Proof of the Feasibility of Using an Angled Fiber-Optic Probe for Depth-Sensitive Fluorescence Spectroscopy of Turbid Media," *Optics Letters* 29, no. 17 (2004): 2034–2036.

22. C. E. R. Weber, R. A. Schwarz, E. N. Atkinson, et al., "Model-Based Analysis of Reflectance and Fluorescence Spectra for In Vivo Detection of Cervical Dysplasia and Cancer," *Journal of Biomedical Optics* 13, no. 6 (2008): 64016.

23. M. S. Nair, N. Ghosh, N. S. Raju, and A. Pradhan, "Determination of Optical Parameters of Human Breast Tissue From Spatially Resolved Fluorescence: A Diffusion Theory Model," *Applied Optics* 41, no. 19 (2002): 4024–4035.

24. D. Stasic, T. J. Farrell, and M. S. Patterson, "The Use of Spatially Resolved Fluorescence and Reflectance to Determine Interface Depth in Layered Fluorophore Distributions," *Physics in Medicine & Biology* 48, no. 21 (2003): 3459–3474.

25. P. Singh, P. Pandey, S. Shukla, N. Naik, and A. Pradhan, "Modelling, Design and Validation of Spatially Resolved Reflectance Based Fiber Optic Probe for Epithelial Precancer Diagnostics," *Applied Sciences* 10, no. 24 (2020): 8836.

26. S. Shukla, P. Singh, P. K. Pandey, et al., "Extraction of Thickness and Fluorophore Concentration of the Upper Layer in a Two-Layered Solid Phantom Using Spatially Resolved Fluorescence Spectroscopy," in *Tissue Optics and Photonics*, vol. 11363 (SPIE, 2020), 176–188.

27. M. M. Saritas and A. Yasar, "Performance Analysis of Ann and Naive Bayes Classification Algorithm for Data Classification," *International Journal of Intelligent Systems and Applications in Engineering* 7, no. 2 (2019): 88–91.

28. R. Geetha, S. Sivasubramanian, M. Kaliappan, S. Vimal, and S. Annamalai, "Cervical Cancer Identification With Synthetic Minority Oversampling Technique and PCA Analysis Using Random Forest Classifier," *Journal of Medical Systems* 43, no. 9 (2019): 1–19.

29. I. Maglogiannis, E. Zafiropoulos, and I. Anagnostopoulos, "An Intelligent System for Automated Breast Cancer Diagnosis and Prognosis Using SVM Based Classifiers," *Applied Intelligence* 30, no. 1 (2009): 24–36.

30. T. Chen and C. Guestrin, "Xgboost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), 785–794.

31. M. Kruczkowski, A. Drabik-Kruczkowska, A. Marciniak, M. Tarczewska, M. Kosowska, and M. Szczerska, "Predictions of Cervical Cancer Identification by Photonic Method Combined With Machine Learning," *Scientific Reports* 12, no. 1 (2022): 1–11.

32. E. Hussain, L. B. Mahanta, C. R. Das, et al., "A Comprehensive Study on the Multi-Class Cervical Cancer Diagnostic Prediction on Pap Smear Images Using a Fusion-Based Decision From Ensemble Deep Convolutional Neural Network," *Tissue and Cell* 65 (2020): 101347.

33. T. Xu, H. Zhang, X. Huang, et al., "Multimodal Deep Learning for Cervical Dysplasia Diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2016), 115–123.

34. M. R. Mohebian, H. R. Marateb, M. Mansourian, M. A. Mañanas, and F. Mokarian, "A Hybrid Computer-Aided-Diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning," *Computational and Structural Biotechnology Journal* 15 (2017): 75–85.

35. X. Y. Liew, N. Hameed, and J. Clos, "An Investigation of Xgboost-Based Algorithm for Breast Cancer Classification," *Machine Learning With Applications* 6 (2021): 100154.

36. D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine Learning Classification Techniques for Breast Cancer Diagnosis," in *IOP Conference Series: Materials Science and Engineering*, vol. 495(1) (IOP Publishing, 2019), 12033.

37. S. Shukla, B. S. Deo, C. Vishwakarma, et al., "A Smartphone-Based Standalone Fluorescence Spectroscopy Tool for Cervical Precancer Diagnosis in Clinical Conditions," *Journal of Biophotonics* e202300468 (2024).

38. S. Majumder, "Support Vector Machine for Optical Diagnosis of Cancer," *Journal of Biomedical Optics* 10, no. 2 (2005): 24034.

39. G. R. Sahoo, P. Singh, K. Pandey, et al., "Improving Diagnosis of Cervical Pre-Cancer: Combination of Pca and Svm Applied on Fluorescence Lifetime Images," in *Photonics*, vol. 5(4) (MDPI, 2018), 57.

40. P. Singh, G. R. Sahoo, and A. Pradhan, "Spatio-Temporal Map for Early Cancer Detection: Proof of Concept," *Journal of Biophotonics* 11, no. 8 (2018): e201700181.

41. W. Wu and H. Zhou, "Data-Driven Diagnosis of Cervical Cancer With Support Vector Machine-Based Approaches," *IEEE Access* 5 (2017): 25189–25195.

42. M. Amrane, S. Oukid, I. Gagaoua, et al., "Breast Cancer Classification Using Machine Learning," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)* (IEEE, 2018), 1–4.

43. P. R. Jeyaraj and E. R. Samuel Nadar, "Computer-Assisted Medical Image Classification for Early Diagnosis of Oral Cancer Employing Deep Learning Algorithm," *Journal of Cancer Research and Clinical Oncology* 145, no. 4 (2019): 829–837.

44. G. Nayak, S. Kamath, K. M. Pai, et al., "Principal Component Analysis and Artificial Neural Network Analysis of Oral Tissue Fluorescence Spectra: Classification of Normal Premalignant and Malignant Pathological Conditions," *Biopolymers: Original Research on Biomolecules* 82, no. 2 (2006): 152–166.

45. M. A. Devi, S. Ravi, J. Vaishnavi, and S. Punitha, "Classification of Cervical Cancer Using Artificial Neural Networks," *Procedia Computer Science* 89 (2016): 465–472.

46. A. K. Barik, S. Pavithran, N. Mithun, et al., "Laser Induced Fluorescence of Cervical Tissues: An In-Vitro Study for the Diagnosis of Cervical Cancer From the Cervicitis," *Journal of Optics* 24, no. 5 (2022): 54002.

47. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research* 16 (2002): 321–357.

48. A. George and A. Vidyapeetham, "Anomaly Detection Based on Machine Learning: Dimensionality Reduction Using PCA and Classification Using SVM," *International Journal of Computer Applications* 47, no. 21 (2012): 5–8.

49. D. Kashyap, A. Somani, J. Shekhar, et al., "Cervical Cancer Detection and Classification Using Independent Level Sets and Multi SVMs," in *2016 39th International Conference on Telecommunications and Signal Processing (TSP)* (IEEE, 2016), 523–528.

50. R. P. C. Gamara, R. Q. Neyra, and K. H. A. Recto, "Behavior-Based Early Cervical Cancer Risk Detection Using Artificial Neural Networks," in *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (IEEE, 2021), 1–6.

51. M. A. Rahman and R. C. Muniyandi, "An Enhancement in Cancer Classification Accuracy Using a Two-Step Feature Selection Method Based on Artificial Neural Networks With 15 Neurons," *Symmetry* 12, no. 2 (2020): 271.

52. M. Chen, U. Challita, W. Saad, et al., "Machine Learning for Wireless Networks With Artificial Intelligence: A Tutorial on Neural Networks," 2017, arXiv preprint arXiv:1710.02913.