# MACHINE-LEARNING ALGORITHMS FOR MULTISPECTRAL AUTOFLUORESCENCE LIFETIME IMAGING (maFLIM) BASED DETECTION OF ORAL AND SKIN CANCER LESIONS

A Dissertation

by

PRIYANKA VASANTHAKUMARI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Javier A. Jo |
| Co-Chair of Committee, | Vladislav Yakovlev |
| Committee Members, | Gerard L. Cote |
| | Ulisses Braga-Neto |
| Head of Department, | Mike McShane |

May 2022

Major Subject: Biomedical Engineering

ABSTRACT

Accurate early diagnosis of malignant skin lesions is critical in providing adequate and timely treatment. Unfortunately, initial clinical evaluation of similarly looking benign and malignant skin lesions can result in missed diagnosis of malignant lesions and unnecessary biopsy of benign ones. Image-guided tools capable of objectively discriminating malignant from benign skin lesions could potentially assist with the clinical evaluation of suspicious lesions and identify those patients benefiting the most from biopsy examination. The first part of this study tested the hypothesis that maFLIM-derived autofluorescence global features can be used in machine-learning models to discriminate malignant from benign pigmented skin lesions. Clinical widefield multispectral autofluorescence lifetime (maFLIM) dermoscopy imaging of suspicious skin lesions were acquired from 30 patients. Different pools of global image-level maFLIM features: multispectral intensity, time-domain bi-exponential, and frequency-domain phasor features, were extracted and compared. Ensemble combinations of quadratic discriminant analysis (QDA) models trained with phasor and bi-exponential features yielded a sensitivity of 84% and specificity of 90%. Simple classification models based on time-resolved autofluorescence global features extracted from maFLIM dermoscopy images have the potential to objectively discriminate malignant from benign pigmented skin lesions.

In the second part of this work, a deep learning model using Long Short-Term Memory (LSTM) networks is trained on the maFLIM images from the skin lesions. The model is trained at the pixel level and generates posterior probability maps for each image. An

average area under the curve (AUC) of the ROC curves constructed on the median posterior probabilities of 0.82±0.04 is obtained. The third part of this study classifies oral cancer lesions using LSTM based deep learning models. Early diagnosis of oral cancer is critical as 70% of lesions are presented at their advanced stages, causing a drastic decrease in the five-year survival rate. The LSTM network is trained on the maFLIM images collected from the cancerous and healthy sites on the oral cavity. The pixel-level classification produced classification maps indicating the regions of malignancy for each patient. The overall sensitivity of 86.4% and specificity of 82% were obtained, with standard deviations less than 4%. Therefore, LSTM based deep learning models have the capability to classify benign and malignant skin and oral lesions at the pixel level.

# DEDICATION

I dedicate this dissertation to my son, Aadhvik Rahul, my husband, Rahul Prasanna Kumar, my parents, K P Ramachandran Nair and Vasantha Kumari, and my mother-in-law, Beena S who have fully supported me throughout this journey.

# ACKNOWLEDGMENTS

I want to express my sincere gratitude to my doctoral committee chair, Dr. Javier Jo, for advising and motivating me during my doctoral research at Texas A&M University. I thank my committee co-chair, Dr. Vladislav Yakovlev, for continuously supporting me during this endeavor. Moreover, I cannot thank them enough for helping me sail through the challenges I faced during the duration of the research. I also thank the committee members, Dr. Gerard Cote and Dr. Ulisses Braga-Neto, for their constant encouragement and for sharing valuable inputs that helped shape the direction of my research.

I thank my collaborators at the University of São Paulo, Dr. Cristina Kurachi and Dr. Renan Romano, for acquiring the skin cancer data I used in my research and providing the information necessary to achieve my research objectives. I further acknowledge the efforts of Dr. Rodrigo Cuenca-Martinez, Dr. Shuna Cheng, and Dr. Bilal Malik in acquiring and organizing the oral cancer data I used in my research. I am grateful to Dr. Elvis Duran for guiding me especially during my initial months in the lab. I also thank the members of my research group, Dr. Michael Serafino, Dr. Ronald Juarez, and Dr. Jorge Palma, for supporting me wholeheartedly. Their suggestions and constructive criticisms have helped me significantly improve the quality of my research findings.

Finally, I thank my son, husband, parents, mother-in-law, sister, brothers-in-law, sister-in-law and nephew for their unconditional love, motivation, and support, without which this dissertation would not have been possible.

CONTRIBUTORS AND FUNDING SOURCES

## Contributors

The work in this dissertation was supervised by a committee composed of Dr. Javier Jo (chair) of the Department of Electrical Engineering, University of Oklahoma, Dr. Vladislav Yakovlev (co-chair) and Dr. Gerard Cote of the Department of Biomedical Engineering, Texas A&M University, and Dr. Ulisses Braga-Neto of the Department of Electrical Engineering, Texas A&M University.

The data acquired for this work was provided by Dr. Renan Romano, Dr. Ramon G. T. Rosa, Dr. Rodrigo Cuenca-Martinez, Dr. Shuna Cheng, and Dr. Bilal Malik. Data analysis and interpretation were conducted by the student and by Dr. Javier Jo. All other work for the dissertation was conducted by the student independently.

## Funding Sources

This work was supported by the National Institutes of Health (NIH/NCI grant R01CA218739, 1R01GM127696, 1R21GM142107) and the Cancer Prevention and Research Institute of Texas (CPRIT grant RP180588). This work was also made possible by the grant NPRP8-1606-3-322 from the Qatar National Research Fund (a member of Qatar Foundation). Research reported in this dissertation was also supported in part by the National Science Foundation (NSF) (DBI-1455671, CMMI-1826078), the Air Force Office of Scientific Research (AFOSR) (FA9550-15-1-0517, FA9550-20-1 0366, FA9550-20-1-0367), and the Army Medical Research Grant (W81XWH2010777), and by the Oklahoma Tobacco Settlement Endowment Trust awarded to the University of

Oklahoma, Stephenson Cancer Center. This study was also supported by the following Brazilian funding agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001; CNPq-PVE (401150/2014-3 and 314533/2014-1); CNPq-PQ (305795/2016-3) and São Paulo Research Foundation (FAPESP) grants: 2013/07276-1 (CEPOF); 2014/50857-8 (INCT).

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

# 1. INTRODUCTION

In 2021, approximately 608,570 people are estimated to die of cancer in the United States[1]. Of all the types of cancer, skin cancer is the most common with melanoma being the fifth most prevalent among men and women[2]. Basal cell carcinoma (BCC) is the most commonly occurring type of skin cancer, affecting 2.8 million people in the US in 2019[3]. Although BCC is considered curable, delayed diagnosis and incomplete surgical resection are two significant factors associated with the ~2,000 deaths caused by BCC reported each year in the US. The most common diagnosis for skin cancer lesions is visual examination followed by biopsy tissue resection and histopathological evaluations. One major drawback of this practice is the inability to visually distinguish between clinically similar lesions; in particular, melanoma is often mistaken for benign lesion pigmented seborrheic keratosis (pSK), and BCC for intradermic nevus. In addition, it is known that the accuracy of melanoma diagnosis with an unaided eye is only about 60%[4]. Therefore, clinical tools that could provide objective, in situ, and accurate non-invasive discrimination between cancerous and benign lesions could significantly facilitate the early detection of skin cancer. One of the most common tools physicians use to diagnose skin cancer lesions is a dermoscope[4], [5] which helps the unaided eye by magnifying the features on the skin. This allows the doctors to conduct a much-detailed examination of the morphological characteristics of the skin. Although dermoscopy has been known to improve the diagnostic sensitivity of skin lesions by ~10-30%[4], its performance largely depends on the level of experience of the dermatologists and the type of lesions. Therefore,

this method's highly subjective nature and poor reproducibility led to the emergence of several computer-aided diagnostics (CAD) systems [5], [6]. CAD systems are becoming vastly popular in diagnosing and prognosis of various diseases as they allow an automated and non-invasive analysis of the tissue conditions. Numerous predictive models are developed based on training datasets obtained from imaging modalities such as ultrasound imaging, MRI [7], coherence tomography (CT) [8], confocal microscopy [9] and even total body photographs [10]. CAD systems for skin cancer diagnosis based on dermoscopy [5], [11], [12] are the most popular ones. However, the amount of information extracted from dermoscopy images is limited as they contain only the enhanced morphological information of the skin lesions.

In addition to accurately identifying skin lesions to ensure that no malignancies are missed, it is also essential to reduce the number of unnecessary biopsies performed on benign skin lesions. This prevents unwanted pain and minimizes the costs incurred by the patients. The average sensitivity of detecting melanoma skin lesions using dermoscopic images is <80%[13]. Haenssle et al.,[13] has conducted a study with 58 dermatologists to identify melanoma using dermoscopic images and other patient clinical information. Although an average sensitivity of 88.9% was achieved, the specificity was only 75.7%. This shows that many of the existing techniques require experienced physicians to make the diagnosis and result in low specificities. This implies that the number of unnecessary biopsies is still high. Thus, the overall goal of building a CAD system for automated discrimination of benign and malignant skin lesions will be to achieve almost 85% sensitivity and specificity without needing experienced oncologists.

Oral cancer is among the most prevalent types of cancer, particularly in developing countries, due to increased tobacco and alcohol use, poor diet, weak immune system, and sun exposure. Despite their ease of accessibility in the oral cavity, almost 70% of oral cancer lesions are presented in their advanced stages, causing a drastic decrease in the five-year survival rate to ~52%[14]. However, early diagnosis could increase the five-year survival rate to ~84%. According to the American Cancer Society (ACS), the 5-year survival rate for oral cancer is about 65%. However, early diagnosis can increase the overall 5-year survival rate to 84%. More than 90% of the oral cancer lesions are squamous cell carcinoma (SCC). The doctors visually examine the lesions, and only those that appear suspicious are sent for further biopsy and histopathological evaluation. This is because biopsy in the oral cavity is a painful procedure, and there is an increased risk of infections. In addition, many pre-cancerous lesions which manifest as either leukoplakia (white) or erythroplakia (red) have possibilities of transforming to malignancy. This brings in the need to regularly monitor the prognosis of such lesions, and biopsy tissue resection is not a viable option. Therefore, it is crucial to accurately discriminate between benign, pre-cancer, and malignant oral lesions. The use of vital staining with agents such as toluidine blue [15] and brush cytology [16] has improved diagnostic accuracy in the hands of experienced oncologists. However, they exhibit low specificity, and the interpretation is highly subjective. Photosensitizers were used as exogenous fluorophores in fluorescence imaging to differentiate between different lesion types [17]. However, the usefulness of this procedure is limited by the prolonged sensitivity of the patients towards sunlight and the possible toxicity associated with the accumulation of agents in the tissue. Several

imaging techniques like magnetic resonance imaging (MRI), computed tomography (CT) and positron emission tomography (PET) [5] have been utilized in detecting oral tumors. The complex and expensive instrumentation and their use of contrast agents are some of the limitations in these methods. Although the imaging techniques mentioned above help doctors by providing an alternative to visual examination by the unaided eye, they still require experienced physicians to decide the type of lesion. The accuracies of most of the techniques are no better than just visual examination if an inexperienced person from a screening camp is establishing the diagnosis.

In one previously reported study [18], where 32 dental practioners were asked to identify 50 oral lesions from their photographs, only 57.8% sensitivity and 53% specificity was achieved. In another study [19] conducted in 2012, physicians were asked to perform conventional oral examination on lesions to identify dysplasia and carcinoma. Although a sensitivity of 93% was obtained, the specificity was only 31%. This indicates a significantly higher number of unwanted biopsies performed on the patients. Thus, the goal of developing an automated CAD system for identifying benign and malignant oral lesions will be to achieve a considerably high sensitivity and specificity of at least ~85% each.

The presence of endogenous fluorophores in the tissue enables the monitoring of autofluorescence upon excitation with specific wavelengths. This eliminates the need for staining the target area with dyes or exogenous fluorophores. Pathological conditions and neoplastic progressions in the tissue are associated with epithelial changes such as an increase in nuclear size and texture, resulting in changes in incident light's absorption and

scattering characteristics. An increase in metabolism is also observed in cancerous tissue, whereby the relative concentrations of intrinsic fluorophores [20] such as nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD), collagen, and elastin are different from that of the surrounding healthy tissue. These changes can also differentiate between benign, precancer, and malignant lesions. Several works have successfully demonstrated the capability of autofluorescence imaging in distinguishing between the lesion and the surrounding healthy tissue [17], [20]–[25]. However, their potential to differentiate between benign, dysplastic, and malignant lesions is not yet fully explored. This is because measuring the fluorescence intensity alone is not enough to distinguish between the types of lesions and results in very low specificities. For example, all benign, precancerous, and cancerous lesions exhibit a decrease in the collagen content and thus a corresponding reduction in the fluorescence intensity. Monitoring the changes in the temporal characteristics of the fluorophores, particularly NADH and FAD, could provide reliable biomarkers to identify malignant transformation in the lesions. Multispectral autofluorescence imaging (maFLIM) enables the measurement of fluorescence lifetimes of multiple fluorophores upon excitation. Moreover, the broad emission spectral bandwidth of the fluorescence intensity signal has an intrinsic disadvantage: it is difficult to multiplex between the intensities of overlapping emissions from multiple fluorophores. maFLIM overcomes this challenge by enabling the measurement of fluorescence decay lifetimes in addition to their intensities.

This work develops non-invasive clinical diagnostic tools to automatically classify skin and oral lesions using maFLIM datasets. Novel feature extraction techniques will be

explored in both time domain and frequency domain deconvolutions of maFLIM data. The significance of using global feature sets, where a single feature vector is generated per patient, in classical machine learning algorithms like quadratic discriminant analysis (QDA) will be explored. This is particularly interesting as it considerably reduces the computational overload associated with pixel-wise feature sets. Further, much of the computational complexity in pre-processing time-domain maFLIM data is reduced by introducing frequency domain analysis to extract reliable features for classification. This work also explores deep learning algorithms to classify oral and skin lesions. maFLIM data contain biochemical and metabolic tissue information that can be learned using deep learning techniques. Application of deep neural networks on maFLIM data has been previously reported to extract lifetime and intensity information [26]. However, a strategy that could automatically understand the underlying features and efficiently produce a reliable classification is yet to be explored. The technique will be a powerful tool that could potentially replace the current visual inspection through the unaided eye or other devices that enhance the morphological information of the lesions. Although this may not be a replacement for biopsy and histopathology, it could immensely aid in improving the overall accuracy together with both high sensitivity and specificity.

# 2. BACKGROUND

## 2.1. **Skin cancer**

Skin cancer is the most common type of cancer that affects people from all geographic locations, socioeconomic conditions, and ethnicities. The two types of skin cancer are malignant melanoma and non-melanoma skin cancer (NMSC). BCC and squamous cell carcinoma (SCC) are the most common subtypes of NMSC. BCC is the most common skin cancer, with an estimated 3.6 million cases every year in the US[27]. SCC is the next most common skin cancer, with an estimated 1.8 million patients diagnosed every year[27]. More people are diagnosed with skin cancer than all the other cancers combined in the US. Although most skin cancer cases belong to NMSC, most deaths due to skin cancer are accountable to malignant melanoma. In 2021, the number of melanoma deaths is expected to increase by 4.8%[28]. Although most skin cancer cases can be attributed to prolonged sun exposure, patients who underwent organ transplants are 100 times more susceptible than other people[29].

It is essential to understand the layers and constituents of the skin before diving into the causes and mechanisms of skin cancer. Human skin consists of epidermis, dermis, and subcutaneous fat. The epidermis consists of four layers corresponding to different stages of cell maturation which occur every 30 days. The deepest layer of the epidermis stratum basale contains keratinocytes and melanocytes that produce melanin to render pigment to the skin. The subsequent layers of the epidermis, the stratum spinosum, stratum granulosum, and stratum corneum contain other differentiating cells. The underlying

dermis houses different glands and hair follicles. The subcutaneous layer stores fat and helps in regulating body temperature.

The pathophysiology of sun-exposure-induced skin cancer has been widely studied [30]–[32]. The two components of ultraviolet (UV) rays in the sun are – UVA and UVB. Both the components can cause severe skin damage, gene mutations, and DNA damage. UV rays can cause extensive damage to the unprotected skin and promote carcinogenesis. The DNA present in the epidermal keratinocytes acts as photoreceptors that absorb the energy from the incident UV rays. This can initiate skin damage leading to immune suppression, free radical damage, and further, neoplastic progressions.

BCC affects the basal cells of the epidermis and accounts for 80% of all diagnosed skin cancers[31]. They develop as shiny ulcerous papules, which are often neglected as pimples. They are unlikely to metastasize to other parts of the body. However, if ignored, they keep enlarging in size and could be dangerous. SCC affects the squamous cells present in the middle and outer layers of the epidermis and accounts for 16% of all skin cancer lesions. Actinic keratosis is often the precursor to SCC. They have a crusty appearance and could metastasize via the lymph nodes in the advanced stages. They can grow deeper into the tissue and cause significant skin destruction. Although melanoma accounts for only 4% of all skin cancer lesions, more than 50% of deaths due to skin cancer is caused by melanoma[31]. Melanoma affects the melanocytes in the epidermis and appears on the skin's surface. The metrics often used for diagnosing melanoma[33] are ABCD (Asymmetry, Border, Color, and Dermatoscopic structures), Menzies method, 7-point checklist, and CASH (Color, Architecture, Symmetry, Homogeneity). Melanoma

lesions are rigorous, spread rapidly, and have a high chance of metastasizing to other parts of the body.

## 2.2. **Oral cancer**

Oral cancer is a global health problem, with its incidence varying across different countries, genders, and ethnicities. It is most prevalent in South-East Asian countries, and its incidence among men is higher than that among women[34]. The oral cancer cases in the Indian subcontinent alone account for one-third of worldwide cases[34], [35]. The cancers associated with the oral cavity, larynx, and pharynx are categorized together to form the head and neck cancer. Tobacco use is the leading cause of cancer in the oral cavity. It can be either in the form of chewing, smoking, or hookah. The other factors that can cause oral cancer are alcohol use, syphilis, human papillomavirus (HPV) and dietary deficiencies[36]. The mucous membrane lining inside the oral cavity is called the oral mucosa. It consists of a top layer, stratified squamous epithelium (or oral epithelium) with underlying connective tissue, *lamina propria,* separated by a basement membrane. The *lamina propria* and deeper connective tissues form the oral submucosa. The epithelial layer can be either keratinized or non-keratinized, depending on the region of the mouth. The epithelium at the gingiva, hard palate, and the dorsal surface of the tongue is keratinized. In contrast, the epithelium at the soft palate, floor of the mouth, inner lips, inner cheek, and the ventral side of the tongue is non-keratinized.

The most common type of oral cancer is SCC affecting the epithelial layer of the oral cavity[37]. Oral cancer affects the oral epithelium, transforming from healthy to

dysplastic to full-on malignancy. SCC on the oral epithelium occurs when the basement membrane ruptures, and the epithelial cells infiltrate the underlying submucosa. As the infiltration progresses, the characteristics of the tumor change, and it becomes hard, ulcerous, and forms island-like structures on the oral cavity. 40% of the SCCs appear at the dorsal or lateral parts of the tongue, 30% appear at the floor of the mouth and the majority of the remaining occur at the gingiva and buccal mucosa[38]. SCCs often develop at those regions of the oral cavity with non-keratinized epithelium as they are thin, permeable and can be easily affected by carcinogens. The benign lesions in the oral mucosa in the form of leukoplakia, erythroplakia or leukoerythroplakia (red/white lesions) could transform into dysplastic or malignant lesions. These benign lesions often appear in the oral cavity due to the prolonged use of tobacco and cigarettes. However not all oral cancer lesions necessary arise from pre-cancerous lesions.

Several grading metrics[36] have been developed to identify the prognosis of the tumor from its histological information. These include the pattern of the tumor, depth of the tumor, spread along nerves, and spread along the lymphatic or blood vessels. Current method for staging oral cancer lesions called the TNM[38] (tumor, node and metastasis) method helps to assess the severity of the lesion and the extent of local, regional, and distant metastasis. Chances of lymph node metastasis are high in patients with oral cancer and, around 50% of oral cancer patients showed evidence of a tumor at the cervical lymph node[36].

## 2.3. **Multispectral autofluorescence imaging**

The emission of light occurring from the electronic states is broadly categorized into phosphorescence and fluorescence. The emission is called fluorescence when the excited electron is in the singlet state (opposite spin as the ground state electron). In contrast, the emission is called phosphorescence when the excited electron is in the triplet state (the same spin as the ground state electron). The fluorescence lifetime of a molecule is defined as the average time between its excitation and its return to the ground state[39]. Typical fluorescence lifetimes are in the range of 10 ns, while phosphorescence lifetimes are in the range of milliseconds to seconds. Figure 2-1 shows the Jablonski diagram depicting the electronic transitions during fluorescence and phosphorescence. $S_0$, $S_1$, and $S_2$ are the singlet excitation states, and $T_1$ is the triplet excitation state. The energies associated with absorption, fluorescence emission, and phosphorescence emission are $h\nu_A$, $h\nu_F$, and $h\nu_P$ respectively.

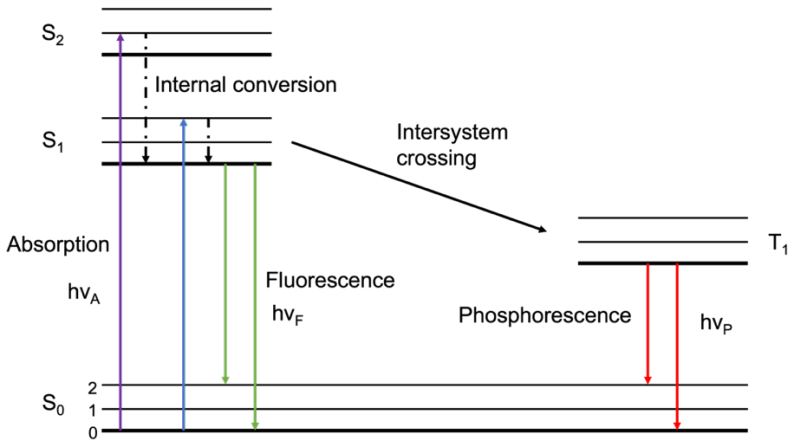



**Figure 2-1** Jablonski diagram showing electron transitions during fluorescence and phosphorescence.

The two most important properties of a fluorophore are lifetime and quantum yield. The ratio of the number of photons emitted to the number of absorbed photons is called the quantum yield. Excited electrons can return to their ground state through fluorescence emission as well as through non-radiative decay. When the number of photons returning to the ground state through the non-radiative decay process decreases, the number of photons involved in fluorescence decay increases, pushing the quantum yield closer to one. The equation for quantum yield, Q, is given by:

$$Q = \frac{\Gamma}{\Gamma + k_{nr}} \tag{2-1}$$

Where, Γ is the emissive rate of the fluorophore and $k_{nr}$ is the rate of non-radiative decay. The lifetime τ of the fluorophore is defined as:

$$\tau = \frac{1}{\Gamma + k_{nr}} \tag{2-2}$$

The fluorescence decay is a random process, and not all molecules in a sample have the same lifetime. For a single exponential decay model, a lifetime of τ means that 63% of molecules decay within τ and the remaining molecules decay after time τ. The fluorescence measurements from a sample can be broadly categorized into steady-state and time-resolved measurements. Since the time scale of fluorescence decay is in the nanosecond range, most fluorescence measurements are steady-state, involving continuous illumination from a light source. Time-resolved measurements are typically employed in applications where the nature and lifetime of fluorescence decay are required. Such measurements require a pulsed excitation of duration shorter than the lifetime of the fluorophore. The instrumentation required for time-resolved measurements is expensive

and complex. Steady-state measurements are essentially the time average of time-resolved measurements. Due to this averaging, most of the relevant information regarding the shape of the fluorescence decay is lost in the steady-state measurements. Single exponential decay for fluorescence intensity can be expressed as:

$$I(t) = I_0 \exp\left(-{}^{t}/_{\tau}\right) \quad (2\text{-}3)$$

where, $I_0$ is the intensity at time 0 and I(t) is the intensity at time t.

Time-resolved measurements can be obtained using time-domain and frequency-domain methods. In the time-domain approach, the sample is excited with a pulsed light source of duration typically less than the sample lifetime. Lifetime is calculated from the slope of log of intensity versus time. In frequency-domain measurements, the sample is excited with an intensity-modulated sinusoidal light of frequency in the range of 100 MHz. In this way, the reciprocal of the modulating frequency matches the range of the decay lifetime. The emitted light tries to follow the excitation light. However, due to the decay lifetime, there will be a delay and attenuation in the emitted light. This delay and attenuation are characterized by the phase and modulation of the emitted light signal. Fluorescence lifetimes of the sample are computed using the phase and modulation information.

The presence of endogenous or intrinsic fluorophores in the tissue enables monitoring their constituent concentrations or change in form through autofluorescence measurements. These fluorophores are produced either within the cell (intracellular) or outside (extracellular). The changes in the concentrations or the physicochemical properties of these fluorophores due to malignant transformations facilitates the use of

autofluorescence to monitor and discriminate the onset or progression of carcinoma. Monitoring tissue autofluorescence is challenging as the endogenous fluorophores can have overlapping excitation and emission bands, making it difficult to distinguish between the individual fluorophore contributions. In addition, the scattering induced by the heterogeneity of the tissues combined with weak fluorescence emission signals could make autofluorescence measurements difficult. Therefore, monitoring the fluorescence decay characteristic such as lifetime along with intensity measurements significantly improves the diagnostic value of the tissue autofluorescence measurements.

### 2.4. **Endogenous fluorophores in the skin**

The endogenous fluorophores identified in the skin are NADH, FAD, collagen, pheomelanin, eumelanin, keratin, elastin, porphyrins, tryptophan, advanced glycation end products (AGE), lipopigment, and lipofuscin[40], [41]. Table 2-1 lists the excitation and emission properties of the common endogenous fluorophores present in the skin. The excitation wavelengths used in several studies that explore skin autofluorescence ranged from 260 to 1000 nm[41]. NADH is a high energy electron carrier present in the cell's mitochondria. Nicotinamide adenine dinucleotide phosphate (NADPH) and NADH are structurally similar and have similar photophysical properties. Therefore, they are often referred to as NAD(P)H. NAD(P)H can be efficiently excited at ~340 nm. The amount of free state and protein-bound NAD(P)H indicates several biochemical changes in the tissue. Protein binding causes a significant increase in the mean lifetimes, which can be efficiently monitored using autofluorescence lifetime measurements. FAD belongs to a class of protein cofactors called flavins. They can be efficiently excited at ~450 nm, with

peak emission at ~525 nm. Unlike NADH, the average lifetime of FAD decreases upon protein binding with respect to its free state. Melanin can be excited at ~400 nm, with peak emission at ~520 nm. The fluorophores present in the extracellular matrix of the cell are elastin and collagen. Keratin is present at the surface of the epithelium, and its thickness varies with the level of keratinization. The fluorescence excitation and emission properties of keratin and collagen are very similar[42]. Changes in the extracellular matrix initiate variations in the autofluorescence responses from collagen and elastin. These changes can be due to benign lesions as well as malignant transformations. The autofluorescence emission from porphyrin occurs at longer wavelengths greater than 600 nm.

### 2.5. **Endogenous fluorophores in the oral mucosa**

The endogenous fluorophores present in the oral mucosa are NADH, FAD, collagen, elastin, and porphyrins. Keratin is also present in areas of the oral cavity with keratinized epithelium. The excitation wavelength range used in several autofluorescence studies at the oral mucosa is 320 to 460 nm, while the emission wavelength ranges from 350 to 750 nm[41]. NADH and FAD are the two most important intrinsic fluorophores indicative of the metabolic changes within the target area. An increase in cellular metabolism is the hallmark of neoplastic transformations within the tissue. The redox ratio, defined as the ratio of fluorescence intensities of NADH to FAD, is very useful in monitoring metabolic changes. A decrease in the redox ratio indicates possible malignant changes in the tissues[43]. In addition, the onset and progression of malignancies can also cause variations in the lifetimes of NADH and FAD. Thickening of the epithelium and

changes in the extracellular matrix in the underlying connective tissues can induce changes in the collagen level and hence its autofluorescence response.

**Table 2-1** Properties of endogenous fluorophores present in the skin and oral mucosa

| **Fluorophores** | **Biological constituents** | **Excitation wavelength (nm)** | **Emission wavelength (nm)** | **Lifetimes (ps)** | **References** |
|---|---|---|---|---|---|
| Elastin | Extracellular proteins | 350 to 420 | 420 to 510 | 260, 1960 | [44], [45] |
| Keratin | Structural protein | 355-405 | 420-480 | 1400 | [22] |
| Collagen | Extracellular proteins | 330 to 340 | 400 to 410 | 1000 to 1500 | [46]–[48] |
| FAD | Enzyme co-factor | 350 to 370<br>440 to 450 | 480/540 | 2300 to 2800 (free)<br>40 to 400 (bound) | [20], [49], [50] |
| NAD(P)H | Enzyme co-factor | 330 to 380 | 440 to 462 | 300 to 700 (free)<br>2500 to 3000 (bound) | [20],[46], [48], [49] |
| Melanin | Cellular pigment | 470 | 540 | 1400 | [52] |
| Porphyrins | Protein prosthetic group | 405 | 630 to 700 | 5200, 18000 to 20000 | [53] |
| Lipofuscin | Proteins, lipids | 400 to 500 | 480 to 700 | 320, 1200 | [54] |
| Aromatic amino acids | Functional proteins | 240 to 280 | 280 to 350 | 500 to 2800 | [55], [56] |

## 2.6. **Cellular respiration and autofluorescence**

Cellular respiration is a set of metabolic processes to produce energy in the form of adenosine triphosphate (ATP) from nutrients such as sugar, amino acids, and fatty acids. These processes take place in the cytosol and the mitochondria of the cell. The steps

involved in cellular respiration are glycolysis, Krebs cycle (or citric acid cycle), and oxidative phosphorylation. Glycolysis is a set of reactions involving the conversion of glucose into two molecules of pyruvate. Glycolysis does not require oxygen and takes place in the cytosol of the cell. This process also produces ATP and NADH in addition to pyruvate. The Krebs cycle occurs within the mitochondria and produces NADH and $FADH_2$, along with ATP and carbon dioxide. In oxidative phosphorylation, NADH and $FADH_2$ are oxidized through electron transport mechanisms in the mitochondria to produce significant energy as ATP. As a result of this step, NADH and $FADH_2$ are converted into $NAD^+$ and FAD. Figure 2-2 shows a schematic of the steps involved in cellular respiration. Of all the steps in cellular respiration, oxidative phosphorylation via electron transport chain is the most efficient way of producing energy in the cells.

Due to rapid cell division, neoplastic cells have a higher metabolic rate than healthy cells[20]. According to Warburg hypothesis[57], cancer cells produce energy primarily through glycolysis and subsequent lactic acid fermentation, even in the presence of sufficient oxygen. This means that the rate of mitochondrial respiration decreases while that of glycolysis increases. This is particularly helpful for those cells situated deep inside the malignant tumor to generate energy even in oxygen deficit conditions. It has been observed that the levels of NADH and FAD are higher in neoplastic cells than healthy cells. The high level of NADH is attributed to the higher glycolysis rate in comparison to oxidative phosphorylation. The high FAD level is due to the overall increase in metabolism, whereby oxidative phosphorylation produces a higher amount of FAD than

normal cells. In addition, a decrease in the redox ratio is observed in cells undergoing malignant transformation.

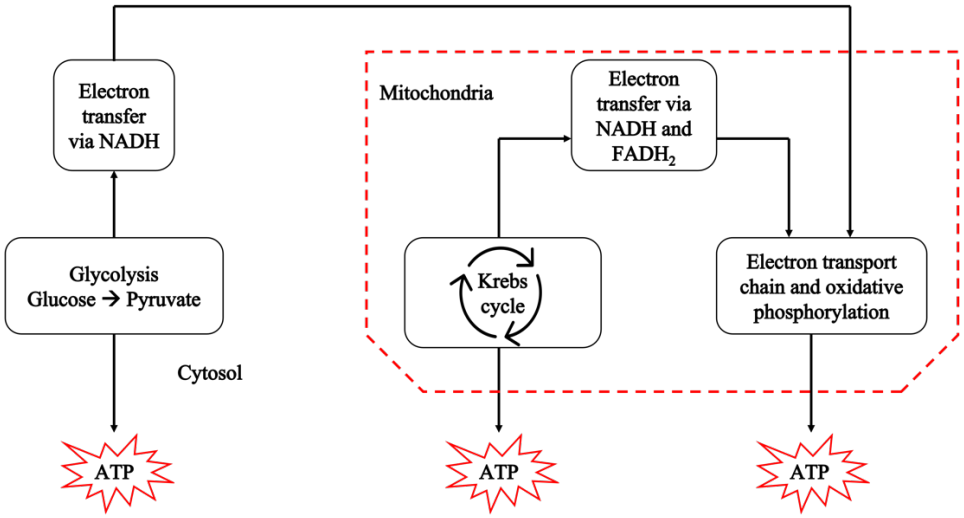



**Figure 2-2** Schematic of the steps involved in cellular respiration showing glycolysis, Krebs cycle and oxidative phosphorylation

In addition to the changes in the relative concentrations of NADH and FAD, neoplastic transformations can induce alterations in the metabolic pathways. The microenvironment of the metabolic co-factors is affected by the variation in the metabolic pathways[58]. These alterations cause changes in the relative amounts of the free and protein-bound metabolic co-factors and modifications in their binding sites. Such variations are reflected in the fluorescence decay lifetimes of NADH and FAD. The short lifetime component of NADH is attributed to its free state, while the longer lifetime component is attributed to its protein-bound state. In contrast, the short lifetime of FAD is due to its protein-bound state, while its long lifetime is due to its free state. One study[20] reported a decrease in the longer lifetime of NADH for cancerous and precancerous cells

compared to healthy cells. It also reported that the short lifetime of FAD increased with high-grade cancer compared to normal cells.

# 3. DISCRIMINATION OF CANCEROUS FROM BENIGN PIGMENTED SKIN LESIONS BASED ON MULTISPECTRAL AUTOFLUORESCENCE LIFETIME IMAGING DERMOSCOPY AND MACHINE LEARNING*

## 3.1. **Introduction**

Skin cancer is the most common type of cancer in the US, with melanoma being the fifth most prevalent among men and women[1]. The 5-year survival rate of patients with early-stage skin melanoma is ~94%; however, ~13% of skin melanoma patients are diagnosed with lesions already at intermediate or advance stages[59], which are associated with 5-year survival rates of ~61% and ~27%, respectively. The most common diagnosis strategy for skin cancer is clinical evaluation of suspicious lesions, followed by biopsy tissue resection and histopathological evaluations of selected lesions based on the clinical assessment. One major drawback of this practice is the inability to clinically distinguish between similar lesions; in particular, melanoma is often mistaken for benign pSK. In addition, it is known that the accuracy of melanoma diagnosis with unaided eye is only about 60%[4]. Therefore, clinical tools that could provide objective, in situ and accurate noninvasive discrimination between malignant and benign skin lesions during clinical evaluation could significantly facilitate early detection of skin cancer.

One of the most common tools used by physicians to diagnose skin cancer lesions is the dermoscope[4], [5], [60] which helps the unaided eye by magnifying the features on the skin. This allows doctors to conduct a much-detailed examination on the morphological features of the skin. Depending on the visual information from the dermoscopy images, using the ABCD metric, which is asymmetry, border, color, and diameter, or the recent CASH metric, which is color, architecture, symmetry, and homogeneity, the dermatologists decide whether it is necessary to conduct a biopsy procedure on the lesion or not[61]. Although dermoscopy has known to improve the diagnostic sensitivity of skin lesions by ~10-30%, its performance largely depends on both the level of experience of the dermatologists and the type of lesions[4]. The highly subjective nature and poor reproducibility of this method have led to the emergence of several proposed CAD systems[5], [11], [62], [63].

CAD systems are becoming largely popular in both diagnosis and prognosis of various diseases as they allow automated and non-invasive analysis of the tissue conditions. Table 3-1 summarizes some of the published works that reports the diagnosis and classification of pigmented skin lesions. Most of the works used dermoscopic images which were either collected by the authors or from publicly available datasets (e.g. ISIC archive, ISBI, Atlas, HAM10000 or PH2). Celebi et al.[61] explored the clinically significant colors in dermoscopic images using K-means clustering and employed symbolic regression to classify the lesions. Ramlakhan et al.[64] designed a melanoma recognition system using smart phone photographs which are classified using k-nearest neighbor (kNN) algorithm. Satheesh et al.[65] examined computerized 3D dermoscopy

features of skin cancer lesions to develop multiclass classifiers using Adaboost, bag of features (BoF) and support vector machine (SVM) techniques. Khristoforova et al.[66] used logistic regression to classify benign and malignant skin lesions using spectral features from Raman and autofluorescence spectroscopy measurements.

Classification of dermoscopy images of benign and malignant skin lesions using different deep learning approaches has also been reported. Harangi et al.[67] used an ensemble of different convolutional neural network (CNN) classifiers, while Lopez et al.[68] used transfer learning with pre-trained VGGNet CNN architecture. Majtner et al.[69] combined CNN with SVM classifier using hand crafted RSurf features and local binary patterns to classify melanomas from other benign skin lesions. Lee et al.[70] developed the WonDerm pipeline that segments the skin cancer dermoscopic images using neural network architectures and classifies it using an ensemble approach. J Amin et al.[71] extracted features using pre-trained AlexNet and VGG16 deep learning architectures, performed feature selection using principal component analysis (PCA), and applied traditional machine learning models including SVM, kNN, and discriminant analysis. J. Acosta et al.[72] utilized transfer learning with ResNet-152 architecture to classify benign and malignant skin lesions using dermoscopic images.

Visual inspection of the lesions only examines their morphological features during the pre-screening stage. However, incorporating more information about the constituents of the tissue or the rate of metabolism in the affected area could decrease the number of misdiagnosed lesions during pre-screening. It has been widely established that autofluorescence responses of intrinsic fluorophores vary significantly between normal

and neoplastic tissues[20], [25], [43], [53], [71]. Neoplastic progressions in the epithelial tissue are associated with morphological, biochemical, and functional alterations which can cause changes in the autofluorescence responses from the tissue[43], [73], [74].

**Table 3-1** Summary of previously reported works on pigmented skin lesion classification using machine learning. Reprinted with permission from P. Vasanthakumari et al, 2022.

| **Imaging modality** | **Distribution of patients or images** | **Classification task** | **Algorithm** | **Performance** | **Validation or testing technique** | **Reference** |
|---|---|---|---|---|---|---|
| Dermoscopy | Total images – 914<br><br>Melanoma - 272<br>Blue nevi - 28<br>Dysplastic nevi - 405<br>Congenital nevi - 17<br>Dermal nevi - 33<br>Dermatofibroma - 20<br>Reed/Spitz nevi - 79<br>Seborrheic keratoses - 47 | Benign vs Malignant | K-means clustering and symbolic regression | Sensitivity – 62%<br>Specificity – 76% | Train-test sets | M. Celebi, 2014[61] |
| Smart phone photograph | Benign - 37<br>Malignant - 46 | Benign vs Malignant | kNN | Accuracy -66.7%<br>Sensitivity - 60.7%<br>Specificity - 80.5% | Train-test sets | K. Ramlakhan, 2011[64] |
| Raman and Autofluorescence spectroscopy | Total patients - 56<br>Melanoma - 19<br>BCC - 18<br>Benign - 19 | Benign vs Malignant | Binary logistic regression | Accuracy - 87%<br>Sensitivity - 84%<br>Specificity - 89% | No independent validation | Y. Khristoforova, 2017[66] |

Table 3-1 Continued

| Imaging modality | Distribution of patients or images | Classification task | Algorithm | Performance | Validation or testing technique | Reference |
|---|---|---|---|---|---|---|
| Computerized dermoscopy (3-D) | PH2 dataset<br>Total images - 200<br>Melanoma<br>In-situ melanoma<br>Atypical nevus<br>Common nevus<br>ATLAS dataset<br>Total images - 63<br>Melanoma<br>In-situ melanoma<br>BCC<br>Blue nevus<br>Dermatofibroma<br>Haemangioma<br>Seborrheic keratosis<br>Normal mole | Multi-class classifier (PH2 - 4 classes and ATLAS - 8 classes) | SVM, AdaBoost, Bag of Features (BoF) | PH2 dataset<br>Sensitivity – 96%,<br>Specificity – 97%<br>ATLAS dataset<br>Sensitivity – 98%,<br>Specificity – 99% | Leave-one out cross-validation | T. Satheesha, 2016[65] |
| FLIM | Melanoma – 43<br>BCC – 28<br>SCC - 67 | Early-stage cancer vs advanced-stage cancer | RF, kNN, SVM, LDA | Accuracy – 84.62%<br>AUC - 1 | Bootstrapping | Q. Yang, 2020[75] |
| Dermoscopy | Total images – 2000<br><br>Nevus - 1372<br>Melanoma - 374<br>Seborrheic keratosis - 254 | 3 classes: Nevus, melanoma, seborrheic keratosis | Ensemble of CNNs | AUC - 0.891 | Training, validation, and test sets | B. Harangi, 2018[67] |
| Dermoscopy | ISBI 2016 database<br>Training images - 900<br>Test images - 379 | Benign vs Malignant | CNN | Accuracy - 81.33%<br>Sensitivity - 78.6%<br>Precision – 79.74% | Train-test sets | Lopez, 2017[68] |

Table 3-1 Continued

| Imaging modality | Distribution of patients or images | Classification task | Algorithm | Performance | Validation or testing technique | Reference |
|---|---|---|---|---|---|---|
| Dermoscopy | ISIC database<br>Training images - 900<br>Benign – 727<br>Melanoma - 173<br>Test images - 379<br>Benign – 304<br>Melanoma - 75 | Benign vs Malignant | CNN + SVM | Accuracy – 80.5%<br>Sensitivity – 53.3%<br>Specificity – 87.2% | Train-test sets | Majtner, 2016[69] |
| Dermoscopy | HAM10000 dataset<br>Training images – 10,015<br>Validation images – 193<br>Test images – 1512 i | 7 classes<br>Melanoma,<br>Melanocytic nevus,<br>BCC,<br>actinic keratosis,<br>Bowens disease,<br>Benign keratosis,<br>Dermatofibroma,<br>Vascular lesion | WonDerm Fine-tuned neural networks (Ensemble) | Validation accuracy – 89.9%<br>Test accuracy – 78.5% | Training, validation, and test sets | Y. Lee 2018[70] |
| Dermoscopy | ISBI and PH2 datasets<br>Benign – 3319<br>Melanoma –830 | Benign vs melanoma | Feature extraction using AlexNet and VGG16<br>Classification – ensemble of kNN, discriminant analysis, SVM and Tree | Accuracy – 99.0%<br>Sensitivity - 99.52%<br>Specificity - 98.59% | 5-fold CV and 0.5 hold out CV | J. Amin (2020)[71] |
| Dermoscopy | ISIC dataset<br>Malignant - 117<br>Benign - 481 | Benign vs malignant | CNN - ResNet 152 | Accuracy – 90.4%<br>Sensitivity – 82.0%<br>Specificity – 92.5% | Train-test sets | J.M. Acosta (2021)[72] |

CNN – Convolutional neural network
RF – Random forests
LDA – Linear discriminant analysis
SVM – Support vector machine
kNN – k-nearest neighbor
FLIM – Fluorescence lifetime imaging

The skin has several intrinsic fluorophores such as NADH, FAD, collagen, elastin, keratin, melanin, and porphyrins[23], [41], [76]. The levels of the two metabolic cofactors, NADH and FAD has been shown to vary depending on the progression of cancer[20], [74]. Optical redox ratio, defined as the ratio of fluorescence intensity of NADH to FAD is sensitive to metabolic changes which can indicate malignant transformations in the tissue. In addition, changes in the connective tissue and extracellular matrix due to the neoplastic transformations can also cause variations in the autofluorescence response from collagen and/or keratin. Thus, preferential excitation of these endogenous fluorophores in the tissue by multiple excitation sources could shed light on the biochemical changes in the target lesion area[22]. The broad emission spectral bandwidth of the fluorescence intensity signal has an intrinsic disadvantage in that it is difficult to differentiate between the intensities of overlapping emissions from multiple fluorophores. Time-resolved technique like maFLIM overcomes this challenge by monitoring the rate of decay or the fluorescence lifetime in addition to the intensities.

Several animal and human tissue studies have been published on the autofluorescence properties of skin cancer lesions[23]–[25], [40], [77]–[79]. Pastore et al.[80] conducted experiments with mouse models to study the autofluorescence response from melanoma skin lesions using multiphoton excitation at 740 nm and 900 nm, and emission spectral bands at 447 nm and 540 nm. A significant difference in the bound and free NADH ratio between cancerous and non-cancerous sites was observed, while the fluorescence decay obtained from targeting FAD remained almost the same between the two regions. It was also mentioned that the presence of melanin in the deeper layers of the skin tissue could

interfere with the overall fluorescence response from the lesions. Miller et al.[78] studied the autofluorescence emission properties between SCC bearing and normal mice skin under 480 nm excitation, and a decrease in the short lifetime component for SCC in comparison to normal skin was observed for 535 nm emission band. Drakaki et al.[79] studied the autofluorescence responses from mouse, chicken, and pig skins under UV excitation, and the structural differences and the variations in tissue constituents were investigated between the different animal species for an emission spectral band between 340 to 950 nm. De Beule et al.[40] investigated the autofluorescence response from ex-vivo biopsy skin lesions under 355 and 440 nm excitations, and the average fluorescence lifetime was found to be useful in discriminating BCC from normal skin tissues at the emission band between 390 and 600 nm. Galletly et al.[77] imaged unstained human biopsy samples using maFLIM under a 355 nm pulsed laser excitation, and significant differences in the mean fluorescence lifetimes for the emission wavelengths 375 and 455 nm were observed between the autofluorescence responses from BCC skin lesions and healthy skin. Lohmann et al.[81] analyzed the fluorescence intensities at the emission wavelength 470 nm, from human skin with melanoma, nevi, and dysplastic nevi lesions under 365 nm excitation, and a significant difference in fluorescence intensities was observed for melanoma and nevi lesions, while melanoma and dysplastic nevi lesions did not show much difference.

In this work, we developed and validated a label-free and objective image-guided strategy for the clinical evaluation of suspicious pigmented skin lesions based on maFLIM dermoscopy. In addition, a computationally efficient frequency-domain deconvolution of

maFLIM data is explored, and three different pools of global image-level maFLIM features were evaluated for machine-learning based objective discrimination between malignant and benign pigmented skin lesions.

### 3.2. **Methods**

A summary of the complete methodology performed in this study is shown in Figure 3-1.

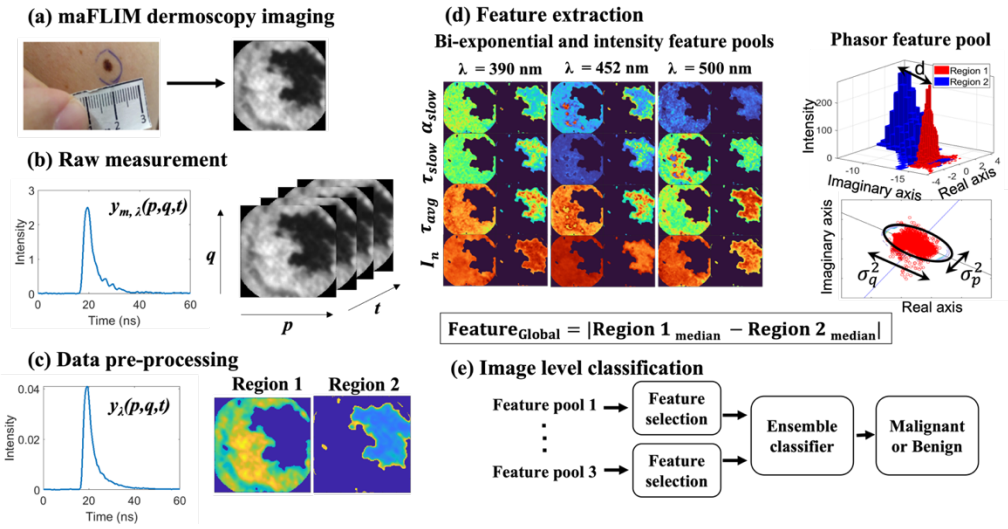



**Figure 3-1** Summary of methodology showing maFLIM image acquisition, pre-processing, feature extraction and classification. Reprinted with permission from P. Vasanthakumari et al, 2022.
maFLIM – multispectral autofluorescence lifetime imaging

#### 3.2.1. **maFLIM instrumentation**

The in-house developed time-domain maFLIM dermoscope used in this study is previously described by Romano et al., 2020[82]. The instrument consists of a handheld probe of length 140 mm and weight less than 350 g. The tube houses all the optical components and has 4 m length fibers and cables connected to it. The schematic of the instrument is provided in Fig 1. of the paper [82]. Briefly, a frequency tripled 355 nm

Nd:YAG Q-switched laser is used for exciting the target area through a 4 m long multimode excitation fiber. A custom-built raster scanning relay system helps to collimate, scan, and focus the beam on to the focal plane. Each image pixel is excited by a single excitation pulse and therefore, the pixel acquisition rate is same as the laser repetition rate (<10 kHz).

The autofluorescence emission signal from the tissue is collected by the same scanning relay system. The excitation beam is separated by using dichroic mirrors. The emission signal is sent to a multispectral module by a collection fiber and separated into three spectral emission channels by using combinations of dichroic mirrors and optical filters. The three separated channels are then time multiplexed by three optical fibers of lengths 1, 13 and 25 meters. The different lengths of the fibers introduce time delays to the signals from different spectral channels, which are then coupled to a microchannel plate photomultiplier tube (MCP-PMT). The output of the MCP-PMT is then amplified and sampled at 2.5 GHz. The laser trigger, acquisition operations and scanning controls are performed using a PC system equipped with the LABVIEW software.

### 3.2.2. **maFLIM dermoscopy imaging of skin lesions**

A total of 30 patients ($n_{patients}$ = 30) from the Dermatology Department of the Amaral Carvalho Cancer Hospital (Jahu, Sao Paulo, Brazil) were recruited for this study, following a human study protocol approved by the Internal Review Board of that institution (CAAE: 71208817.5.00005434). Only patients presenting at least one pigmented skin lesion undergoing biopsy examination for skin cancer diagnosis were recruited. The pigmented skin lesions considered in this work are solar lentigo, pSK,

pigmented superficial BCC, pigmented nodular BCC and melanoma. maFLIM images were obtained from clinically suspicious lesions using the in-house developed time-domain maFLIM dermoscope described in section 3.2.1. With this maFLIM dermoscope, skin tissue autofluorescence is simultaneously imaged at three emission bands (390±20 nm, 452±22.5 nm, and >496 nm, preferentially targeting collagen, NADH and FAD autofluorescence emission, respectively) with a temporal resolution of 0.4 ns, field-of-view (FOV) of 8.65×8.65 $mm^2$, and lateral resolution of 120 μm. For the rest of the paper, the emission wavelengths at the three spectral channels will be more conveniently referred to as 390 nm, 452 nm, and 500 nm. After signing the corresponding written informed consent form, each patient underwent the following imaging protocol right before the scheduled biopsy examination procedure. First, the lesion was gently cleaned with a gauze soaked in a saline solution. Then, the tip of the maFLIM dermoscope, previously disinfected using a gauze soaked in 70% ethanol, was placed in contact with the lesion, and an maFLIM image was acquired. The imaging site was selected so that regions of both lesion and apparent healthy skin tissue were present within the FOV of the maFLIM dermoscope. Right after maFLIM imaging, lesion tissue biopsy was performed following standard procedures. Each maFLIM image was labeled based on the histopathological evaluation of the lesion biopsy, which was considered the gold standard in this study. All images were acquired with an average laser excitation power of 10 mW measured at the sample, 140×140 pixels per image, and at a pixel rate of 10 kHz. These image acquisition parameters corresponded to an acquisition time of 1.96 s per image and an excitation energy exposure of 1.96 mJ at the sample, which is significantly lower than the maximum

permissible exposure (MPE) levels for skin based on guidelines from the American National Standards Institute – ANSI.69 The total number of lesions imaged from the 30 patients was 60 (i.e., $n_{lesions} = 60$).

3.2.3. **maFLIM data pre-processing**

Pixel-level preprocessing: The maFLIM data measured at each image pixel $(p, q)$ is composed of three fluorescence intensity temporal decay signals $y_{m,\lambda}(p, q, t)$ measured at the three targeted emission spectral bands (λ). The preprocessing steps applied to each pixel maFLIM temporal signal is shown in Figure 3-2(a). First, offset and background subtraction was applied to the raw maFLIM signal, $y_{m,\lambda}(p, q, t)$, followed by spatial averaging (order 5x5) to increase the signal-to-noise ratio (SNR) of the time-dependent signal. Second, the duration of the temporal decay signals for all emission bands was adjusted to the length of the longest signal among the three emission channels, which is 149 temporal samples (59.6 ns) by applying zero padding to the short signals. Finally, the signals from the three emission channels, $y_{\lambda}(p, q, t)$, are concatenated to form $y(p, q, t)$ as shown in equation 3-1. Signal concatenation is essential for cluster analysis in image level preprocessing, explained later in this section, as well as for frequency domain deconvolution explained in section 3.2.4.2. The concatenated signal at each pixel location can be represented as:

$$y(p, q, t) = \sum_{n=0}^{2} y_{\lambda_{n+1}}(p, q, (t - M.n)) \quad (3\text{-}1)$$

where, $y(p, q, t)$ is the pre-processed concatenated maFLIM decay signal; $y_{\lambda_1}(p, q, t)$, $y_{\lambda_2}(p, q, t)$, and $y_{\lambda_3}(p, q, t)$ are the pre-processed maFLIM decay signals from each of the three spectral channels, $\lambda_1 = 390\ nm$, $\lambda_2 = 452\ nm$, and $\lambda_3 = 500\ nm$;

M is the temporal spacing between the signals from the three channels; $(p, q)$ indicates the pixel locations. The value of M is equal to 149, which is the length of the fluorescence emission decays in each channel.

Image-level preprocessing: Pixels presenting either signal saturation or low SNR (<15dB) were detected and masked. The majority of the acquired maFLIM images contain pixels from within and outside the skin lesion region; thus, in order to group pixels based on their region of origin, the following cluster analysis was performed on the concatenated pixel-level signal, $y(p, q, t)$. First, an unsupervised K-means clustering algorithm was applied to generate two cluster masks. Then, each cluster mask is applied to the maFLIM image to define two regions within the FOV of the maFLIM image. It should be noted that since the K-means clustering algorithm involves random initialization of cluster centroids, it is difficult to identify which cluster mask belongs to within or outside the skin lesion region; thus, the identified regions were taken as two arbitrary regions – Region-1 and Region-2. Figure 3-2(b) shows an example of the cluster mask and the two separated regions generated from a representative maFLIM image.

### 3.2.4. **Feature extraction**

#### 3.2.4.1. **Features based on time-domain deconvolution parameter estimation**

In the context of time-domain maFLIM data analysis, the fluorescence decay $y_\lambda(p, q, t)$ measured at each emission spectral band (λ) and spatial location $(p, q)$ can be modeled[39] as the convolution of the fluorescence impulse response (FIR) $h_\lambda(p, q, t)$ of the sample and the measured instrument response function (IRF) $u_\lambda(t)$:

$$y_\lambda(p, q, t) = u_\lambda(t) * h_\lambda(p, q, t) \quad (3\text{-}2)$$

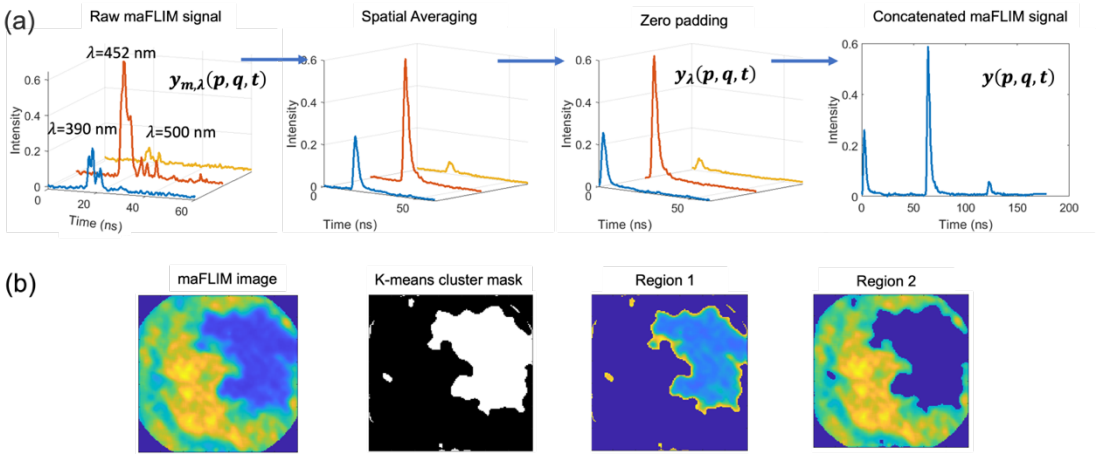



**Figure 3-2** (a) Transformations in a single pixel multispectral maFLIM data during pixel-level pre-processing. (b) Example maFLIM image with K-means cluster mask and the two separated regions. The images map the total integrated intensity of the maFLIM signals at each pixel location. Reprinted with permission from P. Vasanthakumari et al, 2022.
maFLIM – multispectral autofluorescence lifetime imaging

The standard method for time-domain maFLIM data analysis proceeds by first deconvolving the IRF of each spectral band ($u_\lambda(t)$) from the corresponding measured time-resolved fluorescence signal $y_\lambda(p,q,t)$ in order to estimate the sample FIR for each image pixel, $h_\lambda(p,q,t)$[39], which is usually modeled as a multi-exponential decay. The model order (number of exponential components) can be selected by analyzing the model-fitting mean squares error (MSE) as a function of the model order. For the maFLIM data of this study, a model order of two was selected, since the addition of a third component did not reduce the MSE; thus, the FIR was modeled as:

$$h_\lambda(p,q,t) = \alpha_{fast,\lambda} e^{\frac{-t}{\tau_{fast,\lambda}(p,q)}} + \alpha_{slow,\lambda} e^{\frac{-t}{\tau_{slow,\lambda}(p,q)}} \quad (3\text{-}3)$$

where, $\tau_{fast,\lambda}$ and $\tau_{slow,\lambda}$ represent the time-constant (lifetime) of the fast and slow decay components, respectively; while $\alpha_{fast,\lambda}$ and $\alpha_{slow,\lambda}$ represent the contribution of the fast

and slow decay components, respectively. The average fluorescence lifetime for each spectral band at each pixel location is computed using equation 3-4.

$$\tau_{avg,\lambda}(p,q) = \frac{\int t h_\lambda(p,q,t)dt}{\int h_\lambda(p,q,t)dt} \quad (3\text{-}4)$$

The parameters of the bi-exponential decay model are estimated for each pixel by nonlinear least squares iterative reconvolution[39] . After deconvolution, the bi-exponential parameters estimated at each pixel can be used as features representing the temporal dynamics of the fluorescence decays at each emission spectral band: $\alpha_{fast,\lambda}(p,q)$, $\alpha_{slow,\lambda}(p,q)$, $\tau_{fast,\lambda}(p,q)$, $\tau_{slow,\lambda}(p,q)$ and $\tau_{avg,\lambda}(p,q)$. Since the sum of $\alpha_{fast,\lambda}(p,q)$ and $\alpha_{slow,\lambda}(p,q)$ is equal to one, only one of them is kept as a feature.

In addition, the following spectral features can be also estimated from the deconvolved FIR, $h_\lambda(p,q,t)$. Absolute fluorescence intensities $I_\lambda(p,q)$ for each emission spectral bands are simply computed by time integrating the FIR $h_\lambda(p,q,t)$:

$$I_\lambda(p,q) = \int h_\lambda(p,q,t)dt \quad (3\text{-}5)$$

The normalized fluorescence intensities $I_{\lambda,n}(p,q)$ can be also computed from the multispectral absolute fluorescence intensities $I_\lambda(p,q)$ as follows:

$$I_{\lambda,n}(p,q) = \frac{I_\lambda(p,q)}{\sum_\lambda I_\lambda(p,q)} \quad (3\text{-}6)$$

Lastly, the ratio of absolute intensities from the three spectral channels is computed at each pixel location resulting in three additional features:

$$I_{390,n}/I_{452,n}(p,q) = \frac{I_{390,n}(p,q)}{I_{452,n}(p,q)} \quad (3\text{-}7)$$

$$I_{452,n}/I_{500,n}(p,q) = \frac{I_{452,n}(p,q)}{I_{500,n}(p,q)} \quad (3\text{-}8)$$

$$I_{390,n}/I_{500,n}(p,q) = \frac{I_{390,n}(p,q)}{I_{500,n}(p,q)} \quad (3\text{-}9)$$

In general, features can be computed at the pixel or the image level. In this study, image-level global features were explored, whereby one set of features, a single feature vector, is estimated to represent the whole image. Each global feature was computed from the corresponding pixel-level FLIM feature map as follows. Based on the two regions identified per image using the cluster analysis described in section 3.2.2, the feature median value for each region was computed, and the absolute value of their difference was taken as the global feature:

$$Feature_{Global} = |median(Feature_{pixel-level,Region-1}) - median(Feature_{pixel-level,Region-2})| \quad (3\text{-}10)$$

These defined global features have the advantage of reducing patient-to-patient variability in the extracted features, which is particularly important as the color and texture of skin varies considerably with ethnicity and age. This feature extraction approach based on time-domain deconvolution of the maFLIM data generates a total of six intensity and twelve bi-exponential global maFLIM features, as summarized in Table 3-2.

**Table 3-2** Feature set showing both intensity and bi-exponential global maFLIM feature pools. Reprinted with permission from P. Vasanthakumari et al, 2022.

| Intensity features | | Bi-exponential features | | |
|---|---|---|---|---|
| $I_{390,n}$ | $I_{390,n}/I_{452,n}$ | $\alpha_{fast,390}$ | $\alpha_{fast,452}$ | $\alpha_{fast,500}$ |
| $I_{452,n}$ | $I_{452,n}/I_{500,n}$ | $\tau_{slow,390}$ | $\tau_{slow,452}$ | $\tau_{slow,500}$ |
| $I_{500,n}$ | $I_{390,n}/I_{500,n}$ | $\tau_{fast,390}$ | $\tau_{fast,452}$ | $\tau_{fast,500}$ |
| | | $\tau_{avg,390}$ | $\tau_{avg,452}$ | $\tau_{avg,500}$ |

3.2.4.2. **Phasor-based features from frequency-domain deconvolved signals**

As mentioned in section 3.2.3.1, the parameters of the bi-exponential decay model are estimated for each pixel by nonlinear least squares iterative reconvolution, which is computationally expensive and time consuming. This brings about the need to develop a much simpler algorithm for extracting maFLIM features with comparatively similar discriminative capability. An alternate fitting-free strategy explored in this work is inspired by Campos-Delgado et al.[84] where a model free representation of maFLIM data was developed utilizing the frequency domain properties of the phasor representations. Here, we aim to replace the iterative time-domain deconvolution process by a simple division operation in the frequency domain. The computational overload is further reduced by processing the fluorescence decays of all the three spectral channels together, unlike in the traditional method where the maFLIM signal for each spectral channel must be processed separately. Subsequently, several features can be extracted from the frequency-domain phasor representation of the maFLIM data[85], [86]. The presented method proceeds in three steps: 1) performing frequency-domain deconvolution of the instrument response from the concatenated pre-processed fluorescence decays from all the three spectral channels, 2) constructing phasor plots for the maFLIM data, and 3) extracting global features from the phasor plots representing each maFLIM image. A detailed description of this method is presented as follows.

In this approach the pre-processed and concatenated maFLIM signals at each pixel location, $y(p, q, t)$ are normalized to sum one for further processing and feature

extraction. Similarly, the concatenated IRF from all the three spectral channels can be mathematically represented as:

$$u(t) = \sum_{n=0}^{2} u_{\lambda_{n+1}}(t - M.n) \quad (3\text{-}11)$$

where, u(t) is the concatenated IRF; $u_{\lambda_1}(t)$, $u_{\lambda_2}(t)$and $u_{\lambda_3}(t)$ are the IRF signals from the three spectral channels; M is the temporal spacing between the signals from the three channels; $(p, q)$ indicates the pixel locations.

The first step of the algorithm is to compute the Fourier Transform (FT) of both the pre-processed concatenated signal and the concatenated IRF. The FT of the signals $y(p, q, t)$ and $u(t)$ can be represented as $Y(p, q, \omega)$ and $U(\omega)$, respectively, where $\omega$ is the angular frequency.

$$Y(p, q, \omega) = FT\{y(p, q, t)\} \quad (3\text{-}12)$$

$$U(\omega) = FT\{u(t)\} \quad (3\text{-}13)$$

If the effective FIR from all the three fluorescence emission channels is denoted as $h\,(p, q, t)$, the effective fluorescence frequency response $H(p, q, \omega)$ is obtained from the FT of $h\,(p, q, t)$ as:

$$H(p, q, \omega) = FT\{h\,(p, q, t)\} \quad (3\text{-}14)$$

Therefore, the convolution in equation 3-1 can be represented as a multiplication in the frequency domain as:

$$Y(p, q, \omega) = H(p, q, \omega)\, U(\omega) \quad (3\text{-}15)$$

To uniformly scale all the frequency components of $Y(p, q, \omega)$, normalization with respect to the DC response $Y(p, q, 0)$ can be applied as follows:

$$\frac{Y(p,q,\omega)}{Y(p,q,0)} = \frac{H(p,q,\omega)\, U(\omega)}{H(p,q,0)\, U(0)} \quad (3\text{-}16)$$

Subsequently, the normalized fluorescence frequency response $P(p,q,\omega)$ can be estimated as:

$$P(p,q,\omega) = \frac{H(p,q,\omega)}{H(p,q,0)} = \frac{Y(p,q,\omega)\,U(0)}{U(\omega)\,Y(p,q,0)} \tag{3-17}$$

A phasor representation for the normalized frequency response $P(p,q,\omega)$ at specific values of the frequency $\omega$ can be generated by plotting the real $Re\,[P(p,q,\omega)]$ vs. the imaginary $Im\,[P(p,q,\omega)]$ components of $P(p,q,\omega)$. Therefore, each pixel of the maFLIM image is mapped to a point in the corresponding phasor plot generated for a specific frequency. This transformation is illustrated in Figure 3-3, where a representative maFLIM image is mapped to its corresponding phasor plot for an arbitrary frequency. Region-1 and Region-2 marked on the maFLIM image represent the regions obtained after clustering. The pixels of each region are mapped into a 2D-histogram distribution on the phasor plot, as shown in Figure 3-3.

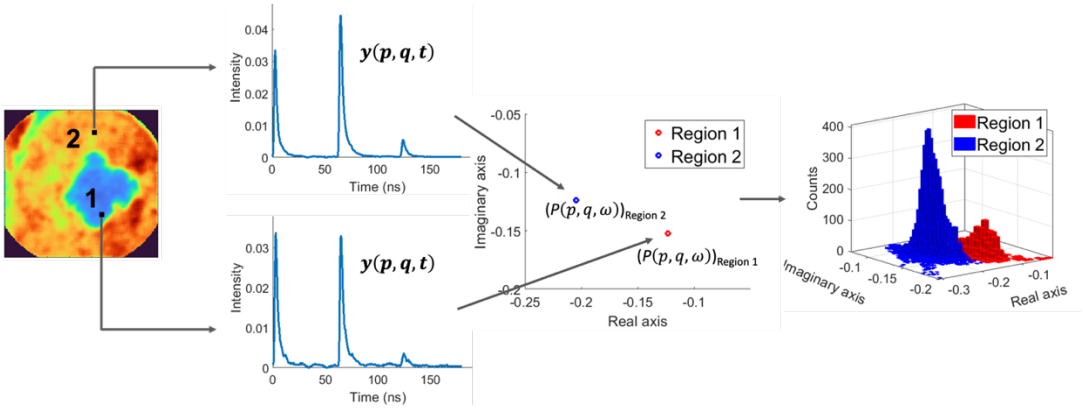



**Figure 3-3** Transition of a sample maFLIM image to the corresponding 2D histogram distribution on the phasor plot. Figure also shows the transformation of pixels from both regions 1 and 2 on the maFLIM image into corresponding points on the phasor plot computed at an arbitrary frequency component. Reprinted with permission from P. Vasanthakumari et al, 2022.

From the phasor representations of the maFLIM images at specific frequency components, $\omega = 2\pi f$, the following features were computed as follows. First, a bivariate Gaussian function was fitted to the phasor distribution of each region (Region-1, Region-2) of a given maFLIM image: $f_1 = N(\mu_1, \Sigma_1), f_2 = N(\mu_2, \Sigma_2)$ (Figure 3-4). The '*distance*' between the phasor distributions of the two regions was then estimated as the magnitude of the difference of the distribution means: $d = |\mu_1 - \mu_2|$ (Figure 3-4(a)). The determinant of the covariance matrix $|\Sigma|$ from the fitted Gaussian distribution provides a measure of the '*spread*' of the distribution. The difference in spread of the phasor distributions of the two regions was thus estimated as: $\Delta\Sigma = |\Sigma_1 - \Sigma_2|$ (Figure 3-4(b)). The '*angl*e' $\theta$ between major axes of the phasor distributions of the two regions was estimated as the acute angle between the eigenvectors of maximum variance of the multivariate gaussian distributions (Figure 3-4(b)). Finally, the '*symmetry*' of the Gaussian distribution can be quantified as the ratio of the variances along the orthogonal directions, $s = \frac{\sigma_p^2}{\sigma_q^2}$ (Figure 3-4(c)). The difference in symmetry of the phasor distributions of the two regions was thus estimated as: $\Delta s = |s_1 - s_2|$.

The fluorescence frequency response $H(p, q, \omega)$ is bandlimited with a bandwidth of ~60 MHz. To cover the bandwidth of $H(p, q, \omega)$, only the first nine frequency components of $H(p, q, \omega)$ were selected, corresponding to the frequencies 5.6 MHz, 11.2 MHz, 16.8 MHz, 22.4 MHz, 28 MHz, 33.6 MHz, 39.2 MHz, 44.8 MHz and 50.4 MHz. These frequency values are the first nine harmonics of the signal Fourier spectrum, which was calculated at a frequency resolution of 5.6 MHz (sampling frequency/#samples = 2.5

GHz/(3x149)). For each of these frequency components, the four phasor features were computed, resulting in a total of 36 phasor features.

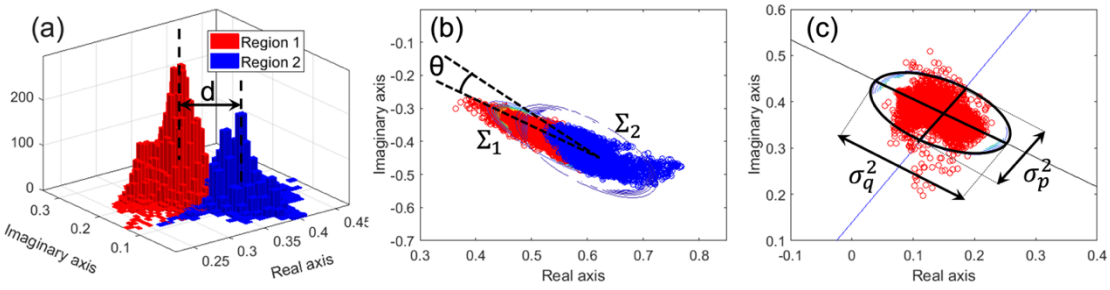



**Figure 3-4** (a) 2D histogram phasor distributions from the pixels corresponding to the two regions in an maFLIM image. The distance between the distributions is indicated by 'd'. (b) Phasor distribution scatter plots with bivariate gaussian fits on regions 1 and 2. The covariance matrices $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$ give a measure of spread of the two regions and $\boldsymbol{\theta}$ represents the angle between their major axes (c) Phasor distribution scatter plot showing the variances $\boldsymbol{\sigma_p^2}$ and $\boldsymbol{\sigma_q^2}$ along the major axes. The ratio of the variances indicates the symmetry of the distribution. Reprinted with permission from P. Vasanthakumari et al, 2022.

### 3.2.5. **Feature selection**

For maFLIM dermoscopy based automated classification of benign vs. malignant skin lesions, a simple QDA model was explored with the global features described in the section 3.2.3. Three different pools of global features were considered: 1) intensity ($n_{features}$=6), 2) bi-exponential ($n_{features}$=12), and 3) phasor ($n_{features}$=36) features, where $n_{features}$ is the number of features in each feature pool. In addition to these individual feature pools, different combinations of the feature pools were also considered: phasor ∪ bi-exponential ($n_{features}$=48), phasor ∪ intensity ($n_{features}$=42), intensity ∪ bi-exponential ($n_{features}$=18), phasor ∪ bi-exponential ∪ intensity ($n_{features}$=54). Feature selection using

sequential forward search (SFS)[87] was performed on each feature pool independently using a leave-one-patient-out cross-validation (LOPO-CV) strategy, whereby the data of one patient is left out at each fold (Figure 3-5(a)). This assures that the left-out patient data is independent of the feature selection and model training. Unlike exhaustive search where every possible combination of features is examined, SFS is computationally simpler and provides an efficient strategy to investigate the importance of the available features. The steps involved in the SFS feature selection process is shown in Figure 3-5(b). The feature subset yielding the highest area under the curve (AUC) value in Receiver Operator Characteristics (ROC) analysis was selected ($f_{SFS}$) at each iteration of the LOPO-CV. Subsequently, the QDA classifier is retrained using $f_{SFS}$ at each LOPO-CV iteration and then tested using the left-out patient data. A threshold of 0.5 on the generated posterior probability classifies the left-out patient data into either benign or malignant. Upon accomplishing all folds, each patient data had been used as test, and a confusion matrix was constructed to estimate the classification model performance. The number of features selected during SFS ($n_{SFS}$) is kept constant in each fold of the LOPO-CV. Therefore, to determine the number of features that can produce the best classifier, the experiments are repeated for different values of $n_{SFS}$, varying from 1 to 7 (1 to 6 for intensity feature pool). The maximum number of features in the feature set is chosen based on Hua et al.[88], where the optimal number of features for feature sets with some degree of correlation is $\sqrt{n_{data}}$, where $n_{data}$ is the number of data points. In this case, the maximum number of features is chosen as the closest and lower integer value to $\sqrt{n_{lesions}} = \sqrt{60} = 7$. The value of $n_{SFS}$ that produced the best F-score computed from the final confusion matrix was

then chosen as the number of features in the final feature set ($n_{selected}$). This is because F-score gives a combined estimate of both the sensitivity and specificity. However, in cases when two values of $n_{SFS}$ produce the same F-score, a higher sensitivity is given more preference. This is because it is critical to ensure that the malignant lesions are correctly classified to provide adequate and timely treatment to the patients.

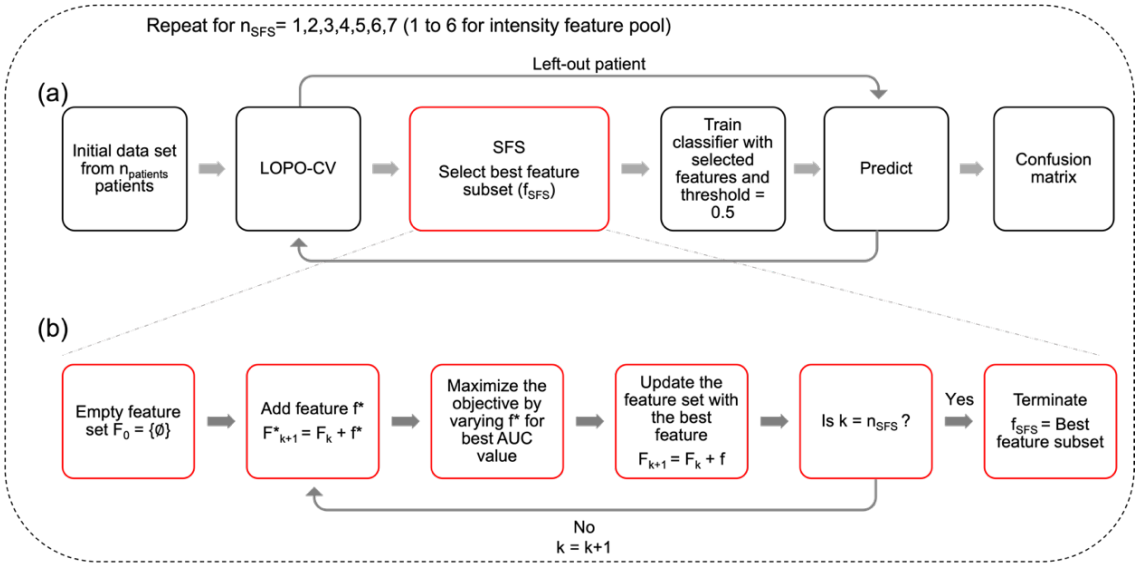



**Figure 3-5** Flow diagram showing (a) feature selection process using leave-one-patient-out cross-validation along with sequential forward search algorithm, and (b) Detailed steps involved in the sequential forward search algorithm. The number of features selected, $n_{SFS}$ is varied from 1 to 7 for all feature pools and 1 to 6 for intensity feature pool. Reprinted with permission from P. Vasanthakumari et al, 2022.
SFS - Sequential forward search
LOPO-CV – Leave-one-patient-out cross-validation
AUC – Area under the curve

Since the LOPO-CV iterates $N_{patient}$ times, where $N_{patient}$ is the number of patient data, the features $f_{SFS}$ selected in each iteration for a particular value of $n_{SFS}$ depends on the ($N_{patient}$-1) patient data that are not left out by that iteration. Thus, there can be some

variation in the features that are picked out in each iteration of the LOPO-CV. Therefore, the selection frequency, which is defined as the number of times each feature becomes part of the $f_{SFS}$ during all iterations of the LOPO-CV is noted. This allowed to identify the most frequent (thus most relevant) features ($f_{selected}$) from each feature pool. It is to be noted that the number of features in each $f_{selected}$ is denoted as $n_{selected}$.

### 3.2.6. **Classification of skin cancer lesions using selected features**

QDA classifiers trained on the best features, $f_{selected}$, from the three feature pools as explained in the section 3.2.4 were also combined in an ensemble fashion as shown in Figure 3.6(a). Separate classifiers are trained on 'k' feature pools with a LOPO-CV loop. If all the feature pools are used, k=3, otherwise, k=2. The left-out patient data from the LOPO-CV is tested on each of the individual classifiers, generating a set of posterior probabilities, $P_{pool\text{-}1}$, $P_{pool\text{-}2}$, .., $P_{pool\text{-}k}$, corresponding to classifiers trained on each feature pool. Here, Pool-k is either phasor, intensity, or bi-exponential feature pool. Subsequently, a weighted average of the posterior probabilities is computed as:

$$P = w_1 P_{pool-1} + w_2 P_{pool-2} \ldots + w_k P_{pool-k} \tag{3-18}$$

$$w_1 + w_2 + \cdots + w_k = 1 \tag{3-19}$$

where, $w_1$, $w_2$, .., $w_k$ are the weights on the posterior probabilities generated from each feature pool and the sum of weights equals one. A threshold of 0.5 on the weighted average of the posterior probabilities assigns a label for the left-out patient data. Figure 3-6(b) shows the process of weight optimization for the ensemble classifiers.

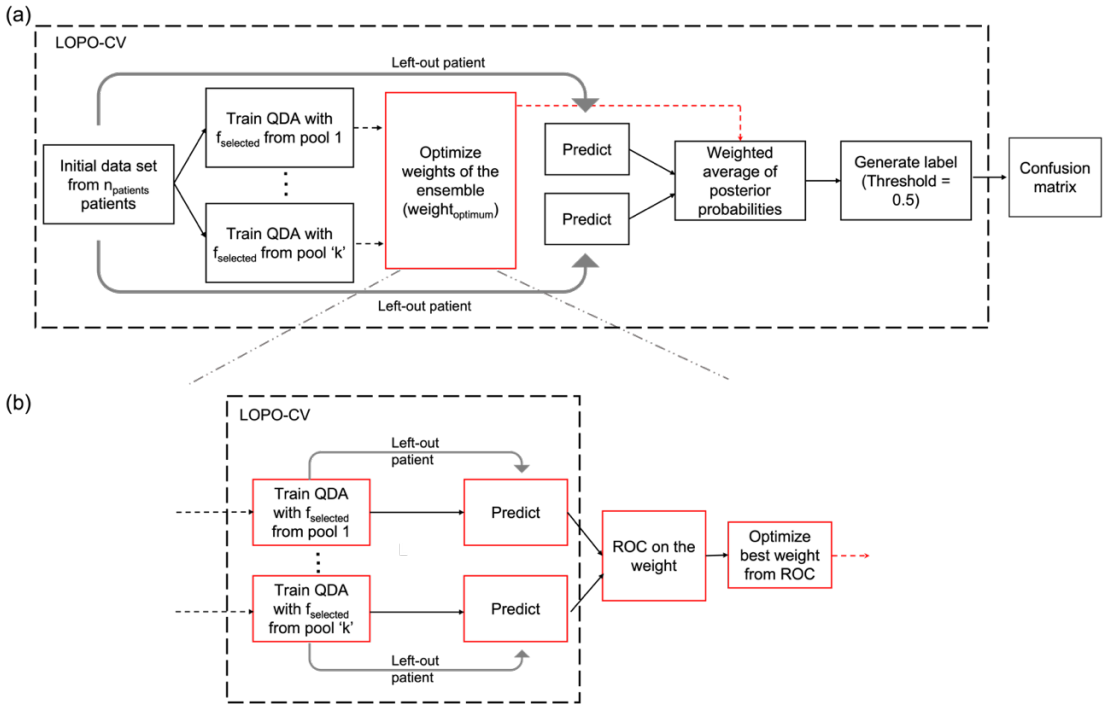



**Figure 3-6** (a) Schematic of classification of skin lesions. The posterior probabilities from the individual classifiers are combined in an ensemble fashion. (b) Weight optimization for the ensemble classifier. The optimum weight is selected from the ROC curve. Reprinted with permission from P. Vasanthakumari et al, 2022.
LOPO-CV – leave-one-patient-out cross-validation
QDA – Quadratic discriminant analysis
ROC – Receiver operator characteristics

Weight optimization is performed within another LOPO-CV loop. The classifiers are trained on the data from $(n_{patients} - 2)$ patients and the left-out patient data are predicted to generate corresponding posterior probabilities. These posterior probabilities are combined using a weighted average. The weights are varied from 0 to 1 in steps of 0.1. The corresponding sensitivities and specificities are obtained using a threshold of 0.5 on the weighted-average probability. An ROC curve is constructed on these weights, and the weight closest to the ideal point ([0,1]) is selected.

## 3.3. **Results**

### 3.3.1. **maFLIM dermoscopy clinical imaging of skin lesions**

A total of 30 patients undergoing tissue biopsy examination of skin lesions for cancer diagnosis participated in this study. maFLIM dermoscopy images of multiple skin lesions were acquired on each patient. The distribution of patients ($n_{patients}$=30) and lesions ($n_{lesions}$=60) imaged in this study showing benign and malignant conditions is provided in Table 3-3. Benign lesions included solar lentigo and pSK, while malignant lesions included pigmented superficial BCC, pigmented nodular BCC and melanoma.

**Table 3-3** Distribution of imaged benign and malignant lesions. Reprinted with permission from P. Vasanthakumari et al, 2022.

| | **Type** | **No. patients** | **No. lesions** |
|---|---|---|---|
| Benign | Solar lentigo | 2 | 10 |
| | Pigmented seborrheic keratosis | 15 | 31 |
| Malignant | Pigmented superficial BCC | 2 | 6 |
| | Pigmented nodular BCC | 5 | 5 |
| | Melanoma | 6 | 8 |

Figure 3-7(a) shows a handheld maFLIM dermoscope imaging the forearm of a patient. The clinical photograph of a sample melanoma skin lesion is shown in Figure 3-7(b)**,** and its corresponding maFLIM feature maps are shown in Figure 3-7(c). The scales of the feature maps across the three spectral wavelengths are kept the same for comparison purposes. Most of the feature maps (including $\alpha_{fast,390}$ , $\alpha_{fast,452}$ , $\tau_{fast,452}$ , $\tau_{fast,500}$, $\tau_{slow,390}$, $\tau_{slow,452}$, $\tau_{slow,500}$ , $\tau_{avg,390}$ , $\tau_{avg,452}$ $\tau_{avg,500}$ , $I_{n,452}$ , $I_{n,500}$ , $I_{452,n}/I_{500,n}$ and $I_{390,n}/I_{500,n}$) clearly shows two distinguishable regions – the center and

the surrounding regions. The cluster masks for this sample image are also shown in the last row of Figure 3-7(c). The maps showing the two clustered regions – Region-1 and Region-2 are plotted using the normalized integrated intensities ($I_{Integrated}$) from the three spectral emission channels.

The relative contributions of fast lifetime ($\alpha_{fast}$) for the spectral channels 390 nm and 452 nm exhibit a higher value in the central region compared to the surrounding regions. Fast lifetime ($\tau_{fast}$) values for the spectral channels 452 nm and 500 nm are lower in the central region than the surrounding. Slow lifetime ($\tau_{slow}$) values for all the three spectral channels are higher in the center than the surrounding regions. The average lifetimes ($\tau_{avg}$) of the pixels from the central region are higher than the surrounding parts. The relative intensity ($I_n$) values for spectral channel 452 nm is lower in the center compared to the surrounding, while that for the spectral channel 500 nm is higher in the center compared to the surrounding. The ratios of the intensities, $I_{452,n} / I_{500,n}$ and $I_{390,n}/I_{500,n}$ are lower in the central region compared to the surrounding regions.

### 3.3.2. **Feature selection**

As explained in the feature selection process in section 3.2.4, the multiple lesions from a single patient constitute the left-out data in each iteration; thus, the number of lesions that are tested in each iteration varies depending on the number of lesions that are imaged for the left-out patient. In this way, every patient becomes part of the testing, and a confusion matrix is generated after all LOPO-CV iterations are completed.

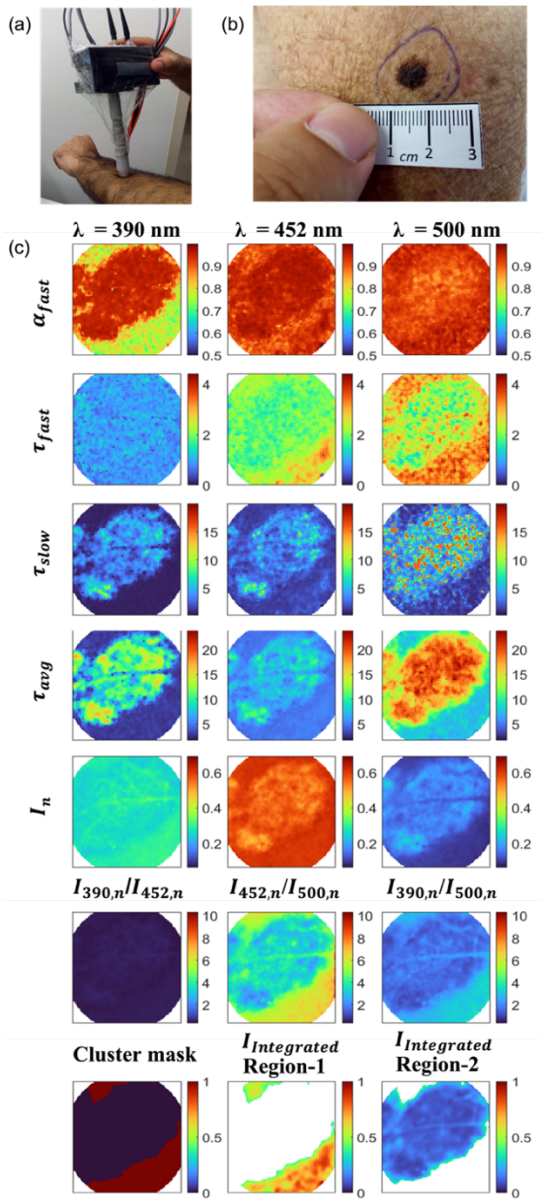



**Figure 3-7** (a) Hand-held maFLIM dermoscope imaging the forearm of a patient. (b) Clinical photograph of a melanoma lesion. (c)Time domain maFLIM feature maps of a melanoma lesion. The columns show the feature maps corresponding to the three emission channels. First row shows the weight of the fast decay. Second row shows the fast lifetime maps, while the third row shows the slow lifetime maps. Average lifetime maps are shown in the fourth row. Fifth row shows the integrated intensity maps of each spectral emission channel, and the ratio of the intensities are shown in the sixth row. The last row shows the cluster mask generated for the lesion and the integrated intensities from all the channels for the clustered regions 1 and 2. Reprinted with permission from P. Vasanthakumari et al, 2022.

As mentioned in section 3.2.4, the number of features $n_{SFS}$ was varied from 1 to 6 for intensity feature pool, and 1 to 7 for phasor and biexponential feature pools. The optimum number of features $n_{selected}$ is chosen based on the highest F-score. Table 3-4 shows values of accuracy, sensitivity (Sn), specificity (Sp) and F-score obtained when classifying benign and malignant skin lesions trained individually on the three feature pools (phasor, bi-exponential and intensity). The table shows the results for $n_{selected}$ number of features in each feature pool. QDA models trained on five bi-exponential features yielded the best performance with 75% accuracy, 84.21% sensitivity, 70.73% specificity, and 68.09% F-score. QDA models trained on six phasor features yielded the next best performance with 76.67% accuracy, 68.42% sensitivity, 80.49% specificity, and 65% F-score. QDA models trained on the intensity feature pool resulted in 84.21% sensitivity and only 31.71% specificity. The results from the combined feature pools show poor performance with low sensitivities and specificities. This is because the features selected by the SFS algorithm in the earlier iterations may not be the best when combined with those selected in the later iterations. From these results it can be inferred that the five bi-exponential features and the six phasor features have potential in classifying benign and malignant skin lesions. The confusion matrices of the classifiers are also shown in the Table 3-4.

**Table 3-4** Performance metrics and confusion matrices obtained during feature selection with phasor, bi-exponential and intensity feature sets. Reprinted with permission from P. Vasanthakumari et al, 2022.

| Feature pool (Total no. of features) | $n_{Selected}$ | Accuracy (%) | Sn (%) | Sp (%) | F-score (%) | Confusion matrices | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Predicted Benign | Predicted Malignant |
| Phasor (36) | 6 | 76.67 | 68.42 | 80.49 | 65.00 | True | Benign | 33 | 8 |
| | | | | | | | Malignant | 6 | 13 |
| Bi-exponential (12) | 5 | 75.00 | 84.21 | 70.73 | 68.09 | | Benign | 29 | 12 |
| | | | | | | | Malignant | 3 | 16 |
| Intensity (6) | 1 | 48.33 | 84.21 | 31.71 | 50.79 | | Benign | 13 | 28 |
| | | | | | | | Malignant | 3 | 16 |
| Phasor ∪ Bi-exponential (48) | 4 | 56.67 | 63.17 | 56.10 | 48.98 | | Benign | 23 | 18 |
| | | | | | | | Malignant | 7 | 12 |
| Phasor ∪ Intensity (42) | 6 | 53.33 | 63.16 | 48.79 | 46.15 | | Benign | 20 | 21 |
| | | | | | | | Malignant | 7 | 12 |
| Bi-exponential ∪ Intensity (18) | 7 | 63.33 | 63.16 | 63.41 | 52.17 | | Benign | 26 | 15 |
| | | | | | | | Malignant | 7 | 12 |
| Phasor ∪ Bi-exponential ∪ Intensity (54) | 6 | 61.67 | 47.37 | 68.29 | 43.90 | | Benign | 28 | 13 |
| | | | | | | | Malignant | 10 | 9 |

### 3.3.3. **Frequency plot of features**

To identify the important features in each feature pool, the number of times each feature becomes a part of $f_{SFS}$ during all iterations of the LOPO-CV is noted. If a feature is selected at least 50% times during all the iterations, it will be considered an important feature and added to the $f_{selected}$ of that feature pool. Table 3-5 lists the important features from each feature pool and their selection frequencies during LOPO-CV. '*Symmetry*' at frequencies 16.8 MHz, 33.6 MHz, 39.2 MHz and 50.4 MHz, '*spread*' at frequency 33.6

MHz, and '*distance*' at 50.4 MHz are the six most important features in the phasor feature pool. The relevant features from the bi-exponential feature pool are $\alpha_{fast,390}$, $\tau_{slow,390}$, $\alpha_{fast,452}$, $\tau_{fast,452}$ and $\alpha_{fast,500}$. In the intensity feature pool, $I_{452,n}$ is the most important feature. Thus, we can summarize the selected features from each feature pool as:

$$(f_{selected})_{phasor} = [symmetry_{16.8\,MHz}, symmetry_{33.6\,MHz}, symmetry_{39.2\,MHz},$$
$$symmetry_{50.4\,MHz}, spread_{33.6\,MHz}, distance_{50.4\,MHz}]$$

$$(f_{selected})_{bi-exponential} = [\alpha_{fast,390}, \tau_{slow,390}, \alpha_{fast,452}, \tau_{fast,452}, \alpha_{fast,500}]$$

$$(f_{selected})_{intensity} = [I_{452,n}]$$

### 3.3.4. **Classification of skin cancer lesions using selected features**

The methodology in classifying skin lesions using the selected features are explained in section 3.2.5. As shown in Figure 3-6, 'k' QDA classifiers trained separately on the $f_{selected}$ features from 'k' feature pools are combined in an ensemble fashion. Four different combinations of feature pools, are used for constructing ensemble classifiers, as shown in Table 3-6. A weighted average of the posterior probabilities generated from the 'k' QDA classifiers is calculated to predict the label of the left-out patient data. Since the weights are optimized within a LOPO-CV loop, there can be some variation in the weights selected during the $n_{patients}$ iterations. Figure 3-8 shows the histograms of the weights obtained during all the iterations for the ensemble classifiers using the four feature pool combinations. Since the sum of weights of 'k' feature pools is one, it is sufficient to show the weights of (k-1) feature pools. It can be seen from Table 3-6 that the ensemble combination of QDA classifiers trained on phasor and bi-exponential features as well as the ensemble combination of all the three feature pools, produced the best performance

with 88.33% accuracy, 84.21% sensitivity, 90.24% specificity, and F-score of 82.05%. The next best performance is obtained by the ensemble combination of phasor and intensity feature pools with an accuracy of 86.67%, sensitivity of 78.95, specificity of 90.4% and an F-score of 78.95%. Ensemble combination of intensity and bi-exponential features produced an accuracy of 71.67%, sensitivity of 84.21%, specificity of 65.85% and F-score of 65.31%.

**Table 3-5** Summary of important features selected from each feature pool along with their ranks. Reprinted with permission from P. Vasanthakumari et al, 2022.

| Feature pool | $f_{selected}$ | Selection frequency percentage (%) |
|---|---|---|
| Phasor | Symmetry at 39.2 MHz | 93.3 |
| | Spread at 33.6 MHz | 90.0 |
| | Symmetry at 16.8 MHz | 90.0 |
| | Symmetry at 50.4 MHz | 86.7 |
| | Symmetry at 33.6MHz | 56.7 |
| | Distance at 50.4 MHz | 50.0 |
| Bi-exponential | $\tau_{fast,452}$ | 96.7 |
| | $\tau_{slow,390}$ | 93.3 |
| | $\alpha_{fast,452}$ | 80.0 |
| | $\alpha_{fast,390}$ | 76.7 |
| | $\alpha_{fast,500}$ | 50.0 |
| Intensity | $I_{452,n}$ | 96.7 |

Table 3-6 also shows the confusion matrices for all the classifiers. While analyzing the weights on the feature pools during the ensemble combination, it can be seen in Figures 3-8(a) and 3-8(b) that intensity features have a higher weightage when combined with phasor features. When phasor and bi-exponential features are combined, both the feature pools have similar weightage as seen in Figure 3-8(c). Similarly, when bi-exponential and intensity feature pools are combined, both the pools have comparable weightage. When

all the feature pools are combined, it can be seen from Figure 3-8(d) that the phasor and bi-exponential features pools have similar weight distribution in the range [0.1, 0.5]. The weights on the intensity feature pool are widely dispersed in the range [0, 0.8].

**Table 3-6** Performance metrics of ensemble classifiers trained with multiple combinations of feature pools. Reprinted with permission from P. Vasanthakumari et al, 2022.

| Feature sets for ensemble | Accuracy (%) | Sn (%) | Sp (%) | F-score (%) | Confusion matrices | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Predicted | |
| | | | | | | | Benign | Malignant |
| $(f_{selected})_{bi\text{-}exponential} + (f_{selected})_{intensity}$ | 71.67 | 84.21 | 65.85 | 65.31 | True | Benign | 27 | 14 |
| | | | | | | Malignant | 3 | 16 |
| $(f_{selected})_{phasor} + (f_{selected})_{bi\text{-}exponential}$ | 88.33 | 84.21 | 90.24 | 82.05 | | Benign | 37 | 4 |
| | | | | | | Malignant | 3 | 16 |
| $(f_{selected})_{phasor} + (f_{selected})_{intensity}$ | 86.67 | 78.95 | 90.24 | 78.95 | | Benign | 37 | 4 |
| | | | | | | Malignant | 4 | 15 |
| $(f_{selected})_{phasor} + (f_{selected})_{bi\text{-}exponential} + (f_{selected})_{intensity}$ | 88.33 | 84.21 | 90.24 | 82.05 | | Benign | 37 | 4 |
| | | | | | | Malignant | 3 | 16 |

## 3.4. **Discussion**

In this study, clinical widefield autofluorescence imaging of benign and malignant pigmented skin lesions was successfully performed in 30 patients using a recently developed multispectral autofluorescence lifetime imaging (maFLIM) dermoscope[82]. The resulting maFLIM images from 60 pigmented lesions enabled exploring novel steady-state (intensity) and time-resolved (bi-exponential, phasor) autofluorescence global features. Results based on rigorous cross-validation methods demonstrate that simple machine-learning classification models (QDA) based on selected time-resolved

autofluorescence global features have the potential to provide discrimination of malignant from benign pigmented skin lesions.

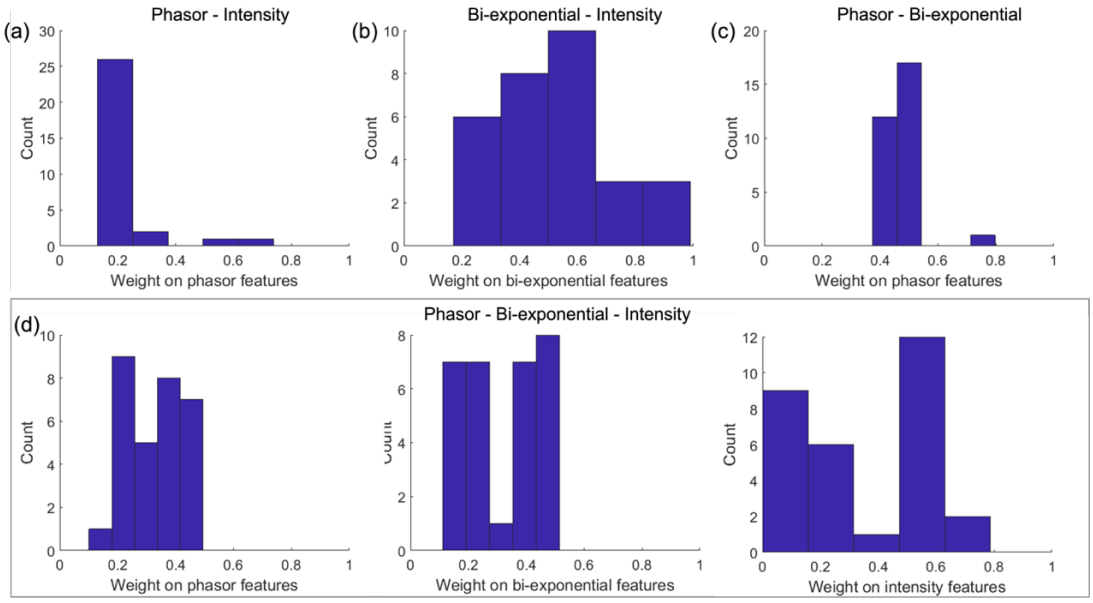



**Figure 3-8** Histogram of weights on one of the feature pools, when combined in an ensemble fashion for (a) Phasor-intensity (b) Bi-exponential-intensity, (c) Phasor-bi-exponential, and (d) Phasor-bi-exponential-intensity feature pools. Reprinted with permission from P. Vasanthakumari et al, 2022.

To the best of our knowledge, only one published work has reported the use of machine learning models based on autofluorescence lifetime imaging features for the classification of pigmented skin lesions[75]. In that study, however, only skin melanoma lesions were imaged, and the classification task was restricted to discriminate early-stage from advanced-stage skin melanoma. In contrast, a more comprehensive set of pigmented skin lesions were imaged in this work (two benign and three malignant lesion categories). Moreover, the classification task focused on discriminating malignant from benign

pigmented lesions, which might be clinically more relevant for early detection of skin cancer.

In multidimensional imaging data, such as in maFLIM, image features can be extracted at the pixel or the image level. We have recently explored pixel-level maFLIM features for the classification of oral dysplasia and early-stage cancer[89]. Pixel-level features, however, require the labeling of each pixel which is generally impractical. In this work, the maFLIM data was labeled at the lesion level based on the histopathology diagnosis obtained from the lesion biopsy samples; therefore, an image-level global feature extraction strategy was preferred. As illustrated in Figure. 3-7, two regions were frequently observed in the maFLIM images, corresponding to pixels either within or outside the lesion extension. In an attempt to reduce inter-patient variability, the explored relative features were defined in terms of difference in autofluorescence properties between the two regions identified in each lesion maFLIM image. This strategy of using global (image-level) and relative features can find applications in many other classification tasks based on optical imaging data.

The performance of machine-learning classification models needs to be carefully estimated when trained on limited sample size. To minimize overfitting and avoid overoptimistic performance estimations, a rigorous strategy was adopted for feature selection, model training, and performance estimation. First, the maximum number of features allowed (seven) was limited based on the sample size[88]. Second, a simple non-linear classification model (QDA) was adopted. Third, cross-validation was applied at the patient-level (LOPO-CV) to ensure that data from the same patient is not used for both

training and validation. Fourth, feature selection was performed together with model training to make sure that the validation data is not used during neither feature selection nor model training. It should be noted that at each fold of the cross-validation strategy, a different classification model (with different selected features and model parameters) is applied to the validation set. Thus, although a single optimal model is not necessarily defined, this approach still enables identifying relevant features and providing unbiased classification performance estimation.

Classification performance was dependent of the feature pool used in the model. The most frequently selected intensity feature was $I_{452,n}$ which is associated to NADH fluorescence contribution. Although classification models using intensity features showed good sensitivity (~84%), their specificity was poor (~31%). The most frequently selected bi-exponential features were associated to NADH ($\tau_{fast,452}$, $\alpha_{fast,452}$) and collagen ($\tau_{slow,390}$, $\alpha_{fast,390}$) fluorescence temporal dynamics. Classification models using bi-exponential features showed similar sensitivity (~84%) than those with intensity features, but significantly higher specificity (~70%). These results indicate that time-resolved properties of pigmented skin lesion autofluorescence could represent novel biomarkers of skin cancer.

The phasor representation of fluorescence lifetime imaging data analysis is a non-iterative, model-free, fast approach to visualize the lifetime components (and their distributions) of the fluorescence emission of a sample[90]–[92]. In this work, a modified version of this method was applied to the clinical multispectral maFLIM data, and a novel set of global image features were extracted from the corresponding phasor representation

of the two regions present in each lesion maFLIM image (Figure. 3-4). Unlike bi-exponential maFLIM features, the phasor features explored cannot be directly interpreted in terms of the skin autofluorescent constituents. Nevertheless, classification models using phasor features showed superior specificity (~80%) than those using bi-exponential features, suggesting that these two different pools of time-resolved maFLIM features might be complementary.

Feature selection starting with combinations of feature pools were also explored, although the resulting models showed lower classification performance overall (Table 3-4). On the other hand, ensemble classifiers based on models using the most frequently selected features of each pool outperformed any other models (Table 3-5). In particular, the ensemble classifier combining the models based on the most frequently selected bi-exponential and phasor features resulted in the best performance overall (~84% sensitivity, ~90% specificity). Moreover, the optimum weights identified for these ensemble models indicate that both the phasor and bi-exponential features contribute equally to the weighted probability (Figure. 3-8(c)). These results further indicate that these two maFLIM feature pools might be complementary, as bi-exponential features seem to contribute to higher sensitivity, while phasor features to higher specificity.

It is plausible that more complex machine learning classification models, such as neural networks, might provide superior performance. Given the limited data available for training and validation, however, a simple QDA model was selected, as it still enables non-linear decision boundaries while reducing the chances of overfitting. In addition, by choosing a simple classification method, it is possible to analyze other aspects of

classification model, such as the feature pools, number of features, and ensemble combinations of the feature pools.

*Study limitations*

Although this preliminary clinical study demonstrates the potential of maFLIM derived autofluorescence features to discriminate malignant from benign pigmented skin lesions, a number of limitations are recognized. First, the database of maFLIM images is limited in both the type of benign and malignant skin conditions, and the number of samples per condition. A more comprehensive and larger database will be needed to fully develop accurate enough classification methods for skin lesion discrimination, and to rigorously quantify their performance in prospective studies. Second, the lack of histopathology-based assessment of the maFLIM imaging data at the pixel-level prevented to specifically quantify the capabilities of maFLIM dermoscopy as a tool for not only detecting malignant skin lesions, but also determining their true extension and margins. Third, the current maFLIM dermoscopy system provides nonspecific excitation and spectral detection of skin autofluorescence component emission. Finally, the current implementation of the machine-learning classification models does not allow for real-time processing of maFLIM data. Ongoing research efforts aiming to overcome these limitations include collecting maFLIM dermoscopy images from a plurality of nonpigmented and pigmented skin lesions from patients of various skin tones, performing accurate pixel-level registration between the lesion maFLIM imaging data and histopathology tissue sections, developing improved maFLIM dermoscope systems with multiwavelength excitation and narrow-band emission detection capabilities, and

implementing optimized CAD using Field Programmable Gate Arrays (FPGA) and Graphics Processing Units (GPU) technologies for real-time maFLIM data processing, pixel-level classification, and tissue mapping visualization.

*Clinical perspective*

The incidence of skin cancer including melanoma continues to increase yet most providers are forced to rely on their own visual recognition skills and experience to identify concerning lesions. Additionally, many patients do not have access to a trained dermatologist which can place them in a potentially precarious situation since early detection of skin cancer leads to better survival rates. The importance of early detection cannot be understated as it not only saves lives but reduces the invasiveness of the treatment patients undergo and conserves precious medical resources leading to quality cost-conscious care. A noninvasive, label-free, fast, accurate, and objective tool capable of discriminating most common malignant from benign skin lesions would improve the clinical management of patients. Currently, there is no objective device providers can use to independently identify cancerous skin lesions. In the hand of primary care physicians, such a tool will enable the early identification of patients in need of referral to a dermatologist. Additionally, dermatologists could use such tool to identify lesions in need of a biopsy, thus reducing the rate of unnecessary biopsies and adverse events such as pain, infection, and scarring. Such a tool would also assist with monitoring cancer recurrence without the need of regular and frequent biopsies. This work demonstrates that maFLIM dermoscopy aided by machine-learning models could potentially have the

capabilities of such tool, thus impacting the clinical management of skin cancer patients for the better.

## 3.5. **Conclusions**

The results of this study demonstrate the capabilities of maFLIM dermoscopy to clinically image a plurality of autofluorescence biomarkers of malignant skin pigmented lesions. Moreover, some of these autofluorescence biomarkers were identified as promising for malignant lesion identification, particularly those quantifying the time-resolved fluorescence characteristics of skin lesions. In addition, these relevant autofluorescence biomarkers were successfully used as features in machine-learning models trained to discriminate malignant from benign pigmented skin lesions with promising accuracy (~84% sensitivity, ~90% specificity). Further developments in maFLIM instrumentation and image analysis methods could result in novel clinical tools for noninvasive, label-free, accurate, and objective *in situ* detection of malignant from benign skin lesions, with the potential to impact the clinical management of skin cancer patients.

# 4. DEEP-LEARNING ASSISTED DISCRIMINATION OF CANCEROUS FROM BENIGN PIGMENTED SKIN LESIONS BASED ON MULTISPECTRAL AUTOFLUORESCENCE LIFETIME IMAGING DERMOSCOPY*

## 4.1. **Introduction**

In recent years several works have been published using deep learning techniques for skin cancer detection. Deep learning can be described as neural networks with a deep architecture consisting of many hidden layers – essentially consisting of alternate linear and non-linear transformations of the inputs from preceding layers in the network. Deep learning has the capability to learn features automatically during the training process. In addition, they can make use of information which may otherwise be missed during manual feature extraction. D. Adla et al.[63] used a CNN based method for segmentation and classification of dermoscopic skin lesion images. J Amin et al. [71] extracted features using pre-trained AlexNet and VGG16 deep learning architectures. The features were selected using principal component analysis (PCA) and then subsequently used for classification through traditional machine learning algorithms like SVM, KNN and discriminant analysis. Harangi et al.[67] used an ensemble of different CNN classifiers, while Lopez et al.[68] used transfer learning with pre-trained VGGNet CNN architecture. J.A. Jaleel et al.[12] used 2D wavelet transform to extract features from dermoscopic

images, which were then classified to either cancerous or non-cancerous using neural networks. Majtner et al.[69] combined CNN with SVM classifier using hand crafted RSurf features and local binary patterns to classify melanomas from other benign skin lesions. Dorj et al.[93] extracted features using a pre-trained AlexNet to classify four types of skin cancer using error-correcting output coding SVM. L Yu et al.[94] implemented a fully convolutional residual network (FCRN) for segmenting dermoscopic images for melanoma detection. Lee et al.[70] developed the WonDerm pipeline that segments the skin cancer dermoscopic images using neural network architectures and classifies it using an ensemble approach. J. Acosta et al.[72] utilized transfer learning with ResNet-50 architecture to classify melanoma and non-melanoma lesions using dermoscopic images. Srinivasu et al.[95] implemented a model in which a CNN based MobileNet v2 is used for classification and a Long Short-Term Memory (LSTM) network is used to maintain the state information of the features.

Interestingly, most of the previously published works on skin cancer classification using deep learning are based on standard white light dermoscopy. We use maFLIM images collected from three spectral emission bands to classify benign and malignant skin lesions. In Chapter 3, features were extracted from the maFLIM images after rigorous feature engineering to classify skin lesions using classical machine learning algorithms. This work aims to bypass the feature extraction step by using deep learning for classification. The fluorescence time decay information in the maFLIM images could be efficiently incorporated using deep-learning models typically used for time series applications such as natural language processing (NLP) or trajectory prediction. Such

models can not only learn from the current time step but also from the past time steps. This is especially useful for ma-FLIM images as such models can automatically learn the information contained in the fluorescence time decay without manually calculating the lifetimes and intensities. This idea is explored by designing an LSTM model, an extension of Recurrent Neural Networks (RNN), to classify benign and malignant pigmented skin lesions. The LSTM model is paired with fully connected dense layers and output prediction layers to generate a prediction probability for every pixel of the maFLIM image. In this way, prediction probability maps can be constructed for the entire imaged region, which allows to classify the type of the lesion and identify regions of malignancy.

## 4.2. **Methodology**

A schematic of the summary of the methodology is shown in Figure 4-1.

### 4.2.1. **Skin cancer database**

The pigmented skin cancer dataset used in this work is the same dataset used in Chapter 3. Therefore, the details of data acquisition are given in section 3.2.1.

### 4.2.2. **Data pre-processing**

The steps involved in pre-processing maFLIM data from skin cancer lesions are given in section 3.2.2 in Chapter 3. The pre-processed signal at each pixel location is represented as $y_{\lambda}(p,q,t)$ for the three spectral emission bands. Equations 3-2 and 3-3 from Chapter 3 show how to obtain the deconvolved signal, $h_{\lambda}(p,q,t)$ from $y_{\lambda}(p,q,t)$ using time domain deconvolution. The deconvolved signal at each spectral emission channel can be concatenated to form a single signal as:

$$h(p,q,t) = \sum_{n=0}^{2} h_{\lambda_{n+1}}(p,q,(t-M.n)) \quad (4\text{-}1)$$

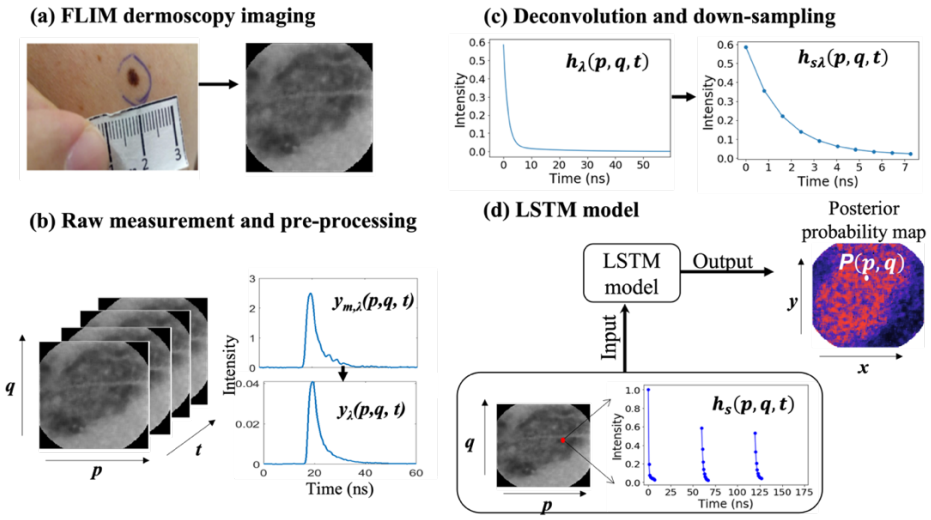



**Figure 4-1** Summary of methodology showing maFLIM image acquisition, pre-processing, deconvolution and signal down-sampling, and the LSTM model. (a) Shows a photograph of a sample melanoma lesion and its maFLIM image mapping the total integrated intensities. (b) Shows the dimensions of an maFLIM image at one spectral emission channel and a fluorescence decay signal at a sample pixel location before and after pre-processing. (c) Shows a deconvolved maFLIM signal before and after down-sampling. (d) Shows the inputs and outputs from the LSTM model.
FLIM – Fluorescence lifetime imaging
LSTM – Long short-term memory network

Where, $h(p,q,t)$ is the concatenated deconvolved maFLIM signal; $h_{\lambda_1}(p,q,t)$, $h_{\lambda_2}(p,q,t)$ and $h_{\lambda_3}(p,q,t)$ are the deconvolved maFLIM signals from the three spectral channels; M is the temporal spacing between the signals from the three channels.

The deconvolved signal at each pixel location is first down-sampled in time in such a way as to represent at least 80% of the fluorescence signal intensity[96]. This considerably reduces the computational overload in comparison to when the entire decay signal is input to the deep learning model. The length of the fluorescence decay signal at every pixel location per spectral emission channel, $h_\lambda(p,q,t)$ is 150 timesteps or 60 ns as

shown in Figure 4-2(a). It was found that the fluorescence intensity decays from 100% to atleast 20% of its peak value within the first 20 time steps, or in 7.6 ns. Therefore the first 20 time steps of the decay signal is downsampled into 10 equally spaced time steps with a down-sampling rate of 0.5. The vertical lines shown in Figure 4-2(a) indicates the sampling intervals. In this way, the signals from all the three spectral channels are down-sampled to form $h_{s\lambda}(p, q, t)$ as seen in Figure 4-2(b). Therefore,the dimension of the signal from each maFLIM image pixel is 3×10; ie., 10 time points corresponding to the 3 spectral emission channels.

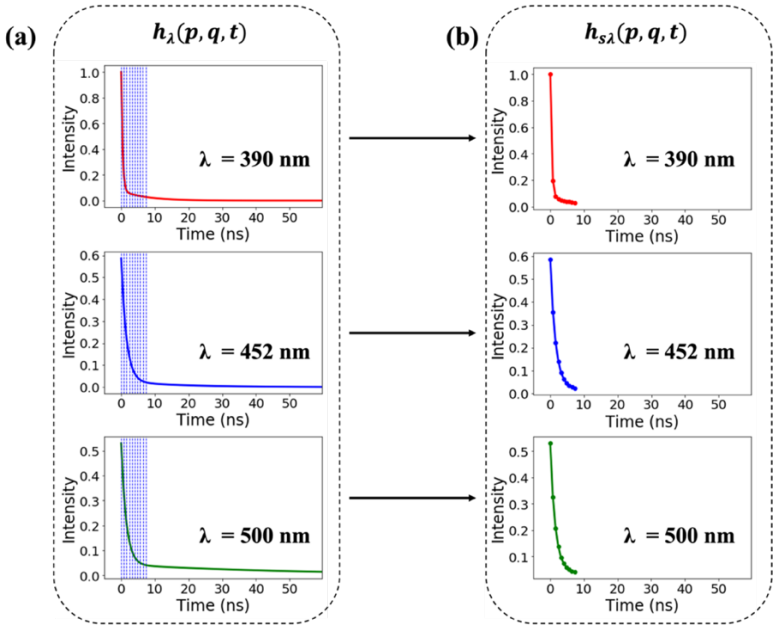



**Figure 4-2** Down-sampling of a sample maFLIM signal. (a) Deconvolved maFLIM signals from the three spectral emission bands. Vertical lines show the time points at which the signals are down sampled. (b) Down sampled signals from the three spectral emission bands. Reprinted with permission from [96].

### 4.2.3. **LSTM model**

Finding appropriate deep learning models for handling multi-dimensional data is a challenging task. CNN[8], [97] is the most popular deep learning technique used in handling image data either in greyscale or in RGB format. As the input dimensions increase, the complexity of the model also increases. For example, 3DCNN models[7], [98], [99] have been widely employed in classifying video data. RNNs[100], [101] and its variants are used for time series applications for predicting future time step values based on the values from previous time steps. Some of the previously published works[102]–[105] combine CNN and RNN based models for classifying multi-dimensional data by exploring spatial and temporal/spectral dimensions. As explained in the previous sections, maFLIM data contains spectral, temporal, and spatial dimensions. Section 4.2.2 illustrated how the spectral and temporal components can be combined to a signal of dimension 3×10 at each pixel location. Many previous works[20], [40], [77], [89], [106], [107] have demonstrated that the fluorescence time decay information such as the lifetimes and intensities are useful in discriminating cancerous lesions from healthy tissue. This indicates that the information contained in the time decay is highly valuable in classifying cancerous lesions. Therefore, this work investigates the fluorescence decay information at each pixel location to classify the lesion as either malignant or benign. The time decay can be efficiently studied by RNN based models, as they have a memory component that can retain information from past time steps. Therefore, a pixel-wise classifier based on LSTM networks is designed to classify the maFLIM image pixels as either benign or malignant.

Unlike traditional neural networks, RNNs have a memory capability that enables them to use the past information for predicting current and future time steps. RNNs use a feedback loop to propagate the information from the previous time steps. However, during backpropagation, the network weights of all the neurons that contributed to that output are updated. This includes the neurons from the current time step and the previous time steps. Repeated multiplications by a small number on the weights of the neurons cause their weights to decrease drastically. Therefore, the contribution of information propagated from much earlier time steps reduces drastically. This is called the problem of vanishing gradients[108] whereby the gradient of the loss function decays exponentially with time. Therefore, RNNs are not suitable for solving problems requiring long-term dependencies.

LSTM networks overcome the problem of vanishing gradients by using memory cells to remember information over more extended time periods. The schematic of an LSTM cell is shown in Figure 4-3. Here $x_t$ is the input to the cell, $s_{t-1}$ is the hidden state or the short-term memory, and $c_{t-1}$ is called the cell state or long-term memory. The memory cell is the straight line that runs through the top of the LSTM unit connecting the cell state to the new cell state, allowing easy flow of information along the network. The network carefully controls this information flow through gates or valves. Each LSTM unit consists of three gates or valves (input, forget, and output) denoted by '×' that regulate the information flowing through the network. $i_t$, $f_t$ and $o_t$ are the outputs of the input, forget and output gates, respectively. Every gate is controlled by a corresponding neural network with sigmoid activation. The output of the sigmoid activation acts like a probability which determines the amount of information flowing through that gate. For example, a gate

output value of 1 permits the free flow of information, and a value of 0, blocks any information to pass. The + operator allows the addition of input information to the memory cell if the corresponding input gate is open. The input and forget gates provide better control over the gradients and prevent the problems of vanishing gradients.

The outputs $i_t$, $f_t$ and $o_t$ can be expressed as:

$$i_t = \sigma(W_i \cdot [s_{t-1}, x_t] + b_i) \tag{4-2}$$

$$f_t = \sigma\big(W_f \cdot [s_{t-1}, x_t] + b_f\big) \tag{4-3}$$

$$o_t = \sigma(W_o \cdot [s_{t-1}, x_t] + b_o) \tag{4-4}$$

Where, $W_f$, $W_i$, and $W_o$ are the weights and $b_f$, $b_i$, and $b_o$ are the biases of the neural networks corresponding to the forget, input and output gates respectively. The cell state $c_t$ of the network is computed using the following equations:

$$\tilde{c}_t = tanh(W_c \cdot [s_{t-1}, x_t] + b_c) \tag{4-5}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{4-6}$$

Where $W_c$ and $b_c$ are the parameters of the neural network corresponding to the cell state. The output of the LSTM unit, $s_t$ is given by:

$$s_t = o_t * tanh(c_t) \tag{4-7}$$

Figure 4-4 shows the schematic of the LSTM model with the inputs, fully connected layer, and the output. In this dataset, the ground truth, or the label for an maFLIM image is obtained from the histopathological evaluation after biopsy resection. A label of 0 indicates that the image is benign, while a label of 1 indicates that the image is malignant. Since the label corresponding to every pixel location of an image is unknown, we assign the label of the entire maFLIM image to its pixels.

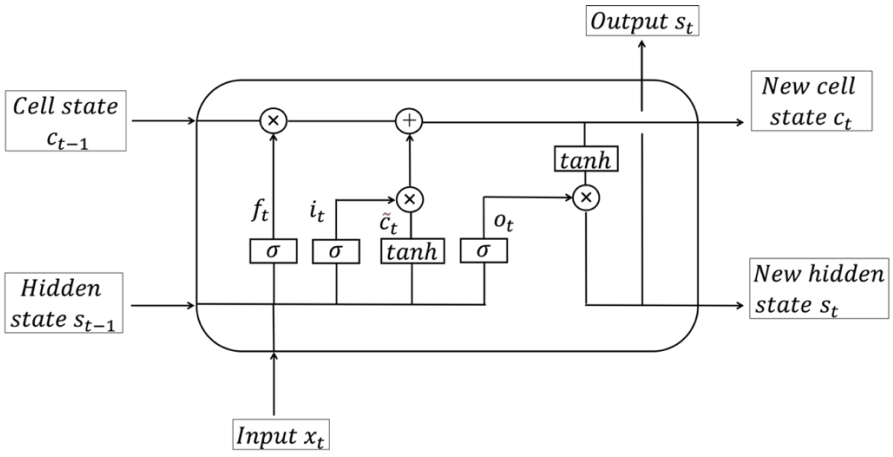



**Figure 4-3** Schematic of an LSTM cell showing the input, cell state, hidden state, output, and the three internal gates.

The input to the LSTM model used in this work is the down-sampled fluorescence decay signal, $h_{s\lambda}(p, q, t)$, from the three spectral channels at each pixel location. The number of LSTM units in the model is a hyperparameter that is tuned for optimum performance. The dimension of outputs from the LSTM layers, $s_1$, $s_2$, and $s_3$ depends on the number of LSTM units used. The output from the LSTM layer is fed to a fully connected dense layer followed by a single unit output prediction layer. The number of neurons in the fully connected dense layer is also a hyperparameter that needs to be tuned. A sigmoid activation at the prediction layer ensures that the output lies in the range [0,1] to resemble posterior probabilities. This model consisting of the LSTM layer, the fully connected layer and the output layer is optimized together by defining a loss function computed from the ground truth and the prediction probability. A binary cross-entropy loss function is used since it performed best for binary classification problems. The loss function can be defined as:

$$loss = \frac{1}{N}\sum_{i=1}^{N} -(y_i \log(pr_i) + (1 - y_i) \log(1 - pr_i)) \quad (4\text{-}8)$$

Where, N is the number of data points, $pr_i$is the predicted probability, and $y_i$ is the truth label. If the truth label is 1, the second part of the equation becomes zero and activates the first part. If the truth label is 1 and the predicted probability is close to 1, the loss function becomes closer to zero. Similarly, if the predicted probability is close to 0, when the truth label is 0, the loss function is again close to zero.

Since deep learning techniques are prone to over-fitting, dropout regularization is added to the LSTM and fully connected dense layers. Dropout regularization is a well-established technique that randomly eliminates a certain number of units from each layer at each iteration with some probability. Dropout can be added to recurrent models like LSTM in two ways[109] – 1) at the recurrent part of the unit or 2) at the non-recurrent part of the unit. The recurrent parts are the horizontal arrows entering and leaving the LSTM unit in Figure 4-3. These refer to the connections from the cell states and the hidden states. Non-recurrent parts are the vertical arrows entering/leaving the LSTM unit in Figure 4-3. These refer to the connections from the input and output. Several previously published works[109], [110] suggest that dropout added to the recurrent part of the network can hamper the ability of the network to remember past states, leading to poor performance. Therefore, in this work, dropouts are added to the non-recurrent part of the LSTM model as well as to the fully connected dense layers. In Keras deep learning framework, dropout rate is defined as the fraction of units to drop. Cheng et al.[110] suggested that the value of dropout with LSTM networks can be upto 0.3. Therefore, the dropout rate on the non-recurrent layer of the LSTM network is fixed to be 0.3. Srivatsava

et al.[111], suggested that the typical values for dropout rates in the fully connected layer is between 0.2 to 0.5. Hence, the dropout rate for the fully connected dense layer is fixed to be 0.4.

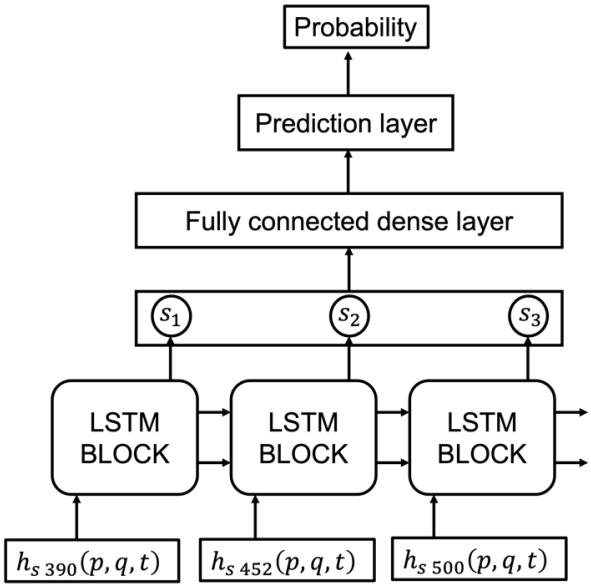



**Figure 4-4** Schematic of the LSTM model architecture showing the LSTM layer, fully connected dense layer and the output prediction layer. Reprinted with permission from [96].

### 4.2.4. **Hyperparameter tuning**

The hyperparameters were tuned using the 'Keras tuner' functionality provided by the Keras library. Keras tuner is an easy-to-use hyperparameter optimization framework that defines a search space and utilizes built-in algorithms to find the best hyperparameter values. Hyperband algorithm[112] is chosen as an effective way to tune multiple hyperparameter configurations in the model simultaneously. Hyperband is a bandit-based approach that belongs to the multi-fidelity group of hyperparameter optimization techniques. It is an improvement to the successive halving technique to address the "n vs. B/n" problem, where 'n' is the number of possible hyperparameter configurations and 'B'

is the budget allocated for the optimization (epochs, time). The successive halving technique uniformly allocates a budget ‘B’ to ‘n’ hyperparameter configurations and eliminates the worst-performing half of configurations depending on its performance. This process is repeated until a single hyperparameter configuration remains. In the successive halving technique, as ‘n’ increases, the budget, ‘B/n’ allocated for each configuration decreases and leads to inefficient optimization. In addition, for a fixed budget ‘B’, it is not intuitive to decide whether to choose a small ‘n’ and train for a long time or a large ‘n’ and train for a short time.

The hyperband algorithm addresses this problem by considering several ‘n’ values for a given budget ‘B.’ A minimum budget ‘b’ is also allocated for each of the ‘n’ values. The algorithm employs early stopping to give a smaller budget to larger values of ‘n’. Essentially, the hyperband algorithm contains two loops: the outer loop iterates over different values of ‘n’ and its corresponding ‘b’; the inner loop performs successive halving with early stopping. As the algorithm proceeds and ‘n’ decreases, more budget is allocated to optimize the hyperparameters. Thus, aggressive early stopping for larger values of ‘n’ at the beginning of the optimization produced faster results than other hyperparameter optimization techniques. In this work the budget allotted for hyperparameter tuning is 50 epochs. The following hyperparameters are tuned by varying them over a range using a specific step size.

1) LSTM units – min value=192, max value=512, step size=16
2) Dense layer neurons – min value = 32, max value = 112, step size = 16
3) Optimizer - ['Adam’, 'SGD', 'RMSprop’, 'Adamax']

4.2.5. **Data splitting strategy**

The dataset was randomly split into five folds patient-wise such that four folds contain maFLIM images from three patients with benign lesions and three patients with malignant lesions. The last fold contains maFLIM images from 5 patients with benign lesions and 1 patient with malignant lesions. Since some patients have multiple lesions, the number of lesions in each fold is variable. When the dataset is split into five folds in this manner, it forms a partition. This random splitting of the dataset allows us to produce many partitions with different combinations of benign and malignant lesions in each fold. In order to obtain a good estimate of the model performance and to eliminate any dependency on the data splitting, the entire process is repeated on 10 different random partitions of the data. Figure 4-5 shows the data splitting strategy. Each fold within a partition becomes the test set during an iteration. Thus the number of iterations equals the number of folds within a partition, which is five. In this way, every lesion data gets a chance to be a part of the test set. The validation fold for every iteration is chosen randomly from the other four folds of the dataset. The remaining three folds are combined to form the training set.

4.2.6. **Classification workflow**

Figure 4-6 shows the schematic of the workflow[96]. The training and validation sets are used to tune the model's hyperparameters using the hyperband algorithm, as explained in section 4.2.4. After choosing the best configuration of hyperparameters, the number of epochs for training is optimized by monitoring the training and validation accuracies over fifty epochs and selecting the one with the highest validation accuracy.

As a model is trained over many epochs, there is a high probability for it to overfit. When this happens, the validation accuracy decreases while the training accuracy increases.

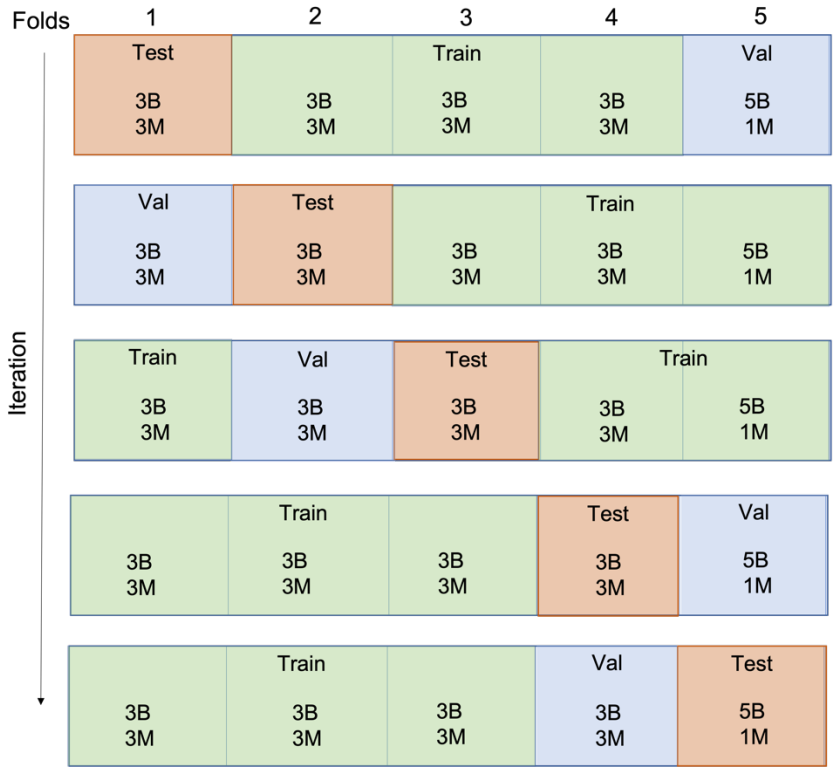



**Figure 4-5** Data-splitting strategy in each partition showing the distribution of benign and malignant patients in each fold. Reprinted with permission from [96].

After optimizing the epoch, the model is retrained using the best set of hyperparameters and best epoch. The next step is to test the model performance using the test set. Since this is a pixel-wise classifier, with the output in the range [0,1], we could generate a posterior probability map for each lesion data. The generated posterior probability maps visualize the region-wise likelihood of the imaged lesion area to be malignant. However, it is necessary to produce an image-level diagnosis of the lesions. The classification performance can be estimated by constructing a ROC curve based on a

discriminating variable. In this case, ROC curves are constructed on the median posterior probability values from the generated maps of the lesion images in each fold. A higher AUC indicates a superior classification performance on the test set. This way, five different AUC values are produced for each partition. The mean and standard deviation of the AUC values is chosen as the metric indicating the overall classification performance on that partition.

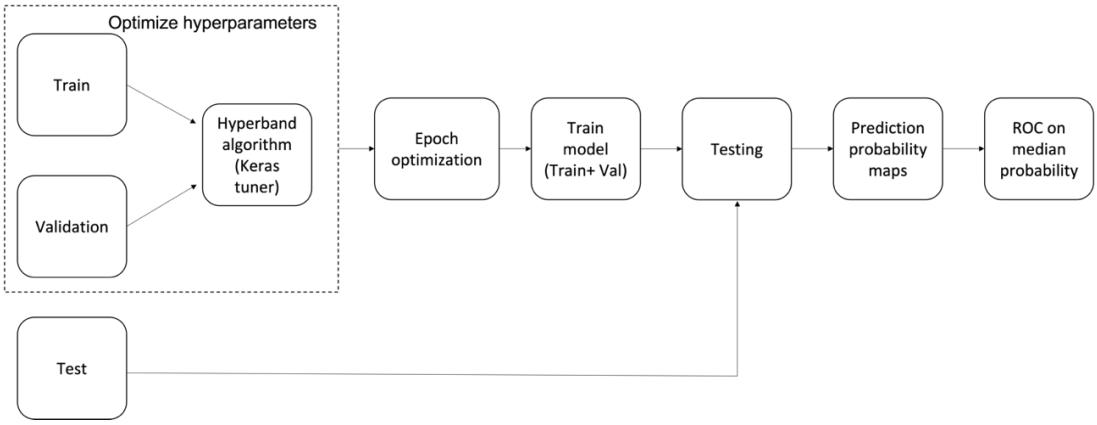



**Figure 4-6** Workflow of the methodology showing hyperparameter tuning with training and validation sets and testing the model with the test set. It is to be noted that this process is repeated five times for each partition. Since there are 10 different partitions, this is executed 50 times. Reprinted with permission from [96].

## 4.3. **Results**

### 4.3.1. **maFLIM dermoscopy of skin lesions**

A total of 30 patients undergoing tissue biopsy examination of skin lesions for cancer diagnosis participated in this study. maFLIM dermoscopy images of multiple skin lesions were acquired on each patient. Figure 4-7(a) shows a handheld maFLIM dermoscope imaging the forearm of a patient and Figure 4-7(b) shows the clinical photograph of a sample melanoma skin lesion. Figure 4-7(c) shows the integrated intensity

maps of a melanoma skin lesion at the three spectral channels. The distribution of patients ($n_{patients}$=30) and lesions ($n_{lesions}$=60) imaged in this study showing benign and malignant conditions is provided in Table 4-1. Benign lesions included solar lentigo and pigmented sebohrreic keratosis, while malignant lesions included pigmented superficial BCC, pigmented nodular BCC and melanoma.

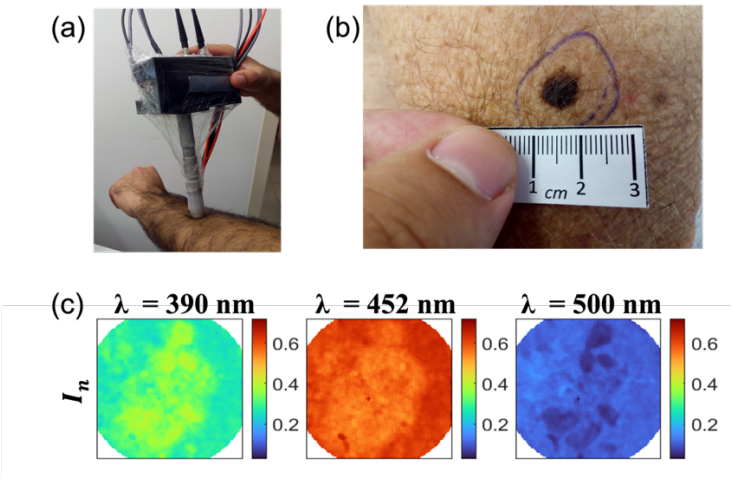



**Figure 4-7** (a) Hand-held maFLIM dermoscope imaging the forearm of a patient. (b) Clinical photograph of a melanoma lesion. (c) Integrated intensity maps of a melanoma skin lesion corresponding to the three spectral channels. Reprinted with permission from [96].

**Table 4-1** Distribution of imaged benign and malignant lesions. Reprinted with permission from [96].

| | **Type** | **No. patients** | **No. lesions** |
|---|---|---|---|
| Benign | Solar lentigo | 2 | 10 |
| | Pigmented seborrheic keratosis | 15 | 31 |
| Malignant | Pigmented superficial BCC | 2 | 6 |
| | Pigmented nodular BCC | 5 | 5 |
| | Melanoma | 6 | 8 |

### 4.3.2. **Hyperparameter tuning**

The hyper-parameters are tuned using the hyperband algorithm provided by the Keras-tuner package in Keras as explained in section 4.2.4. The workflow of the classifier design shows that the hyperparameters are optimized using the training and validation sets. Every train-validation-test set combination produces one set of optimum hyper-parameters. With a total of ten partitions and each partition consisting of five folds, we end up with fifty sets of optimum hyper-parameters. To determine the hyper-parameters that are chosen more often, histograms are constructed to represent the frequency at which different hyper-parameter values are picked up during all 50 iterations. The frequency distributions of the number of dense layer neurons, optimizers and number of LSTM units are shown in Figures 4-8 (a), (b) and (c) respectively. The number of neurons in the dense layer follows a normal distribution with a peak around 80 neurons. Adam is the most frequently selected optimizer in all 50 iterations. The number of LSTM units is almost uniformly distributed over the range 192 to 432.

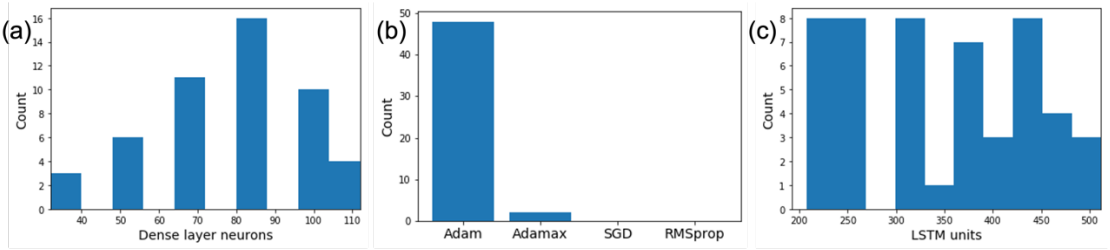



**Figure 4-8** Frequency distributions of the hyper-parameters that are selected using 10 different partitions for (a) number of dense layer neurons, (b) optimizer (c) number of LSTM units. Five folds in each partition make the total number of iterations equal to 50. Reprinted with permission from [96].

### 4.3.3. **Epoch optimization**

The optimum epoch was determined by a modified early stopping technique. Here, the model was trained on the training set using the best set of hyper-parameters selected from the previous step in the workflow. The model is validated on the validation set after each epoch to determine the validation accuracy. The model was trained on 50 epochs irrespective on the validation performance. The optimum epoch is then determined as the value corresponding to the highest validation accuracy. Subsequently, the model is retrained on a new training set obtained by combining the training and validation sets using the optimum epoch and the optimum hyper-parameters. Figure 4-9(a) shows a sample iteration with the training and validation accuracies over 50 epochs. It can be seen that the validation accuracy starts decreasing after 6 epochs, which is chosen as the optimum epoch. Figure 4-9(b) shows the training accuracy when the model is trained on the new training set (combined training and validation sets) using the optimum epoch. Figure 4-9(c) shows the histogram of the values of epochs selected during all the 50 iterations in the 10 partitions. A lower epoch value in the range between 1 to 10 is selected most frequently.

### 4.3.4. **Classification results**

As can be seen in the workflow schematic in Figure 4-6, an ROC curve on the median posterior probability values is generated for every test set. This means that each partition produces five different ROC curves corresponding to the five folds. Table 4-2 shows the mean and standard deviation of the AUC values from all the five folds in every partition.

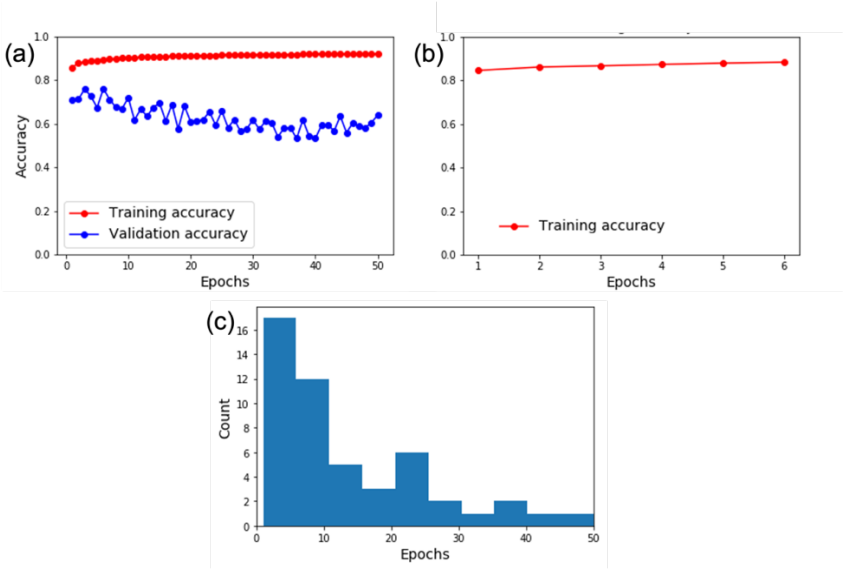



**Figure 4-9** (a) Training and validation accuracies during epoch optimization. (b) Training accuracy using optimum epoch. Reprinted with permission from [96].

A set of posterior probability maps are generated for every test set at each iteration. The entire dataset can be represented by combining the posterior probability maps generated during all the five folds in a partition. Figure 4-10 shows the maps generated during all the iterations in a sample partition of the data. It can be seen that the majority of the benign lesions are darker, while the malignant ones are more red.

**Table 4-2** Mean and standard deviation of AUC values from the test sets in each partition. Reprinted with permission from [96].

| Partition | Test AUC |
|---|---|
| 1 | 0.84±0.16 |
| 2 | 0.83±0.09 |
| 3 | 0.85±0.11 |
| 4 | 0.80±0.25 |
| 5 | 0.77±0.22 |
| 6 | 0.79±0.17 |
| 7 | 0.89±0.06 |
| 8 | 0.79±0.17 |
| 9 | 0.76±0.18 |
| 10 | 0.78±0.20 |

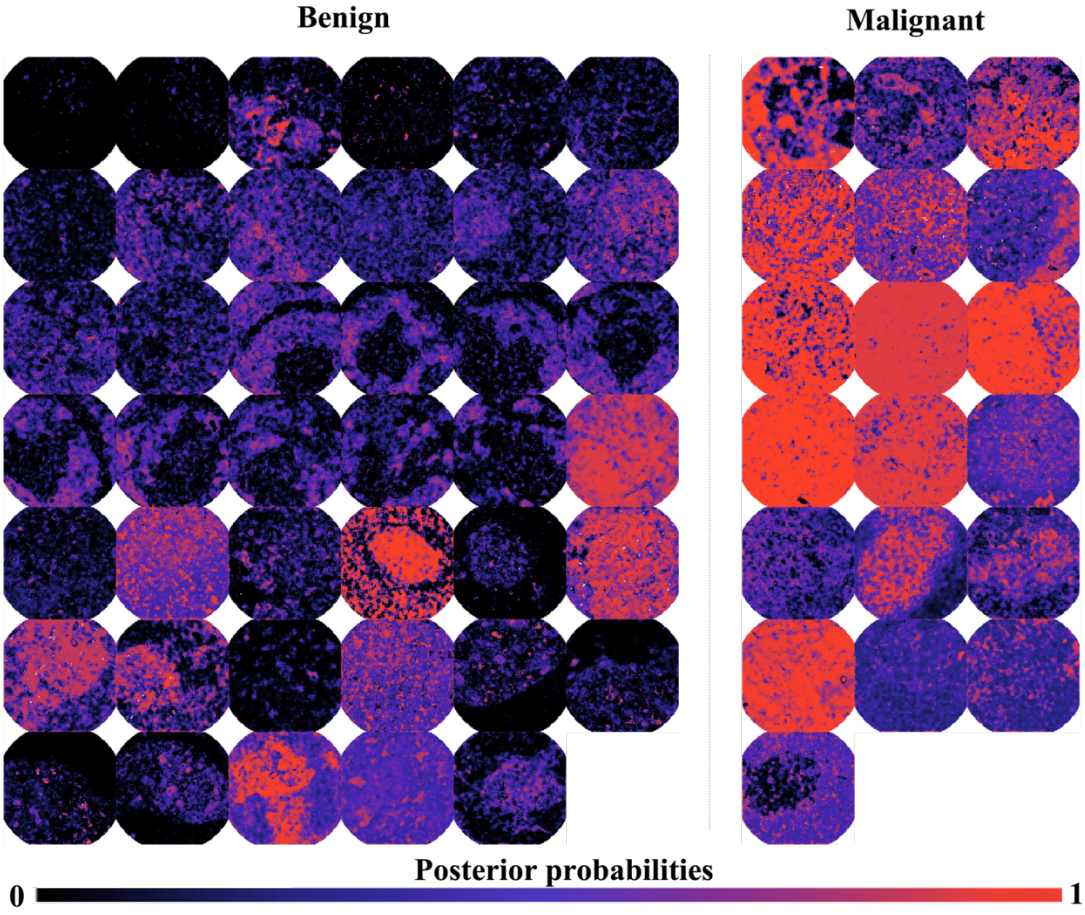



**Figure 4-10** Posterior probability maps generated from all the test sets in a sample partition of the data. Reprinted with permission from [96].

In addition to generating ROC curves on the median posterior probability values for every test set, it is also possible to generate ROC curves for all the test sets in a partition. In this way a single ROC curve can be produced for a partition. Figure 4-11 shows the ROC curves generated for all the 10 partitions with their corresponding AUC values. The mean and standard deviation for the AUC values from all the 10 ROC curves is 0.82±0.04.

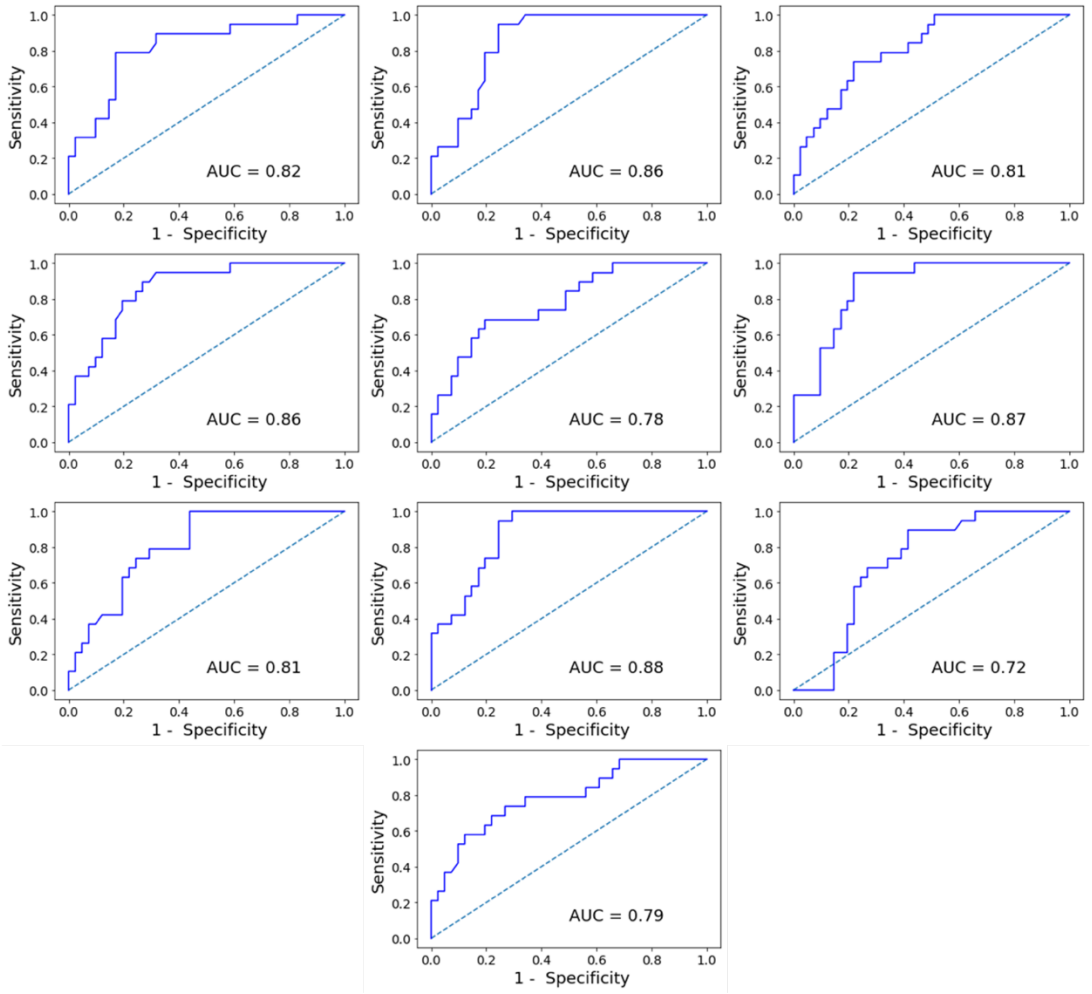



**Figure 4-11** ROC curves for ten partitions obtained by combining the results from all the five folds in each partition. Reprinted with permission from [96].

## 4.4. **Discussion**

Deep learning models offer the advantage of automatically learning from the input without employing extensive feature engineering techniques. They also bring in the additional benefit of identifying the most relevant information from the input data without explicitly conducting any feature selection. Previously published works on deep learning techniques for skin cancer detection have primarily used CNN models to classify

dermoscopic images based on their morphological features such as asymmetry, border, color, and texture. This work explores deep learning techniques on maFLIM data acquired from benign and malignant skin cancer lesions. This is particularly interesting because, unlike dermoscopic images, maFLIM data contain a much higher level of information about the metabolic changes and the composition of co-factors and other tissue constituents. The temporal component in the fluorescence decay signal at each pixel location can be effectively studied using RNN based models. Such models have the ability to monitor the input changes over a range of past time steps. This hypothesis is the basis for employing LSTM networks to classify maFLIM images acquired from skin lesions. LSTMs have the advantage of learning long term time dependencies in the input signal, which can effectively aid in differentiating the fluorescence decay signals at different pixel locations.

The main challenge in deep learning models is that it requires an extensive training set to efficiently learn from the input data. However, in medical settings, where the data is both limited and expensive, the standard method adopted to overcome this hurdle is by using data augmentation to increase the number of available data. However, this comes with the disadvantage of adding information without any significant value to the existing training set. In this work, although the number of patients (30) and the number of imaged lesions (60) are small, the size of the training set is around 650000. This is because the LSTM model is trained at the pixel level. The number of usable pixels from each image (total usable pixels depends on the number of unmasked pixels from each image after pre-processing) significantly increases the number of training samples multifold. The size of

the validation and test sets is ~ 250000. Since the pixel-level truth label is unknown, the label obtained from the histopathological evaluation of the whole lesion is assigned to every image pixel. Although this could cause some variation in the computed loss function and affect the training process to some degree, the assumption that most pixels in a malignant lesion are malignant and vice versa could produce some reliable classification performances.

Another challenge is the complexity of the input data. The number of time points in the deconvolved maFLIM signal from one spectral emission channel is 149, corresponding to 60 ns. As the size of each input signal increases, the complexity of the model also increases, which could lead to overfitting problems. In addition, the information contained in the signal after the initial decay is not significant. Therefore, efficient down-sampling of the signal to include the most important information helps to minimize any redundancy and reduces the complexity of the model. Here a down-sampling rate of 0.5 was applied on the first 20 time steps of the signal to include at least 80% of the fluorescence decay, i.e., the signal intensity dips to at least 20% of its peak value.

The results from the hyperparameter tuning in Figure 4-8 indicate that Adam is the optimum optimizer, and the number of neurons in the dense layer is optimum, around 80 neurons. However, the number of LSTM units is more or less uniformly distributed over the selected range. A lower epoch value of less than ten was optimum for training the model. A thorough evaluation of the model performance is conducted by testing the model on ten different random partitions. The five folds in each partition are such that the ratio

of patients with benign and malignant lesions is kept constant. Every partition produces five independent test sets over all the iterations. In effect, the five independent test sets for each partition correspond to fifty test sets for the entire procedure. The results in Table 4-2 indicate that the mean AUC values over the five folds in each partition are greater than 0.75 and the highest AUC obtained was 0.89. It can also be seen that the maximum variability in the AUCs is 0.25. This is because, since the number of lesions in each fold of the partition is small (less than 15), even a single misclassification can alter the AUC obtained from that fold by a large value. The ROC curves in Figure 4-11 are constructed by combining the results from all the folds in a partition. Here the mean and standard deviation of the AUCs is 0.82 and 0.04, respectively. We can see that the AUC values are greater than 0.72, with variability of only 0.04.

This work demonstrates a unique approach to classify maFLIM data using recurrent deep learning models to distinguish between benign and malignant skin lesions. The limitations are as follows: 1) Single laser excitation at 355 nm may not efficiently excite all the targeted endogenous fluorophores. 2) The signals collected from the broad emission spectral bands may be the combined response from multiple fluorophores. 3) Since the pixel-wise ground truth label is unknown, the label for the whole image is assigned to all the image pixels, thereby limiting efficient model training and validation. However, the assumption that most pixels have the same label as the image resulted in a decent performance. 4) More comparative analysis on different machine learning models is required to thoroughly understand the LSTM model's capabilities. The future work will classify skin cancer lesions by exploring the time decay information from the raw

maFLIM signal. Since LSTM based models can get information directly from the decay signals, it will be even simpler to feed the preprocessed measured signal to obtain a classification. This work can be translated to a clinical setting for margin demarcation and evaluation problems where the physicians can decide the boundaries of the lesions. In addition, extending this work to real-time applications can aid in the classification of the targeted tissue areas during surgery to remove the suspicious lesions.

## 4.5. **Conclusions**

Early diagnosis of pigmented skin lesions such as melanoma is necessary to provide adequate treatment strategies to the patients. This work shows the capabilities of LSTM based deep learning models to directly extract information from the fluorescence time decay signals to classify benign and malignant pigmented skin lesions with an overall ROC-AUC of 0.82±0.04. Posterior probability maps generated by the model helps to identify regions of full-blown malignancy, which has a potential to be used in margin delineation studies.

# 5. DEEP-LEARNING ASSISTED CLASSIFICATION OF ORAL CANCER LESIONS USING MULTISPECTRAL AUTOFLUORESCENCE IMAGING

## 5.1. **Introduction**

The number of new oral cancer cases estimated in the US in 2021 is 54,010, and the estimated number of deaths is10,850[1]. The five-year survival rate of oral cancer lesions diagnosed early without any metastasis is 85%[1]. However, when diagnosed at stages showing regional and distant metastases, the five-year survival rates decrease to 67% and 40%, respectively. Unfortunately, most oral cancer lesions are diagnosed at advanced stages of malignancy. This is because, the pre-cancerous lesions which are often ignored by the patients transform into malignant lesions unless they undergo routine dental checkups. Only 29% of all oral cancer lesions are diagnosed at early stages. 48% and 19% of lesions are diagnosed at stages showing regional and distant metastases[1]. Oral cancer is a major health problem in developing and under-developed countries. It is most prevalent in South-East Asian countries, and its incidence among men is higher than that among women[34]. The oral cancer cases in the Indian subcontinent alone account for one-third of worldwide cases[34], [35]. The major risk factor for oral cancer is the prolonged chewing or smoking of tobacco. Alcohol use, syphilis, HPV and dietary deficiencies[36] are the other causes of oral cancer.

More than 90% of all reported head and neck cancers belong to oral SCC, affecting the epithelial layer of the oral mucosa[113]. As the tumor grows, the basement membrane of the oral mucosa ruptures, causing the infiltration of malignant cells into the lamina

propria and deeper connective tissues. The characteristics of the tumor change at advanced stages of malignancy, and it becomes hard, ulcerous, and forms island-like structures on the oral cavity. Clinical oral examination (COE)[113] is the primary method for oral cancer screening involving visual examination and palpation. The appearance of cancerous lesions at their early stage may be similar to benign conditions due to the absence of swelling or ulceration. Therefore, early diagnosis by COE may be challenging as the risk of misdiagnosing malignant lesions is high. If the lesion persists for more than three weeks, COE is followed by biopsy resection and histopathological evaluation. However, an invasive biopsy inside the oral cavity is painful, and the average healing time is longer when compared to a skin biopsy at an exposed region. In addition, biopsy procedures are not recommended for high-risk populations. Another technique for oral cancer screening is biofluid screening analysis, in which the presence and concentrations of specific biomarkers in the saliva[114] or blood serum[115], [116] is determined. These measures are indicative of possible malignant developments in the oral cavity. Vital staining and brush biopsy[117] are other techniques for oral cancer screening where the regions of interest are stained with specific dyes, and cells from the suspicious areas are exfoliated using a brush respectively.

The two most important endogenous fluorophores in the oral tissue are the metabolic cofactors, NADH and FAD[41]. An increase in cell metabolism is a characteristic of neoplastic progressions. The variations in the concentrations of the metabolic cofactors, NADH, and FAD, can be monitored using autofluorescence measurements. Additionally, deviations in the metabolic pathways cause variations in the

relative amounts of free or enzyme-bound cofactors and their binding sites[17]. These variations are reflected in the fluorophores' long and short lifetime components. Therefore, monitoring the fluorescence decay dynamics adds an extra dimension to the capability of autofluorescence measurements in differentiating malignant and normal cells. Other endogenous fluorophores such as collagen and elastin[41] also show differences in their autofluorescence response at the onset or progression of malignancy. This is due to the thickening of the epithelium and the rupture of the basement membrane at the advanced stages of cancer. Time-resolved techniques such as maFLIM facilitate the measurement of fluorescence lifetimes and the monitoring of decay dynamics.

Several CAD techniques have emerged in the past decade to automate and standardize the pre-screening of oral cancer. The advancements in deep learning have further improved the capabilities of CAD techniques to identify malignant lesions efficiently. Table 5-1 lists the methods adopted in some recently published works using deep learning algorithms to classify oral cancer lesions. Aubreville et al.[118], classified cancerous oral lesions from healthy tissue using LeNet-5 CNN architecture, using confocal laser endomicroscopy images. Bhandari et al.[119], built a multiclass oral lesion classifier using CNN with a modified loss function on magnetic resonance imaging (MRI) images. Song at al.[120], in 2018, developed CNN based classifier using autofluorescence and white light images to classify normal tissue and lesions that exhibit suspicions of malignancy oral lesions. Song et al.[121], in 2021, used Bayesian deep networks, to discriminate benign and healthy tissues from suspicious lesions using a smartphone-based intraoral screening device. Folmsbee et al.[122], classified hematoxylin and eosin (H&E)

stained histological images into seven classes including benign and malignant oral lesions using CNN networks and active learning. Halicek et al.[123], built a CNN based classifier for discriminating healthy and cancerous images using hyperspectral images. Jeyaraj et al.[124], developed a CNN based classifier using hyperspectral images for classifying healthy and benign lesions. Lin et al.[125], developed a multiclass classifier on smartphone images using high resolution representation learning network (HRNet). Nanditha et al.[126], built an ensemble classifier of CNNs based on Resnet-50 and VGG-16 architectures for differentiating premalignant and normal tissues. Welikala et al.[127], collected images from oral lesions using mobile mouth screening anywhere mobile application (MeMoSa) and classified them using Resnet101 CNN model. Xu et al.[128], developed a 3D-CNN based classifier to discriminate benign and malignant oral lesions using computed tomography (CT) images.

**Table 5-1** Summary of previously reported works on oral cancer classification using deep learning

| Imaging modality | Distribution of patients or images | Classification task | Algorithm | Performance | Validation or testing technique | Reference |
|---|---|---|---|---|---|---|
| Confocal laser endomicroscopy | Total images - 7894 | Cancer vs normal | CNN | AUC – 0.96<br>Sensitivity – 86.6%<br>Specificity – 90.0%<br>Accuracy - 88.3% | LOPOCV | Aubreville, 2017[118] |
| MRI | Total images - 3045<br>Total patients - 230 | Four classes<br>1. SCC,<br>2. Oral lymphoma,<br>3. Oral osteosarcoma,<br>4. Oral fibro sarcoma | CNN | Accuracy - 96.5% | 7-fold CV | Bhandari, 2020[119] |

Table 5-1 Continued

| Imaging modality | Distribution of patients or images | Classification task | Algorithm | Performance | Validation or testing technique | Reference |
|---|---|---|---|---|---|---|
| Autofluorescence and white light images | Total patients – 170<br>Normal – 86<br>Suspicious - 84 | Normal vs suspicious | CNN | Sensitivity – 85.0%<br>Specificity – 88.7%<br>Accuracy - 86.9% | 4-fold CV | Song, 2018[120] |
| Smartphone-based intraoral screening device | Total images – 2350<br>Normal and benign – 1510<br>Suspicious - 840 | (Normal + benign) vs suspicious | Bayesian deep network | Accuracy – 90.0% | Train-validation-test sets | Song, 2021[121] |
| H&E stained histology images | Total images - 143 | Seven classes<br>1. Stroma,<br>2. Lymphocytes,<br>3. Tumor,<br>4. Mucosa,<br>5. Keratin pearls<br>6. Blood<br>7. Background or adipose | CNN with active learning | Accuracy – 96.44% | Average of 10 test sets | Folmsbee, 2018[122] |
| Hyperspectral images | Total patients - 50 | Normal vs cancer | CNN | <u>LOPOCV</u><br>Sensitivity – 77.0%<br>Specificity – 78.0%<br>Accuracy – 77.0%<br><u>Train-Test set</u><br>Sensitivity – 96.8%<br>Specificity – 96.1%<br>Accuracy - 96.4% | 1. LOPOCV<br>2. Train-test sets | Halicek, 2017[123] |
| Multidimensional hyperspectral image | Total images - 100 | Cancer vs benign | Partitioned deep CNN | Sensitivity – 94.0%<br>Specificity – 91.0%<br>Accuracy - 91.4% | Train-test sets | Jeyaraj, 2019[124] |

Table 5-1 Continued

| Imaging modality | Distribution of patients or images | Classification task | Algorithm | Performance | Validation or testing technique | Reference |
|---|---|---|---|---|---|---|
| Smartphone images | Total test images – 455 | Five classes<br>1. Normal<br>2. Aphthous ulcer,<br>3. Low risk OPMD<br>4. High risk OPMD<br>5. SCC | HRNet | Sensitivity – 83.0%<br>Specificity – 96.6%<br>Precision – 84.3%<br>F1 score – 83.6% | Train-test sets | Lin, 2021[125] |
| Digital true color images | Total images - 332<br>Benign - 63<br>Precancerous - 269 | Premalignant vs normal | Ensemble of CNNs | Sensitivity – 98.1%<br>Specificity – 94.2%<br>Accuracy – 96.2% | Train-test sets | Nanditha, 2021[126] |
| MeMoSa captured images | Total images – 2155<br>Total patients - 1085 | Five classes Based on lesion, referral, OPMD and cancer | Classification – CNN<br>Object detection – Faster R-CNN | F1 score– 87.07% | Train-validation-test sets | Welikala, 2020[127] |
| CT | Total images - 7000 | Benign vs malignant | 3D-CNN | Sensitivity – 81.8%<br>Specificity – 73.9%<br>Accuracy – 75.4%<br>AUC – 0.796 | No independent validation | Xu, 2019[128] |
| maFLIM | Total patients – 53 | Cancer vs healthy (Region-level and point-level) | 1D-CNN | <u>Region-level</u><br>Sensitivity – 79%<br>Specificity – 76%<br>AUC – 0.79<br><u>Point-level</u><br>Sensitivity – 62%<br>Specificity – 67%<br>AUC – 0.70 | LOPOCV | Marsden et al., 2021[129] |

AUC – Area under the curve
CNN – Convolutional neural network
MRI – Magnetic resonance imaging
SCC – Squamous cell carcinoma
CV – Cross-validation
H&E - Hematoxylin and eosin
CT – Computed tomography
R-CNN – Region based convolutional neural networks
LOPOCV – Leave one patient out cross-validation

As can be seen from Table 5-1, the only study that employed deep learning algorithms for classifying oral cancer lesions was by Marsden et al.[129], in 2021. A 1D-CNN model consisting of four convolutional layers was used to learn from the fluorescence decay signal at each pixel to classify cancerous lesions from healthy regions to obtain a region-level sensitivity of 79%, specificity of 76%, and a point-level sensitivity of 62%, and specificity of 69%. Additionally, it can also be seen that most deep learning models listed in Table 5-1 use different variants of CNNs. In this work, we employ LSTM based deep learning models to study the fluorescence decay dynamics of maFLIM signals. The temporal component of maFLIM data could be efficiently studied using deep learning models typically employed for time series applications. The LSTM based deep learning model developed in this work classifies cancerous and healthy oral cancer lesions. The model generates classification maps of the imaged region to identify the pixels that are potentially malignant.

## 5.2. **Methodology**

### 5.2.1. **Oral cancer database**

The oral cancer database consists of data collected using a handheld multispectral maFLIM oral endoscope stationed at Hamad Medical Corporation in Qatar. maFLIM data were acquired from 34 patients before they underwent tissue biopsy. The collected fluorescence emission signals correspond to three spectral channels: 390±20 nm, 452±22 nm, and >500 nm, preferentially targeting collagen, NADH, and FAD, respectively. Two sets of maFLIM images were collected from each patient, one from the lesion tissue and

the other from the contralateral healthy tissue. Each maFLIM image contains 160×160 pixels with a temporal resolution of 0.25 ns. According to the histopathological information, the lesions are labeled as mild dysplasia, moderate dysplasia, high-grade (HG) dysplasia, and SCC. To classify the maFLIM images as either cancerous or healthy, SCC, mild, moderate, and HG dysplasias are grouped together as cancerous lesions. Since there is a healthy maFLIM image associated with every lesion image, the two classes are perfectly balanced.

5.2.2. **Data pre-processing**

The maFLIM data measured at each image pixel (p,q) is composed of three fluorescence intensity temporal decay signals $y_{m,\lambda}(p,q,t)$ measured at the three target emission spectral band (λ). The steps involved in pre-processing of the maFLIM data are as follows: First, offset and background subtraction was applied to the temporal signal at each pixel of the maFLIM image. Second, pixels presenting temporal signal saturation were detected by setting a threshold on the maximum signal amplitude, and masked. Third, spatial averaging (order 5 × 5) was applied to increase the temporal SNR at each spatial location. Fourth, pixel masking based on SNR was also performed with an SNR threshold value of 15 decibels. The pre-processed signal at each pixel location is represented as $y_\lambda(p,q,t)$ for the three spectral emission bands. The steps involved in obtaining the deconvolved signal, $h_\lambda(p,q,t)$ from $y_\lambda(p,q,t)$ is given in equations 3-2 and 3-3 in Chapter 3. The concatenated deconvolved signal, $h(p,q,t)$ is given in equation 4-1 in Chapter 4.

The deconvolved signal at each pixel location is first down-sampled in time in such a way as to represent at least 80% of the fluorescence signal intensity. This considerably reduces the computational overload in comparison to when the entire decay signal is input to the deep learning model. The length of the fluorescence decay signal at every pixel location per spectral emission channel, $h_\lambda(p, q, t)$ is 300 timesteps or 75 ns as shown in Figure 5-1(a). It was found that the fluorescence intensity decays from 100% to at least 20% of its peak value within the first 40 time steps, or in 10 ns. Therefore, the first 40 time steps of the decay signal is downsampled into 10 equally spaced time steps with a down-sampling rate of 0.25. The vertical lines shown in Figure 5-1(a) indicates the sampling intervals. In this way, the signals from all the three spectral channels are down-sampled to form $h_{s\lambda}(p, q, t)$ as seen in Figure 5-1(b). Therefore, the dimension of the signal from each maFLIM image pixel is 3×10; i.e., 10 time points corresponding to the 3 spectral emission channels.

5.2.3. **LSTM model**

A detailed introduction to LSTM networks and its internal diagram is provided in section 4.2.3 and Figure 4-3 respectively. The input to the LSTM model used in this work is the concatenated and down-sampled fluorescence decay signal from the three spectral channels. The schematic of the LSTM model with the inputs, fully connected layer, and the output is the same as shown in Figure 4-4 in Chapter 4. In this dataset, the ground truth, or the label for an maFLIM image is obtained from the histopathological evaluation after biopsy resection. A label of 0 indicates that the image is healthy, while a label of 1

indicates that the image is cancerous. Since the label corresponding to every pixel location of an image is unknown, we assign the label of the entire maFLIM image to its pixels.

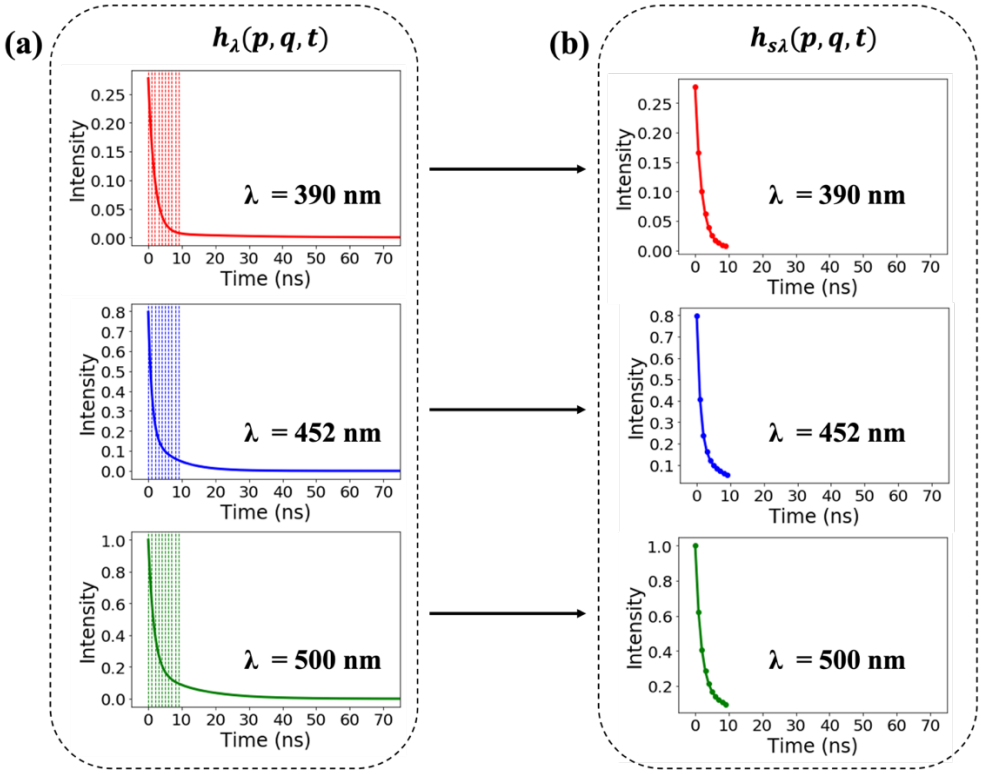



**Figure 5-1** Down-sampling of a sample maFLIM signal. (a) Deconvolved maFLIM signals from the three spectral emission bands. Vertical lines show the time points at which the signals are down sampled. (b) Down sampled signals from the three spectral emission bands.

The input to the LSTM model used in this work is the down-sampled fluorescence decay signal, $h_{s\lambda}(p,q,t)$, from the three spectral channels at each pixel location. The dimension of outputs from the LSTM layers, $s_1$, $s_2$, and $s_3$ depends on the number of LSTM units used. The output from the LSTM layer is fed to a fully connected dense layer followed by a single unit output prediction layer. A sigmoid activation at the prediction

layer ensures that the output lies in the range [0,1] to resemble posterior probabilities. This model consisting of the LSTM layer, the fully connected layer and the output layer is optimized together by defining a loss function computed from the ground truth and the prediction probability. A binary cross-entropy loss function is used since it performed best for binary classification problems.

Since deep learning techniques are prone to over-fitting, dropout regularization is added to the LSTM and fully connected dense layers. As explained in the previous chapter, dropouts are added to the non-recurrent part of the LSTM network and the fully connected dense layer. The dropout rate on the non-recurrent layer of the LSTM network is fixed to be 0.3 and that of the fully connected dense layer is fixed to be 0.4.

#### 5.2.4. **Classifier design**

The dataset is randomly split into seven folds so that each fold contains some lesion images and their corresponding healthy images. In this way, six folds have ten images, and the seventh fold includes eight images. When the dataset is split into seven-folds, it forms a partition. This random splitting of the dataset allows us to produce many partitions with different healthy and cancerous lesion combinations in each fold. The following steps were performed on ten different partitions of the data to obtain a reasonable estimate of the model performance and to eliminate any dependency on the data splitting. Figure 5-2 shows the schematic of the data-splitting and the allocation of train, test and validations sets in each iteration for a partition.

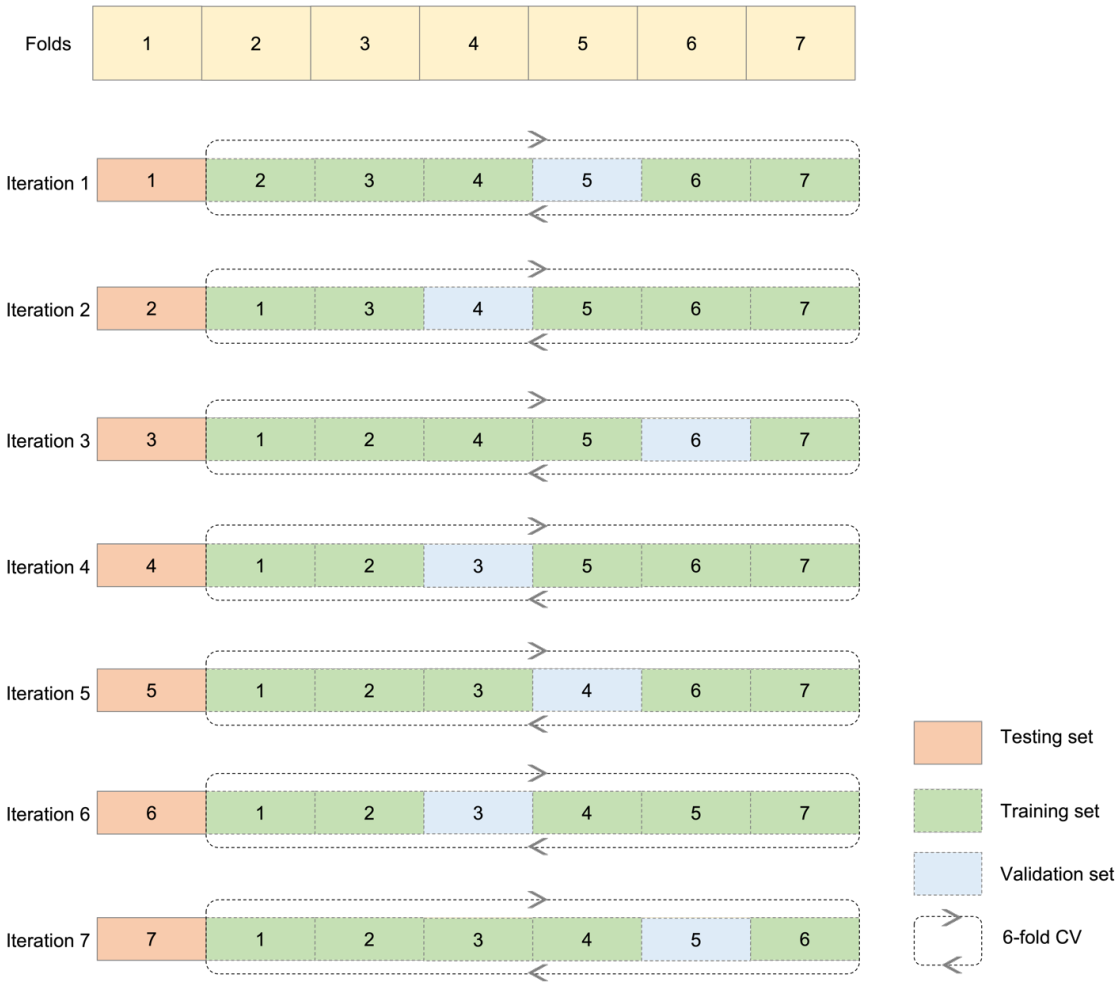



**Figure 5-2** Schematic of data splitting for a single partition of the data. The allocation of train, test and validation sets are also shown. The validation and train sets within the 6-fold cross-validation change during the cross-validation. The boxes with dotted line border indicate that their train or validation set assignment change during the iteration. CV – Cross-validation

Figure 5-3 shows the schematic of the workflow. One fold within a partition becomes the test set during an iteration. Thus the number of iterations equals the number of folds within a partition, seven. In this way, every image gets a chance to be part of the test set. The remaining folds are subsequently used in a 6-fold cross-validation, such that each of the 6 folds become the validation set in each iteration. This validation set is used

to tune the model and fix the hyperparameters. In each cross-validation iteration, the model is first trained on the training set (formed by combining five folds) and validated on the validation set. In this way, a posterior probability is generated for every pixel of every image in the validation set. This creates a posterior probability map for every image in the validation set.

The next step is to optimize a threshold for the posterior probability ($Th_{prob,\ fold}$) to classify the pixels as either cancerous or healthy. A receiver opertator characteristics (ROC) curve is constructed on the median posterior probability values from the generated posterior probability maps in the validation set. The optimum thresholds, $Th_{prob}$ is chosen as the threshold value closer to the ideal point [0,1] on the ROC curve. Classification maps consisting of class label values (0 - healthy or 1 - cancer) for every pixel location are generated for every validation set image. This map can be noisy as it may contain some isolated pixels classified as either 0 or 1. To minimize the noise and identify continous regions in the image that belong to either of the two classes, spatial averaging with a 5×5 median filter is performed on the classification maps. Subsequently, a ROC curve is constructed on the percentage of cancer pixels for every image of the validation set. This helps to identify a threshold on the percentage of cancer pixels ($Th_{pix,\ fold}$) to classify the entire image as either cancerous or healthy. The final values of the two thresholds ($Th_{prob}$, $Th_{pix}$) are chosen as the median from all the six folds. Figure 5-4 shows an example of how the pixel-wise output from the LSTM model is transformed into image-level classification.

Finally, the test set is evaluated on the model trained on both training and validation sets (6 folds combined). The pixel-wise output from each test set image is converted to image-level classification in the same manner as explained in Figure 5-4. Therefore, a confusion matrix is generated for each test set in a partition. This way, seventy confusion matrices are generated from all the test sets from the ten partitions. This gives a reasonable estimate of the model performance on independent tests and the generalization capabilities of the model.

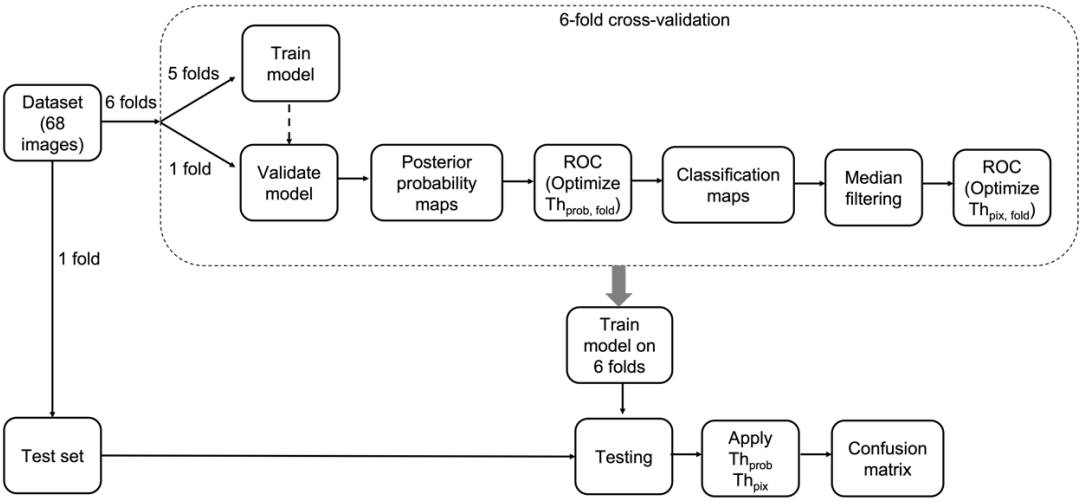



**Figure 5-3** Workflow of the methodology showing training, validation and testing sets. It is to be noted that this process is repeated seven times for each partition. Since there are 10 different partitions, this is executed 70 times. $Th_{prob}$ and $Th_{pix}$ are the median values of $Th_{prob,fold}$ and $Th_{pix,fold}$ over all 6 folds respectively.
ROC – Receiver operator characteristics

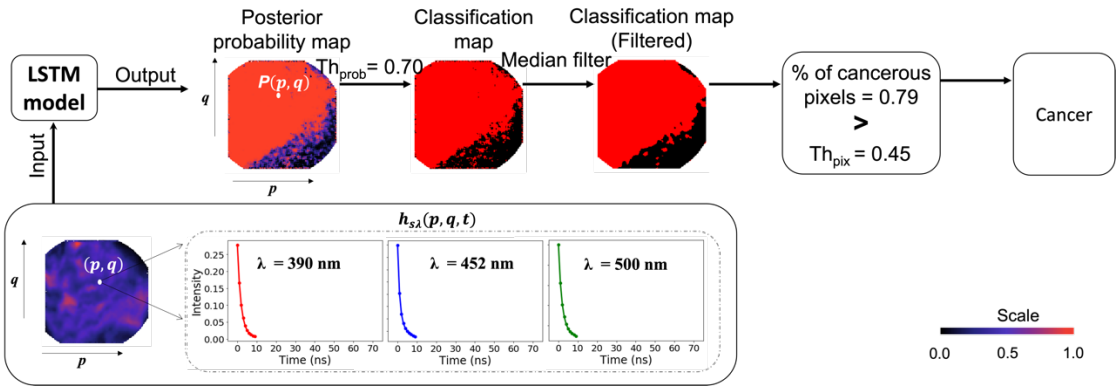



**Figure 5-4** Transformation of the pixel-wise output from the LSTM model to the image level classification
LSTM – Long Short-Term Memory

## 5.3. **Results**

### 5.3.1. **maFLIM measurements from oral lesions**

A total of 34 patients undergoing tissue biopsy examination of oral lesions for cancer diagnosis participated in this study. Figure 5-5(a) shows a handheld maFLIM endoscope and Figure 5-5(b) shows the clinical photograph of a sample oral cancer lesion. Figure 5-5(c) shows the integrated intensity maps of a lesion at the three spectral channels. The lesions contained in the dataset originate from different regions of the oral cavity, such as buccal mucosa, gingiva, tongue, mandible, the floor of the mouth, maxilla, lip, and palate. Table 5-2 shows the distribution of maFLIM data collected from different regions of the oral cavity.

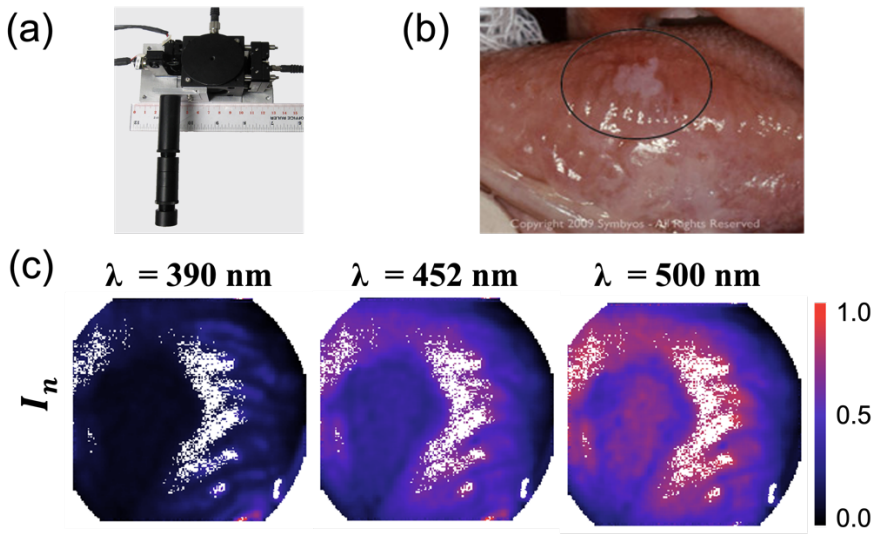



**Figure 5-5(a)** Hand-held maFLIM endoscope. (b) Clinical photograph of an oral cancer lesion. (c) Integrated intensity maps of an oral cancer lesion corresponding to the three spectral channels.

**Table 5-2** Distribution of maFLIM data collected from different regions of the oral cavity

| Lesion type | Lesion Location | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Tongue | Gingiva | Buccal Mucosa | Mandible | Floor of mouth | Maxilla | Lip | Total |
| Mild Dysplasia | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| Moderate Dysplasia | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| HG Dysplasia | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 3 |
| SCC | 12 | 3 | 9 | 1 | 1 | 1 | 2 | 29 |

HG – High grade
SCC – Squamous cell carcinoma

### 5.3.2. **Threshold optimization**

As explained in section 5.2.4, the two thresholds, $Th_{prob}$ and $Th_{pix}$ are optimized for every test set. This means that 70 different values of the thresholds are produced after repeating the experiments on the 10 partitions. Therefore, it is interesting to analyze the

distribution of the threshold values selected during all the 70 iterations. Figure 5-6 shows the distribution of the two thresholds. The mean and standard deviation of $Th_{prob}$ and $Th_{pix}$ over all the iterations are 0.79±0.08 and 0.46±0.06 respectively.

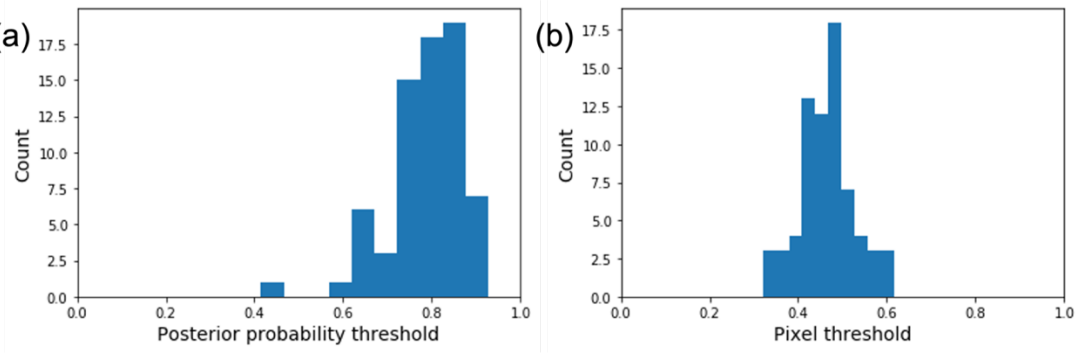



**Figure 5-6** Distribution of the thresholds, $Th_{prob}$ and $Th_{pix}$ selected over all iterations for ten partitions

### 5.3.3. **Classification results**

Table 5-3 shows the mean and standard deviations of the sensitivities, specificities and accuracies obtained from all the test sets for each partition. It can be seen that all the sensitivities are greater than 82% and specificities are greater than 75%. It is also possible to construct a single confusion matrix by combining the results from all the test sets in a partition. The performance metrics computed from the combined confusion matrix in each partition is shown in Table 5-4. The mean and standard deviations of the sensitivities, specificities, and accuracies from all the partitions are 86.4±2.35, 82.0±4.0, and 84.24±1.9 respectively.

**Table 5-3** Mean and standard deviation of the performance metrics computed from the test sets in each partition

| **Partition** | **Sensitivity (%)** | **Specificity (%)** | **Accuracy (%)** |
|---|---|---|---|
| 1 | 88.6±20.0 | 82.8±27.0 | 85.7±15.0 |
| 2 | 87.8±10.0 | 80.0±21.0 | 83.9±11.6 |
| 3 | 85.0±14.3 | 80.0±21.3 | 82.50±10.0 |
| 4 | 87.8±15.1 | 79.9±18.0 | 83.9±8.8 |
| 5 | 87.8±15.1 | 76.4±19.7 | 82.1±9.2 |
| 6 | 82.1±16.8 | 90.7±10.8 | 86.4±9.5 |
| 7 | 84.9±9.6 | 79.2±10.8 | 82.1±7.5 |
| 8 | 87.8±15.1 | 85.7±17.6 | 86.8±6.9 |
| 9 | 87.8±10.6 | 85.7±13.9 | 86.7±4.3 |
| 10 | 82.8±19.8 | 82.8±16.6 | 82.8±10.3 |

A set of posterior probability maps are generated for every test set before applying $Th_{pos}$ at each iteration. The entire dataset can be represented by combining the posterior probability maps generated during all the seven folds in a partition. Figure 5-7 shows the maps generated during all the iterations in a sample partition of the data. It can be seen that the majority of the healthy images are darker, while the cancer images are more red.

Figure 5-8 shows the classification maps after median filtering obtained from all the tests sets in a sample partition. These classification maps are generated from the posterior probability maps in Figure 5-7. The red pixels correspond to those pixels classified as cancer (class 1) and black pixels correspond to those pixels classified as

healthy (class 0). It can be noted that the images belonging to the cancer class contain more red pixels than those belonging to the healthy class.

**Table 5-4** Performance metrics computed from combined confusion matrices in each partition

| **Partition** | **Sensitivity (%)** | **Specificity (%)** | **Accuracy (%)** |
|---|---|---|---|
| 1 | 88.2 | 82.3 | 85.3 |
| 2 | 88.2 | 79.4 | 83.8 |
| 3 | 85.3 | 79.4 | 82.3 |
| 4 | 88.2 | 79.4 | 83.8 |
| 5 | 88.2 | 76.5 | 82.3 |
| 6 | 82.3 | 91.1 | 86.7 |
| 7 | 85.3 | 79.4 | 82.3 |
| 8 | 88.2 | 85.3 | 86.8 |
| 9 | 88.2 | 85.3 | 86.8 |
| 10 | 82.3 | 82.3 | 82.3 |
| Mean ± standard deviation | 86.4±2.3 | 82.0±4.0 | 84.2±1.9 |

## 5.4. **Discussion**

The utility of recurrent deep learning models to learn from the maFLIM fluorescence decay signals is investigated in this work. The maFLIM data collected from the three emission spectral channels is down-sampled to minimize the model complexity.

The pixel-level output from the model is converted to image-level classification by applying two thresholds – one on the posterior probabilities and the other on the percentage of cancer pixels. This resulted in classification maps that help to identify the regions in the image that are cancerous. A thorough model evaluation is conducted with 70 test sets over ten different partitions. The overall sensitivity of 86.4% and specificity of 82% were obtained, with standard deviations less than 4%.

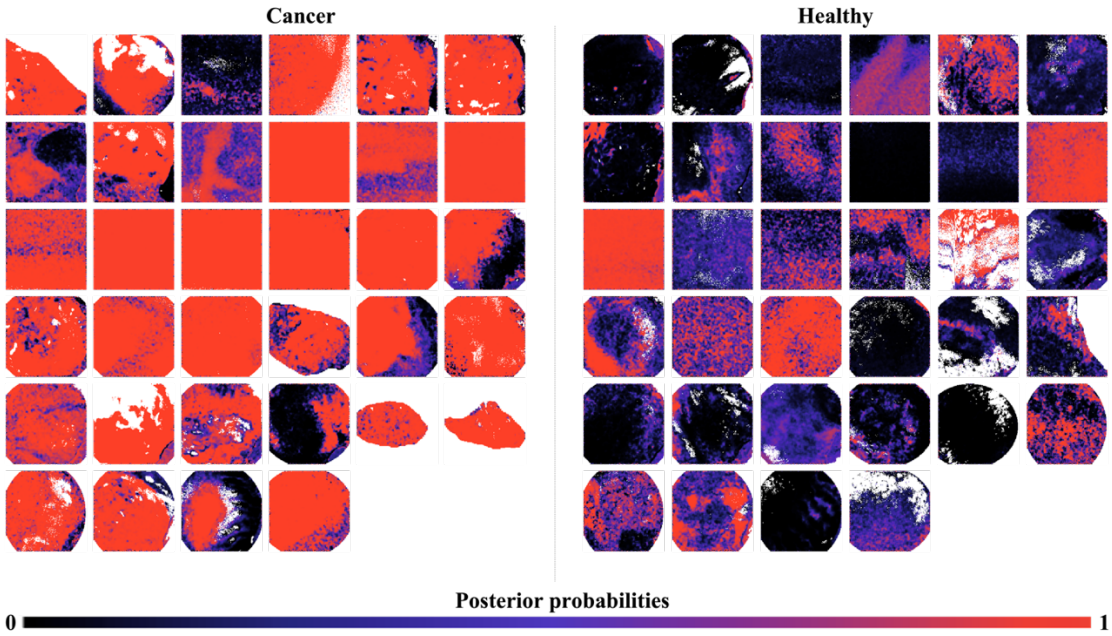



**Figure 5-7** Posterior probability maps generated from all the test sets in a sample partition of the data. These maps are generated before applying $Th_{prob}$ to the test set images.

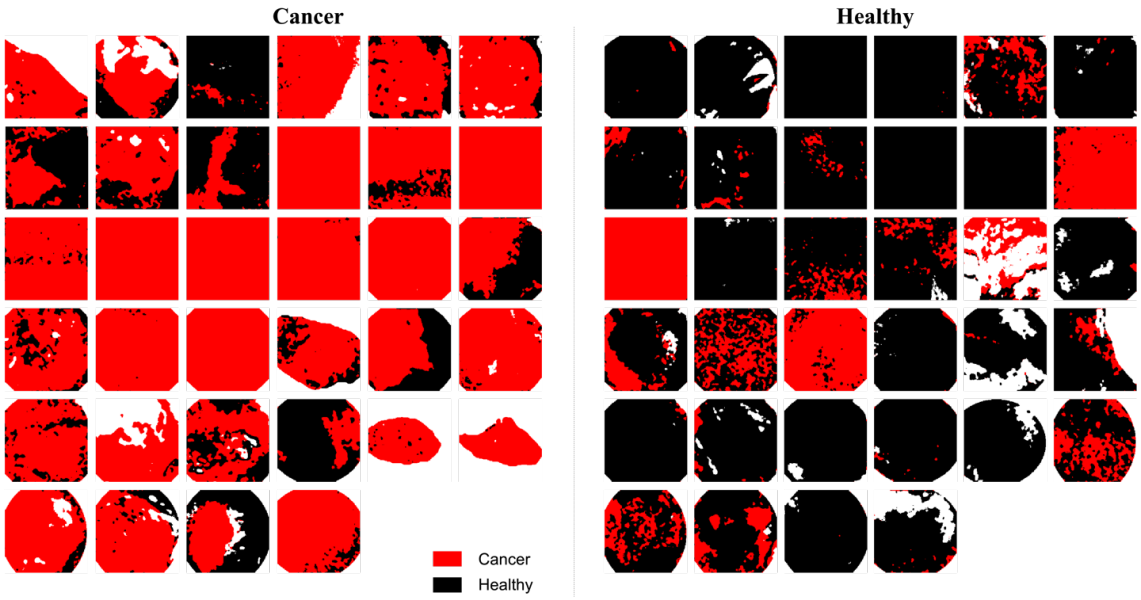



**Figure 5-8** Classification maps generated from all the test sets in a sample partition of the data after median filtering. These maps are generated before applying $Th_{pix}$ to the test set images.

To the best of our knowledge, this is the first work that employed LSTM networks to classify cancerous and healthy oral cancer using maFLIM data. One other work, Marsden et al.122, used 1D CNNs to classify healthy and cancerous oral lesions. However, the sensitivity and specificity obtained at the region-level are 79% and 76%, respectively. In comparison, the classification performance obtained in this work is superior to that obtained using other deep learning models on maFLIM data for oral cancer classification.

The small dataset problem with deep learning models is overcome by training at the pixel level, with 160x160 pixels per image. The assumption that most pixels have the same label as that of the image allows assigning the image label to every pixel. However, this can have its own drawbacks by reducing the model's training efficiency. As seen in Figure 7, the values of both the thresholds are normally distributed over all the iterations

of the 10 partitions. The mean value of $Th_{prob}$ is around 0.79 while that of $Th_{pix}$ is around 0.46. This means that if more than 46% of pixels in an image is classified as cancer, the whole image will be considered as belonging to the cancer class.

The limitations of this work are as follows: 1) Single excitation wavelength at 355 nm may not efficiently excite all the three targeted fluorophores. 2) Due to the broad spectral range of the emission channels, the signals collected from each channel could be the combined effect of the fluorescence decays from multiple fluorophores. 3) Due to the lack of pixel-level ground truth information, the loss calculated at each iteration may not be optimal, which could hinder effective model training. As a future work, comparisons with standard machine learning models is needed for thorough performance evaluation. In addition, this work could be extended on raw maFLIM signals to allow the network to perform classification with minimal pre-processing. This work can be translated to a clinical setting for margin demarcation and evaluation problems where the physicians can decide the boundaries of the lesions. In addition, extending this work to real-time applications can aid in the classification of the targeted tissue areas during surgery to remove the malignant lesions.

## 5.5. **Conclusions**

Classification of oral cancer lesions from healthy tissue using LSTM based deep-learning models has been demonstrated in this work. The classification maps generated by the model helps to identify regions of malignancy, which in turn can be useful in margin delineation applications, especially during trans-oral surgeries where the surgeons can identify the cancerous regions in the oral cavity.

# 6. CONCLUSIONS AND FUTURE WORKS

The primary goal of this work is to develop CAD techniques for identifying benign and malignant skin cancer lesions, as well as oral cancer lesions from the contralateral healthy tissue. Primarily two approaches are considered for this study: 1) traditional machine-learning, and 2) deep-learning based techniques. The efficiencies of the developed models are thoroughly evaluated using multiple cross-validation strategies as well as with several random partitions of the data. The specific contributions to each study are as follows:

**Machine learning based classification of skin cancer lesions**: A non-iterative and simple frequency-domain deconvolution of maFLIM data is implemented. Phasor-based features are extracted after frequency-domain deconvolution of the data. The classification performance of the phasor features is compared with that of the features derived after time-domain deconvolution. Global features are explored, which, unlike pixel-level features, are computationally more straightforward and allow for easier image-level classification. An ensemble approach by combining the classification outputs from two classifiers trained on the phasor and bi-exponential features resulted in an accuracy of 88.33%, sensitivity of 84.21%, and specificity of 90.24%.

**Deep learning based classification of skin cancer lesions**: Demonstrated the utility of a recurrent deep-learning model, such as LSTM, for classifying benign and malignant skin cancer lesions. A pixel-level classification is obtained from the LSTM model, resulting in a posterior probability map for each imaged lesion. Thorough model evaluation and hyperparameter tuning are conducted with fifty test sets from ten different random

partitions of the data. The mean and standard deviation of the AUCs obtained from the ROCs in each partition are 0.82 and 0.04, respectively.

**Deep learning based classification of oral cancer lesions**: Developed a deep learning model using LSTM networks for classifying oral cancer lesions from healthy tissue. The pixel-level model output is transformed to an image-level classification by applying two thresholds on the posterior probability and the percentage of cancer pixels. This method yielded classification maps that help identify regions of malignancy. Thorough evaluation of the model performance using seventy test sets produced an overall sensitivity of 86.4% and specificity of 82%.

**Future work**:

The size of skin and oral cancer datasets are small as they are collected from only 30 and 34 patients respectively. To make realistic conclusions on the viability of the techniques explained in this dissertation, the experiments should be conducted on larger datasets by increasing the number of patients significantly. The dataset used for skin cancer lesions does not include data collected from the normal or healthy skin of the patients. Collecting this data from each patient in future studies will improve the training efficiency of the classification models. In addition, this work can be extended to data collected from maFLIM dermoscopy/endoscopy with dual excitation wavelength to excite the target fluorophores efficiently. The LSTM models developed for oral and skin cancer lesions need further evaluation compared with other machine learning algorithms. The deep learning models can be extended to input the raw measurements directly from the instrument, thus further reducing the preprocessing of the data. Additionally, the CAD

systems can be improved to process and classify skin or oral cancer lesions using the maFLIM signals acquired from the suspicious regions in real-time.

The overall cost of the time domain maFLIM system is around $62,000, which is primarily contributed by the pulsed laser and the MCP-PMT detector. In the future, an alternate system based on frequency domain maFLIM using continuous wave laser excitations and low-cost avalanche photodiode (APD) detectors can decrease the overall cost of the system significantly to around $25000. Other techniques that are either used or being developed for skin cancer pre-screening are the dermoscope, reflectance confocal microscope (RCM) and optical coherence tomography (OCT). Although dermoscope is not very expensive (~$1000) and often employed in routine clinical examinations, the information extracted from them is very limited. On the other hand, RCM provides useful in-vivo information in both lateral and axial dimensions, such systems are very expensive (~$80,000) and is only used in research and teaching hospitals. OCT is another expensive technique that helps to image the skin surface. The cost of the instrument ranges between $50,000 to $120,000. The commercially available device for oral cancer prescreening is the Velscope. Although the device cost only around $5000, it is not widely used in clinical settings owing to the low specificities. In comparison to these techniques, a future low cost version of the maFLIM endoscope will be highly beneficial to extract useful information on the metabolic as well as other biochemical changes in the tissue during malignant transformations.

REFERENCES

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer Statistics, 2021," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, Jan. 2021, doi: 10.3322/caac.21654.

[2] R. Shellenberger, M. Nabhan, and S. Kakaraparthi, "Melanoma screening: A plan for improving early detection," *Annals of Medicine*, vol. 48, no. 3. Taylor and Francis Ltd, pp. 142–148, Apr. 02, 2016. doi: 10.3109/07853890.2016.1145795.

[3] P. Aggarwal, P. Knabel, and A. B. Fleischer, "United States burden of melanoma and non-melanoma skin cancer from 1990 to 2019," *J Am Acad Dermatol*, vol. 85, no. 2, pp. 388–395, Aug. 2021, doi: 10.1016/j.jaad.2021.03.109.

[4] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *Lancet Oncology*, vol. 3, no. 3, pp. 159–165, 2002, doi: 10.1016/S1470-2045(02)00679-4.

[5] A. Masood and A. A. Al-Jumaily, "Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms," *International Journal of Biomedical Imaging*, vol. 2013, 2013, doi: 10.1155/2013/323268.

[6] A.-R. A. Ali and T. M. Deserno, "A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data," *Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment*, vol. 8318, no. February 2012, p. 83181I, 2012, doi: 10.1117/12.912389.

[7] K. Kamnitsas *et al.*, “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Medical Image Analysis*, vol. 36, pp. 61–78, 2017, doi: 10.1016/j.media.2016.10.004.

[8] P. Sahu, D. Yu, M. Dasari, F. Hou, and H. Qin, “A Lightweight Multi-Section CNN for Lung Nodule Classification and Malignancy Estimation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 960–968, 2019, doi: 10.1109/JBHI.2018.2879834.

[9] P. Calzavara-Pinton, C. Longo, M. Venturini, R. Sala, and G. Pellacani, “Reflectance confocal microscopy for in vivo skin imaging,” *Photochemistry and Photobiology*, vol. 84, no. 6, pp. 1421–1430, 2008, doi: 10.1111/j.1751-1097.2008.00443.x.

[10] L. T. Dengel *et al.*, “Total body photography for skin cancer screening,” *International Journal of Dermatology*, vol. 54, no. 11, pp. 1250–1254, 2015, doi: 10.1111/ijd.12593.

[11] B. Rosado *et al.*, “Accuracy of computer diagnosis of melanoma: A quantitative meta-analysis,” *Archives of Dermatology*, vol. 139, no. 3, pp. 361–367, 2003, doi: 10.1001/archderm.139.3.361.

[12] J. Abdul Jaleel and S. Salim, “Artificial Neural Network Based Detection of Skin Cancer,” 2012. [Online]. Available: www.ijareeie.com

[13] H. A. Haenssle *et al.*, “Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in

comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, Aug. 2018, doi: 10.1093/annonc/mdy166.

[14] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019 (US statistics)," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019, doi: 10.3322/caac.21551.

[15] P. Wilder-Smith *et al.*, "Noninvasive imaging of oral premalignancy and malignancy," *Advanced Biomedical and Clinical Diagnostic Systems III*, vol. 5692, no. October 2005, p. 375, 2005, doi: 10.1117/12.591170.

[16] J. J. Sciubba, "Improving detection of precancerous and cancerous oral lesions: Computer-assisted analysis of the oral brush biopsy," *Journal of the American Dental Association*, vol. 130, no. 10, pp. 1445–1457, 1999, doi: 10.14219/jada.archive.1999.0055.

[17] D. C. G. de Veld, M. J. H. Witjes, H. J. C. M. Sterenborg, and J. L. N. Roodenburg, "The status of in vivo autofluorescence spectroscopy and imaging for oral oncology," *Oral Oncology*, vol. 41, no. 2, pp. 117–131, 2005, doi: 10.1016/j.oraloncology.2004.07.007.

[18] J. Seoane, S. Warnakulasuriya, P. Varela-Centelles, G. Esparza, and P. Dios, "Oral cancer: Experiences and diagnostic abilities elicited by dentists in North-western Spain," *Oral Diseases*, vol. 12, no. 5, pp. 487–492, 2006, doi: 10.1111/j.1601-0825.2005.01225.x.

[19] J. B. Epstein, P. Güneri, H. Boyacioglu, and E. Abt, "The limitations of the clinical oral examination in detecting dysplastic oral lesions and oral squamous cell

carcinoma," *Journal of the American Dental Association*, vol. 143, no. 12, pp. 1332–1342, 2012, doi: 10.14219/jada.archive.2012.0096.

[20] M. C. Skala *et al.*, "In vivo multiphoton microscopy of NADH and FAD redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia," *Proc Natl Acad Sci U S A*, vol. 104, no. 49, pp. 19494–19499, 2007, doi: 10.1073/pnas.0708425104.

[21] N. P. Galletly *et al.*, "Fluorescence lifetime imaging distinguishes basal cell carcinoma from surrounding uninvolved skin," *British Journal of Dermatology*, vol. 159, no. 1, pp. 152–161, 2008, doi: 10.1111/j.1365-2133.2008.08577.x.

[22] Y. Wu and J. Y. Qu, "Autofluorescence spectroscopy of epithelial tissues," *Journal of Biomedical Optics*, vol. 11, no. 5, p. 054023, 2006, doi: 10.1117/1.2362741.

[23] H. Zeng, D. I. McLean, C. E. MacAulay, and H. Lui, "Autofluorescence properties of skin and applications in dermatology," in *Biomedical Photonics and Optoelectronic Imaging*, Oct. 2000, vol. 4224, pp. 366–373. doi: 10.1117/12.403940.

[24] E. G. Borisova *et al.*, "Autofluorescence spectroscopy of cutaneous neoplasia under ultraviolet, visible and near infrared excitation," no. April, p. 37, 2020, doi: 10.1117/12.2555982.

[25] P. Diagaradjane, M. A. Yaseen, J. Yu, M. S. Wong, and B. Anvari, "Autofluorescence characterization for the early diagnosis of neoplastic changes in

DMBA/TPA-induced mouse skin carcinogenesis," *Lasers in Surgery and Medicine*, vol. 37, no. 5, pp. 382–395, Dec. 2005, doi: 10.1002/lsm.20248.

[26] R. Yao, M. Ochoa, P. Yan, and X. Intes, "Net-FLICS: fast quantitative wide-field fluorescence lifetime imaging with compressed sensing – a deep learning approach," *Light: Science and Applications*, vol. 8, no. 1, pp. 1–7, 2019, doi: 10.1038/s41377-019-0138-x.

[27] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the us population, 2012," *JAMA Dermatology*, vol. 151, no. 10, pp. 1081–1086, Oct. 2015, doi: 10.1001/jamadermatol.2015.1187.

[28] Cancer Facts and Figures 2021, "American Cancer Society. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf. Accessed December 25, 2021."

[29] B. Lindelo È F, B. Sigurgeirsson, H. Ga, È. Bel[2], and R. S. Stern[3], "Incidence of skin cancer in 5356 patients following organ transplantation," 2000.

[30] C. L. Benjamin and H. N. Ananthaswamy, "p53 and the pathogenesis of skin cancer," *Toxicology and Applied Pharmacology*, vol. 224, no. 3. pp. 241–248, Nov. 01, 2007. doi: 10.1016/j.taap.2006.12.006.

[31] R. Gordon, "Skin Cancer: An Overview of Epidemiology and Risk Factors," *Seminars in Oncology Nursing*, vol. 29, no. 3, pp. 160–169, 2013.

[32] Z. Apalla, D. Nashan, R. B. Weller, and X. Castellsagué, "Skin Cancer: Epidemiology, Disease Burden, Pathophysiology, Diagnosis, and Therapeutic Approaches," *Dermatology and Therapy*, vol. 7, pp. 5–19, 2017, doi: 10.1007/s13555-016-0165-y.

[33] I. Zaqout, "Diagnosis methods of skin lesions in dermoscopic images: A survey," in *Proceedings - 2019 International Conference on Promising Electronic Technologies, ICPET 2019*, Oct. 2019, pp. 102–106. doi: 10.1109/ICPET.2019.00026.

[34] J. M. García-Martín, P. Varela-Centelles, M. González, J. M. Seoane-Romero, J. Seoane, and M. J. García-Pola, "Epidemiology of oral cancer," in *Oral Cancer Detection: Novel Strategies and Clinical Impact*, Springer International Publishing, 2019, pp. 81–93. doi: 10.1007/978-3-319-61255-3_3.

[35] Rivera Cesar, "Essentials of oral cancer," *Int J Clin Exp Pathol*, vol. 8, no. 9, pp. 11884–11894, 2015.

[36] P. M. Speight and P. M. Farthing, "The pathology of oral cancer," *British Dental Journal*, vol. 225, no. 9, pp. 841–847, Nov. 2018, doi: 10.1038/sj.bdj.2018.926.

[37] H. Ram, J. Sarkar, H. Kumar, R. Konwar, M. L. B. Bhatt, and S. Mohammad, "Oral Cancer: Risk Factors and Molecular Pathogenesis," *Journal of Maxillofacial and Oral Surgery*, vol. 10, no. 2. Springer, pp. 132–137, Jun. 01, 2011. doi: 10.1007/s12663-011-0195-z.

[38] D. Kademani, "Oral Cancer," *SYMPOSIUM ON SOLID TUMORS*, vol. 82, no. 7, pp. 878–887, 2007.

[39] J. R. Lakowicz, *Principles of fluorescence spectroscopy*. 2006. doi: 10.1007/978-0-387-46312-4.

[40] P. A. A. de Beule *et al.*, "A hyperspectral fluorescence lifetime probe for skin cancer diagnosis," *Review of Scientific Instruments*, vol. 78, no. 12, 2007, doi: 10.1063/1.2818785.

[41] I. Giovannacci, C. Magnoni, P. Vescovi, A. Painelli, E. Tarentini, and M. Meleti, "Which are the main fluorophores in skin and oral mucosa? A review with emphasis on clinical applications of tissue autofluorescence," *Archives of Oral Biology*, vol. 105. Elsevier Ltd, pp. 89–98, Sep. 01, 2019. doi: 10.1016/j.archoralbio.2019.07.001.

[42] Y. Wu, P. Xi, J. Y. Qu, T.-H. Cheung, and M.-Y. Yu, "Depth-resolved fluorescence spectroscopy reveals layered structure of tissue," *Optics Express*, vol. 12, no. 14, p. 3218, 2004, doi: 10.1364/opex.12.003218.

[43] E. Duran-Sierra *et al.*, "Clinical label-free biochemical and metabolic fluorescence lifetime endoscopic imaging of precancerous and cancerous oral lesions," *Oral Oncology*, vol. 105, Jun. 2020, doi: 10.1016/j.oraloncology.2020.104635.

[44] J. Blomfield and J. F. Farrar, "The Fluorescent Properties of Maturing Arterial Elastin," 1969. [Online]. Available: https://academic.oup.com/cardiovascres/article/3/2/161/286734

[45] J. Wizenty *et al.*, "Recent advances and the potential for clinical use of autofluorescence detection of extra-ophthalmic tissues," *Molecules*, vol. 25, no. 9. MDPI AG, May 01, 2020. doi: 10.3390/molecules25092095.

[46] J.-A. Grimaud, M. Druguet, S. Peyrol, O. Chevalier, D. Herbage, and N. el Badrawy, “Collagen Immunotyping in Human Liver: Light and Electron Microscope Study,” *The journal of histochemistry and cytochemistry*, vol. 28, pp. 1145–1156, 1980.

[47] I. Georgakoudi *et al.*, “NAD(P)H and Collagen as in Vivo Quantitative Fluorescent Biomarkers of Epithelial Precancerous Changes 1,” 2002. [Online]. Available: http://aacrjournals.org/cancerres/article-pdf/62/3/682/2500494/ch0302000682.pdf

[48] Y. Wu, P. Xi, J. Y. Qu, T.-H. Cheung, and M.-Y. Yu, “Depth-resolved fluorescence spectroscopy reveals layered structure of tissue,” *Optics Express*, vol. 12, no. 14, p. 3218, 2004, doi: 10.1364/opex.12.003218.

[49] O. I. Kolenc and K. P. Quinn, “Evaluating cell metabolism through autofluorescence imaging of NAD(P)H and FAD,” *Antioxidants and Redox Signaling*, vol. 30, no. 6. Mary Ann Liebert Inc., pp. 875–889, 2019. doi: 10.1089/ars.2017.7451.

[50] A.-M. Pena *et al.*, “Multiphoton FLIM imaging of NADH and FAD to analyze cellular metabolic activity of reconstructed human skin in response to UVA light,” Feb. 2019, p. 9. doi: 10.1117/12.2508858.

[51] Y. Yuan *et al.*, “Autofluorescence of NADH is a new biomarker for sorting and characterizing cancer stem cells in human glioma,” *Stem Cell Research and Therapy*, vol. 10, no. 1, Nov. 2019, doi: 10.1186/s13287-019-1467-7.

[52] E. Dimitrow *et al.*, “Spectral fluorescence lifetime detection and selective melanin imaging by multiphoton laser tomography for melanoma diagnosis,” *Experimental*

*Dermatology*, vol. 18, no. 6, pp. 509–515, 2009, doi: 10.1111/j.1600-0625.2008.00815.x.

[53] D. M. Harris and J. Werkhaven, "Endogenous Porphyrin Fluorescence in Tumors," 1987.

[54] R. T. Smith, N. L. Gomes, G. Barile, M. Busuioc, N. Lee, and A. Laine, "Lipofuscin and autofluorescence metrics in progressive STGD," *Investigative Ophthalmology and Visual Science*, vol. 50, no. 8, pp. 3907–3914, 2009, doi: 10.1167/iovs.08-2448.

[55] F. P. Koch, P. W. Kaemmerer, S. Biesterfeld, M. Kunkel, and W. Wagner, "Effectiveness of autofluorescence to identify suspicious oral lesions-a prospective, blinded clinical trial," *Clinical Oral Investigations*, vol. 15, no. 6, pp. 975–982, Dec. 2011, doi: 10.1007/s00784-010-0455-1.

[56] B. Lin *et al.*, "Characterizing the origin of autofluorescence in human esophageal epithelium under ultraviolet excitation," *Opt Express*, vol. 18, no. 20, pp. 21074–21082, 2010.

[57] M. G. vander Heiden, L. C. Cantley, and C. B. Thompson, "Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation," *Science (1979)*, vol. 324, pp. 1029–1033, 2009, [Online]. Available: https://www.science.org

[58] A. J. Walsh *et al.*, "Optical metabolic imaging identifies glycolytic levels, subtypes, and early-treatment response in breast cancer," *Cancer Research*, vol. 73, no. 20, pp. 6164–6174, Oct. 2013, doi: 10.1158/0008-5472.CAN-13-0527.

[59] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, “Cancer Statistics, 2021,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, Jan. 2021, doi: 10.3322/caac.21654.

[60] R. M. Bakos, T. P. Blumetti, R. Roldán-Marín, and G. Salerni, “Noninvasive Imaging Tools in the Diagnosis and Treatment of Skin Cancers,” *Am J Clin Dermatol*, vol. 19, no. s1, pp. 3–14, 2018, doi: 10.1007/s40257-018-0367-4.

[61] M. E. Celebi and A. Zornberg, “Automated quantification of clinically significant colors in dermoscopy images and its application to skin lesion classification,” *IEEE Systems Journal*, vol. 8, no. 3, pp. 980–984, 2014, doi: 10.1109/JSYST.2014.2313671.

[62] D. A. Shoieb, S. M. Youssef, and W. M. Aly, “Computer-Aided Model for Skin Diagnosis Using Deep Learning,” *Journal of Image and Graphics*, pp. 122–129, 2016, doi: 10.18178/joig.4.2.122-129.

[63] D. Adla, G. V. R. Reddy, P. Nayak, and G. Karuna, “Deep learning-based computer aided diagnosis model for skin cancer detection and classification,” *Distributed and Parallel Databases*, 2021, doi: 10.1007/s10619-021-07360-z.

[64] K. Ramlakhan and Y. Shang, “A mobile automated skin lesion classification system,” in *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence. IEEE*, 2011, pp. 138–141. doi: 10.1109/ICTAI.2011.29.

[65] T. Y. Satheesha, D. Satyanarayana, M. N. G. Prasad, and K. D. Dhruve, “Melanoma Is Skin Deep: A 3D Reconstruction Technique for Computerized Dermoscopic Skin Lesion Classification,” *IEEE Journal of Translational*

*Engineering in Health and Medicine*, vol. 5, no. December 2016, pp. 1–17, 2017, doi: 10.1109/JTEHM.2017.2648797.

[66] Y. A. Khristoforova *et al.*, "Optical diagnostics of malignant and benign skin neoplasms," *Procedia Engineering*, vol. 201, pp. 141–147, 2017, doi: 10.1016/j.proeng.2017.09.664.

[67] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J Biomed Inform*, vol. 86, pp. 25–32, 2018.

[68] A. Romero Lopez, X. Giro-I-Nieto, J. Burdick, and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," in *017 13th IASTED international conference on biomedical engineering (BioMed). IEEE*, 2017, pp. 49–54. doi: 10.2316/P.2017.852-053.

[69] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Combining deep learning and hand-crafted features for skin lesion classification," in *2016 6th International Conference on Image Processing Theory, Tools and Applications, IPTA IEEE*, 2016, pp. 1–6. doi: 10.1109/IPTA.2016.7821017.

[70] Y. C. Lee, S.-H. Jung, and H.-H. Won, "WonDerM: Skin Lesion Classification with Fine-tuned Neural Networks," Aug. 2018, [Online]. Available: http://arxiv.org/abs/1808.03426

[71] J. Amin *et al.*, "Integrated design of deep features fusion for localization and classification of skin cancer," *Pattern Recognition Letters*, vol. 131, pp. 63–70, Mar. 2020, doi: 10.1016/j.patrec.2019.11.042.

[72] M. F. Jojoa Acosta, L. Y. Caballero Tovar, M. B. Garcia-Zapirain, and W. S. Percybrooks, "Melanoma diagnosis using deep learning techniques on dermatoscopic images," *BMC Medical Imaging*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12880-020-00534-8.

[73] K. Awasthi, D. Moriya, T. Nakabayashi, L. Li, and N. Ohta, "Sensitive detection of intracellular environment of normal and cancer cells by autofluorescence lifetime imaging," *Journal of Photochemistry and Photobiology B: Biology*, vol. 165, pp. 256–265, Dec. 2016, doi: 10.1016/j.jphotobiol.2016.10.023.

[74] O. I. Kolenc and K. P. Quinn, "Evaluating cell metabolism through autofluorescence imaging of NAD(P)H and FAD," *Antioxidants and Redox Signaling*, vol. 30, no. 6. Mary Ann Liebert Inc., pp. 875–889, 2019. doi: 10.1089/ars.2017.7451.

[75] Q. Yang, M. Qi, Z. Wu, L. Liu, P. Gao, and J. Qu, "Classification of skin cancer based on fluorescence lifetime imaging and machine learning," in *Optics in Health Care and Biomedical Optics*, Oct. 2020, pp. 115531Y1-115531Y5. doi: 10.1117/12.2573851.

[76] R. Na, I.-M. Stender, L. Ma, and H. C. Wulf, "Autofluorescence spectrum of skin: component bands and body site variations," *Skin Research and Technology*, vol. 6, pp. 112–117, 2000.

[77] N. P. Galletly *et al.*, "Fluorescence lifetime imaging distinguishes basal cell carcinoma from surrounding uninvolved skin," *British Journal of Dermatology*, vol. 159, no. 1, pp. 152–161, Jul. 2008, doi: 10.1111/j.1365-2133.2008.08577.x.

[78] J. P. Miller, L. Habimana-Griffin, T. S. Edwards, and S. Achilefu, " Multimodal fluorescence molecular imaging for in vivo characterization of skin cancer using endogenous and exogenous fluorophores ," *Journal of Biomedical Optics*, vol. 22, no. 6, p. 066007, Jun. 2017, doi: 10.1117/1.jbo.22.6.066007.

[79] E. Drakaki, E. Borisova, M. Makropoulou, L. Avramov, A. A. Serafetinides, and I. Angelov, "Laser induced autofluorescence studies of animal skin used in modeling of human cutaneous tissue spectroscopic measurements," *Skin Research and Technology*, vol. 13, no. 4, pp. 350–359, Nov. 2007, doi: 10.1111/j.1600-0846.2007.00237.x.

[80] M. N. Pastore, H. Studier, C. S. Bonder, and M. S. Roberts, "Non-invasive metabolic imaging of melanoma progression," *Experimental Dermatology*, vol. 26, no. 7, pp. 607–614, Jul. 2017, doi: 10.1111/exd.13274.

[81] W. Lohmann and R. H. Bodeker, "In Situ Differentiation Between Nevi and Malignant Melanomas by Fluorescence Measurements," *Naturwissenschaften*, vol. 78, pp. 456–457, 1991.

[82] R. A. Romano, R. G. Teixeira Rosa, A. G. Salvio, J. A. Jo, and C. Kurachi, "Multispectral autofluorescence dermoscope for skin lesion assessment," *Photodiagnosis and Photodynamic Therapy*, vol. 30, Jun. 2020, doi: 10.1016/j.pdpdt.2020.101704.

[83] R. Henderson and K. Schulmeister, *Laser safety*, vol. 53, no. 9. 2004. doi: 10.1017/CBO9781107415324.004.

[84] D. U. Campos-Delgado, O. G. Navarro, E. R. Arce-Santana, and J. A. Jo, "Extended output phasor representation of multi-spectral fluorescence lifetime imaging microscopy," *Biomedical Optics Express*, vol. 6, no. 6, p. 2088, Jun. 2015, doi: 10.1364/boe.6.002088.

[85] D. Kim *et al.*, "Phasor plot analysis using low pass filter for high-speed FLIM," no. February, p. 24, 2019, doi: 10.1117/12.2509325.

[86] T. Luo, Y. Lu, S. Liu, D. Lin, and J. Qu, "Phasor-FLIM as a Screening Tool for the Differential Diagnosis of Actinic Keratosis, Bowen's Disease, and Basal Cell Carcinoma," *Analytical Chemistry*, vol. 89, no. 15, pp. 8104–8111, 2017, doi: 10.1021/acs.analchem.7b01681.

[87] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 94, pp. 1119–1125, 1994, doi: 10.1016/S0167-8655(99)00083-5.

[88] J. Hua, Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty, "Optimal number of features as a function of sample size for various classification rules," *Bioinformatics*, vol. 21, no. 8, pp. 1509–1515, Apr. 2005, doi: 10.1093/bioinformatics/bti171.

[89] E. Duran-Sierra *et al.*, "Machine-learning assisted discrimination of precancerous and cancerous from healthy oral tissue based on multispectral autofluorescence lifetime imaging endoscopy," *Cancers (Basel)*, vol. 13, no. 19, Oct. 2021, doi: 10.3390/cancers13194751.

[90] M. A. Digman, V. R. Caiolfa, M. Zamai, and E. Gratton, “The phasor approach to fluorescence lifetime imaging analysis,” *Biophysical Journal*, vol. 94, no. 2, Jan. 2008, doi: 10.1529/biophysj.107.120154.

[91] B. K. Wright, L. M. Andrews, M. R. Jones, C. Stringari, M. A. Digman, and E. Gratton, “Phasor-flim analysis of NADH distribution and localization in the nucleus of live progenitor myoblast cells,” *Microscopy Research and Technique*, vol. 75, no. 12, pp. 1717–1722, Dec. 2012, doi: 10.1002/jemt.22121.

[92] N. Ma, M. A. Digman, L. Malacrida, and E. Gratton, “Measurements of absolute concentrations of NADH in cells using the phasor FLIM method,” *Biomedical Optics Express*, vol. 7, no. 7, p. 2441, Jul. 2016, doi: 10.1364/boe.7.002441.

[93] U. O. Dorj, K. K. Lee, J. Y. Choi, and M. Lee, “The skin cancer classification using deep convolutional neural network,” *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9909–9924, Apr. 2018, doi: 10.1007/s11042-018-5714-1.

[94] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, “Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, Apr. 2017, doi: 10.1109/TMI.2016.2642839.

[95] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, “Classification of skin disease using deep learning neural networks with mobilenet V2 and LSTM,” *Sensors*, vol. 21, no. 8, Apr. 2021, doi: 10.3390/s21082852.

[96] P. Vasanthakumari *et al.*, “AI-driven discrimination of benign from malignant pigmented skin lesions based on multispectral autofluorescence lifetime

dermoscopy imaging," in *Proc. SPIE 11934, Photonics in Dermatology and Plastic Surgery 2022*, Mar. 2022, p. 1193408. doi: 10.1117/12.2610345.

[97] H. C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016, doi: 10.1109/TMI.2016.2528162.

[98] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013, doi: 10.1109/TPAMI.2012.59.

[99] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "3D convolutional neural networks for efficient and robust hand pose estimation from single depth images," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5679–5688, 2017, doi: 10.1109/CVPR.2017.602.

[100] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, Mar. 2020, doi: 10.1016/j.physd.2019.132306.

[101] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent Neural Networks and Robust Time Series Prediction," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 240–254, 1994, doi: 10.1109/72.279188.

[102] K. Kaur and S. K. Mittal, "Classification of mammography image with CNN-RNN based semantic features and extra tree classifier approach using LSTM," *Materials Today: Proceedings*, Oct. 2020, doi: 10.1016/j.matpr.2020.09.619.

[103]A. Tasdelen and B. Sen, “A hybrid CNN-LSTM model for pre-miRNA classification,” *Scientific Reports*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-93656-0.

[104]Ş. Öztürk and U. Özkaya, “Gastrointestinal tract classification using improved LSTM based CNN,” *Multimedia Tools and Applications*, vol. 79, no. 39–40, pp. 28825–28840, Oct. 2020, doi: 10.1007/s11042-020-09468-3.

[105]S. Karimi Jafarbigloo and H. Danyali, “Nuclear atypia grading in breast cancer histopathological images based on CNN feature extraction and LSTM classification,” *CAAI Trans Intell Technol*, vol. 6, no. 4, pp. 426–439, Dec. 2021, doi: 10.1049/cit2.12061.

[106]R. A. Romano, R. G. T. Rosa, J. A. Jo, and C. Kurachi, “Label-free multispectral lifetime fluorescence to distinguish skin lesions,” *Proceedings of SPIE*, vol. 10890, 2019, doi: 10.1117/12.2510498.

[107]P. Vasanthakumari *et al.*, “Classification of skin-cancer lesions based on Fluorescence Lifetime Imaging,” in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, Feb. 2020, p. 34. doi: 10.1117/12.2548625.

[108]P. Le and W. Zuidema, “Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive LSTMs,” Mar. 2016, [Online]. Available: http://arxiv.org/abs/1603.00423

[109]Y. Gal and Z. Ghahramani, “A Theoretically Grounded Application of Dropout in Recurrent Neural Networks.”

[110]G. Cheng, V. Peddinti, D. Povey, V. Manohar, S. Khudanpur, and Y. Yan, "An exploration of dropout with LSTMs," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2017, vol. 2017-August, pp. 1586–1590. doi: 10.21437/Interspeech.2017-129.

[111]N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[112]L. Li, K. Jamieson, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization," *Journal of Machine Learning Research*, vol. 18, pp. 1–52, 2018, [Online]. Available: http://jmlr.org/papers/v18/16-558.html.

[113]S. R. Bisht, P. Mishra, D. Yadav, R. Rawal, and K. P. Mercado-Shekhar, "Current and emerging techniques for oral cancer screening and diagnosis: A review," *Progress in Biomedical Engineering*, vol. 3, no. 4. IOP Publishing Ltd, Oct. 01, 2021. doi: 10.1088/2516-1091/ac1f6c.

[114]M. M. A. Chiamulera, C. B. Zancan, A. P. Remor, M. F. Cordeiro, F. O. Gleber-Netto, and A. R. Baptistella, "Salivary cytokines as biomarkers of oral cancer: a systematic review and meta-analysis," *BMC Cancer*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12885-021-07932-3.

[115]Y. H. Wu, P. Y. Lin, J. H. Yang, Y. S. Kuo, and Y. C. Wu, "Serum levels and positive rates of tumor biomarkers in oral precancer patients," *Journal of the*

*Formosan Medical Association*, vol. 120, no. 6, pp. 1324–1331, Jun. 2021, doi: 10.1016/j.jfma.2020.12.013.

[116]P. Madki, M. L. A. Tejasvi, G. Paramkusam, R. Khan, and S. J., "Evaluation of Serum Immunoglobulins (IgG, IgA, IgM) and Circulating Immune Complexes in Oral Precancer and Cancer Patients," *Global Medical Genetics*, vol. 08, no. 03, pp. 095–099, Sep. 2021, doi: 10.1055/s-0041-1725157.

[117]S. J. Silverman, "Early Diagnosis of Oral Cancer," *Cancer*, vol. 62, no. S1, pp. 1796–1799, 1988.

[118]M. Aubreville *et al.*, "Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning," *Scientific Reports*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-12320-8.

[119]B. Bhandari, A. Alsadoon, P. W. C. Prasad, S. Abdullah, and S. Haddad, "Deep learning neural network for texture feature extraction in oral cancer: enhanced loss function," *Multimedia Tools and Applications*, vol. 79, no. 37–38, pp. 27867–27890, Oct. 2020, doi: 10.1007/s11042-020-09384-6.

[120]B. Song *et al.*, "Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning," *Biomedical Optics Express*, vol. 9, no. 11, p. 5318, Nov. 2018, doi: 10.1364/boe.9.005318.

[121]B. Song *et al.*, "Bayesian deep learning for reliable oral cancer image classification," *Biomedical Optics Express*, vol. 12, no. 10, p. 6422, Oct. 2021, doi: 10.1364/boe.432365.

[122]J. Folmsbee, X. Liu, M. Brandwein-Weber, and S. Doyle, “Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer,” in *Proceedings - International Symposium on Biomedical Imaging*, May 2018, vol. 2018-April, pp. 770–773. doi: 10.1109/ISBI.2018.8363686.

[123]M. Halicek *et al.*, “Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging,” *Journal of Biomedical Optics*, vol. 22, no. 6, p. 060503, Jun. 2017, doi: 10.1117/1.jbo.22.6.060503.

[124]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*, vol. 145, no. 4, pp. 829–837, Apr. 2019, doi: 10.1007/s00432-018-02834-7.

[125]H. Lin, H. Chen, L. Weng, J. Shao, and J. Lin, “Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis,” *Journal of Biomedical Optics*, vol. 26, no. 08, Aug. 2021, doi: 10.1117/1.jbo.26.8.086007.

[126]B. R. Nanditha, A. G. Kiran, H. S. Chandrashekar, M. S. Dinesh, and S. Murali, “An Ensemble Deep Neural Network Approach for Oral Cancer Screening,” *International journal of online and biomedical engineering*, vol. 17, no. 2, pp. 121–134, 2021, doi: 10.3991/ijoe.v17i02.19207.

[127]R. A. Welikala *et al.*, “Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer,” *IEEE Access*, vol. 8, pp. 132677–132693, 2020, doi: 10.1109/ACCESS.2020.3010180.

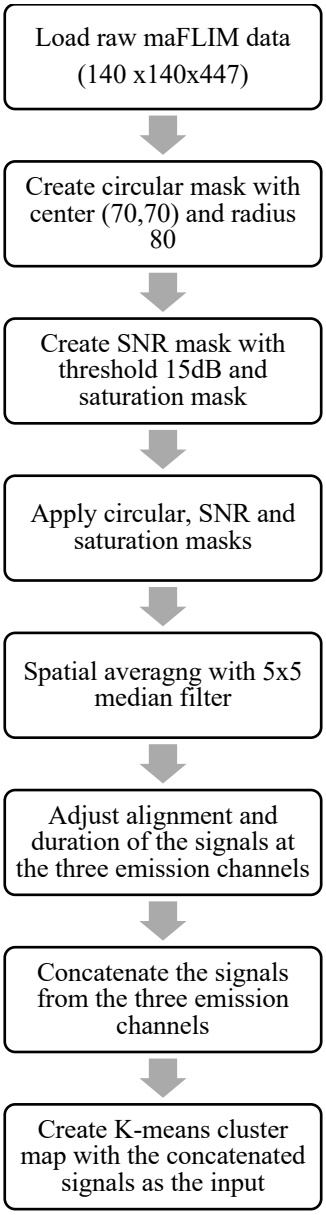
[128]S. Xu *et al.*, “An Early Diagnosis of Oral Cancer based on Three-Dimensional Convolutional Neural Networks,” *IEEE Access*, vol. 7, pp. 158603–158611, 2019, doi: 10.1109/ACCESS.2019.2950286.

[129]M. Marsden *et al.*, “Intraoperative Margin Assessment in Oral and Oropharyngeal Cancer Using Label-Free Fluorescence Lifetime Imaging and Machine Learning,” *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 3, pp. 857–868, Mar. 2021, doi: 10.1109/TBME.2020.3010480.

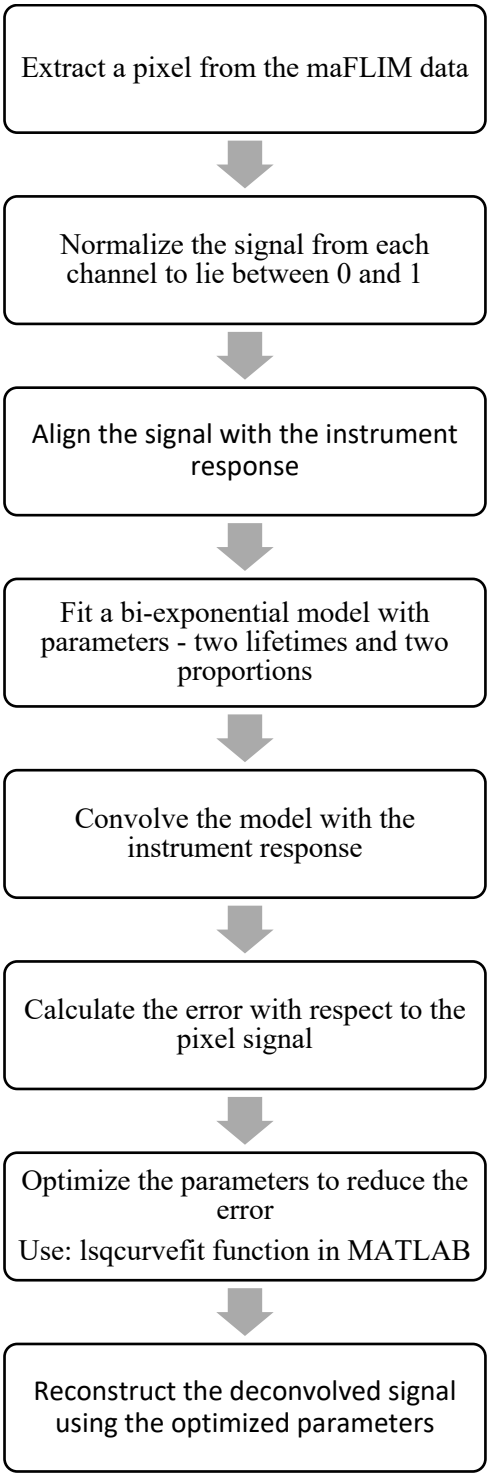
# APPENDIX

## Chapter – 3

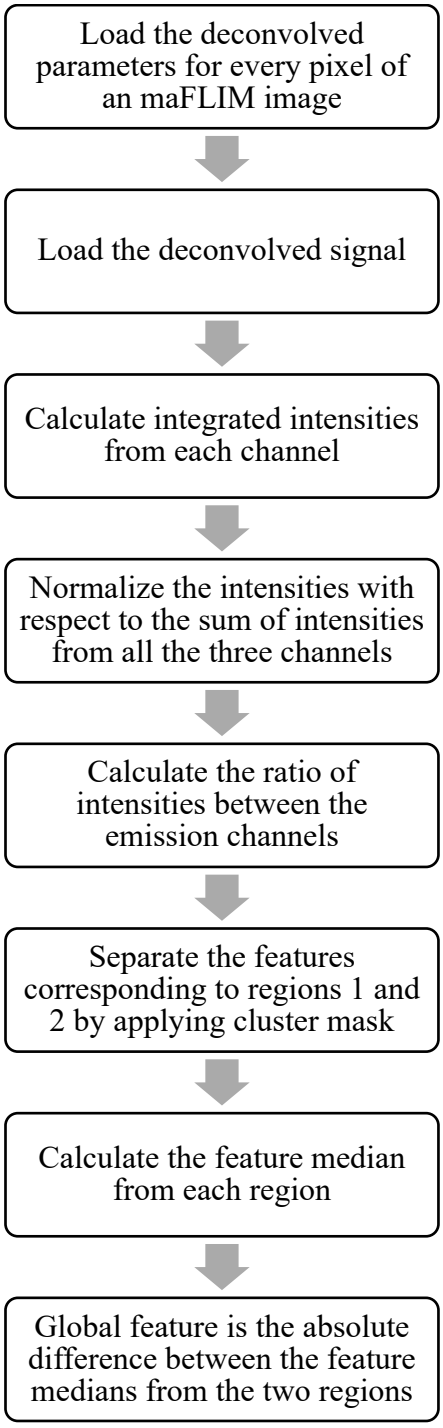
### Pre-processing of maFLIM data



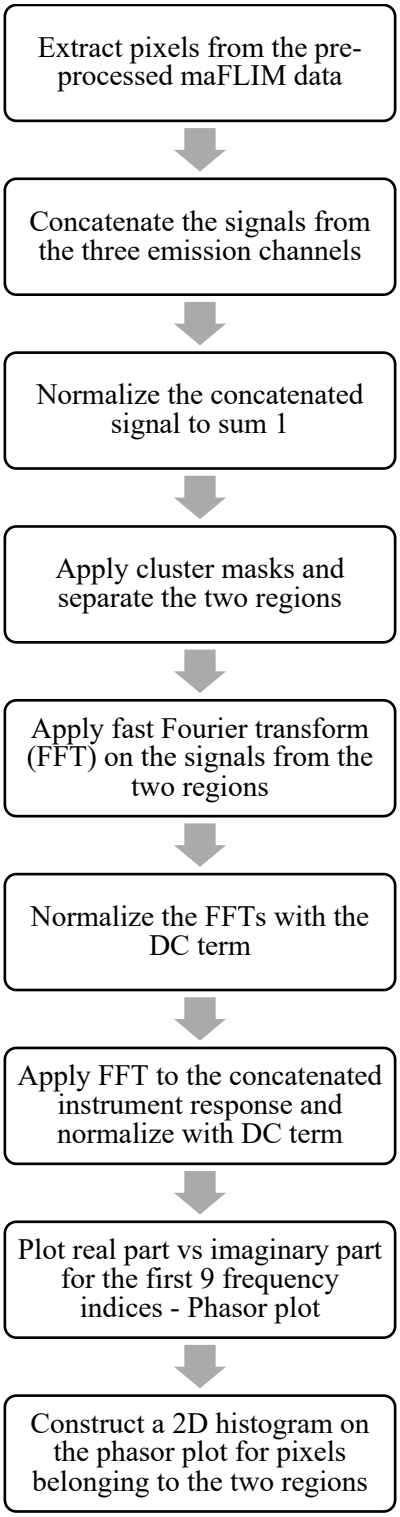
Time domain deconvolution



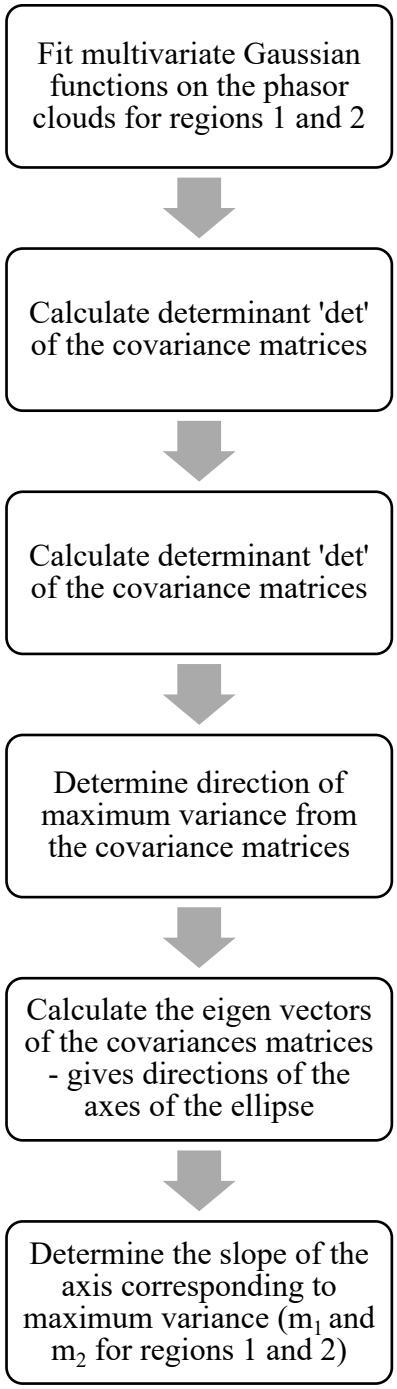
Time domain global feature extraction

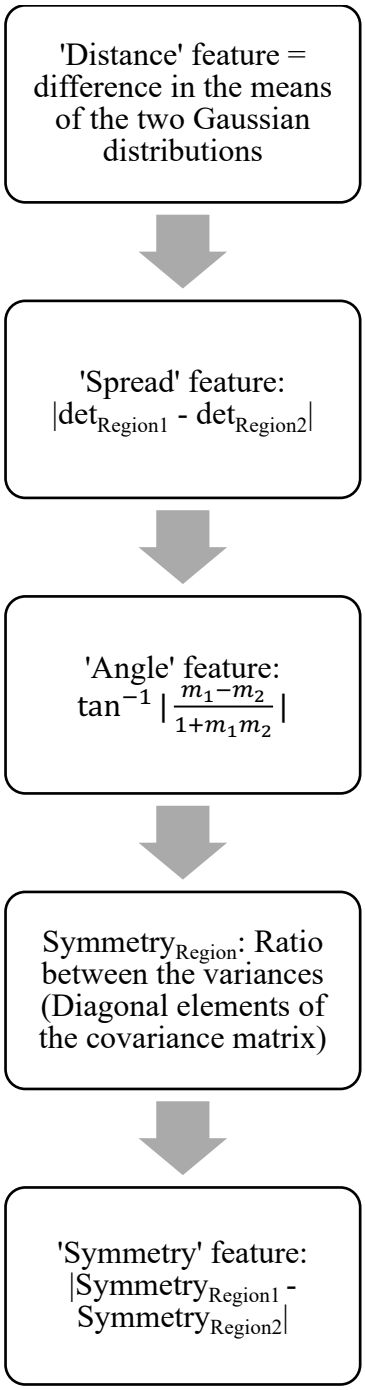


Frequency domain deconvolution and phasor construction



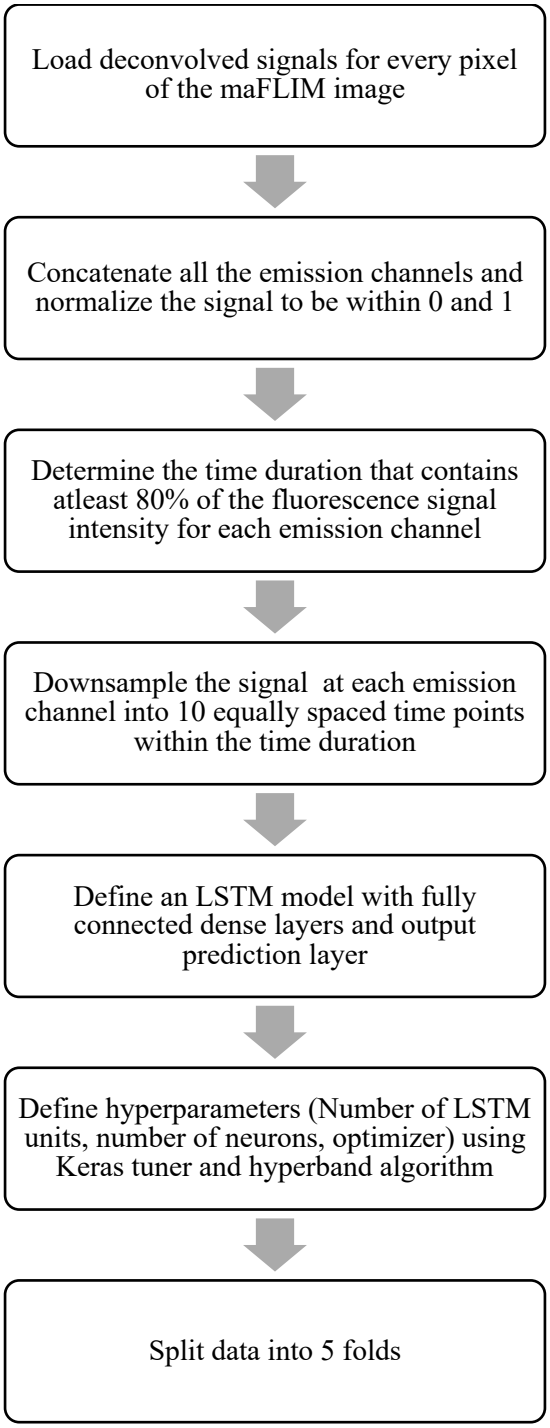
Phasor based feature extraction



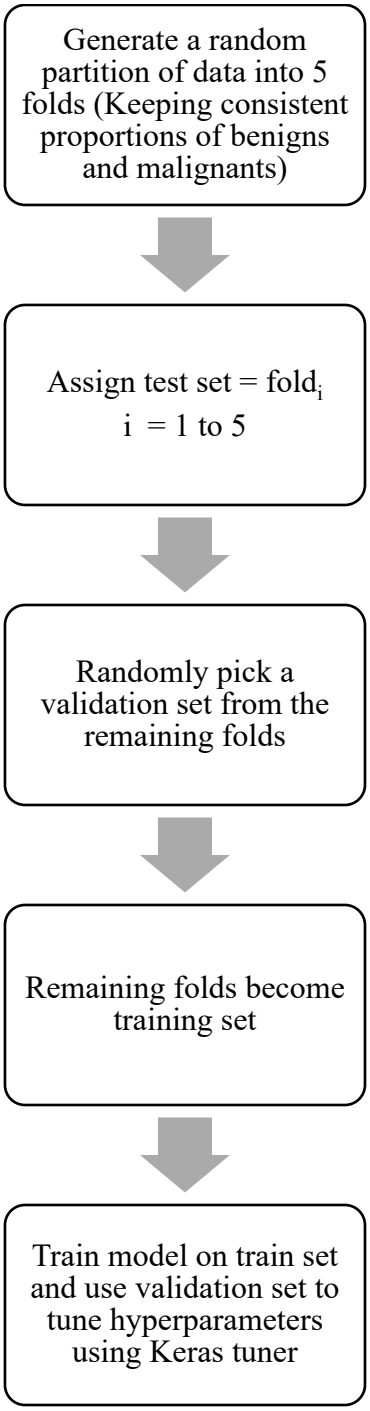

'Distance' feature = difference in the means of the two Gaussian distributions
'Spread' feature: $|det_{Region1} - det_{Region2}|$
'Angle' feature: $\tan^{-1}\left|\frac{m_1 - m_2}{1 + m_1 m_2}\right|$
$Symmetry_{Region}$: Ratio between the variances (Diagonal elements of the covariance matrix)
'Symmetry' feature: $|Symmetry_{Region1} - Symmetry_{Region2}|$

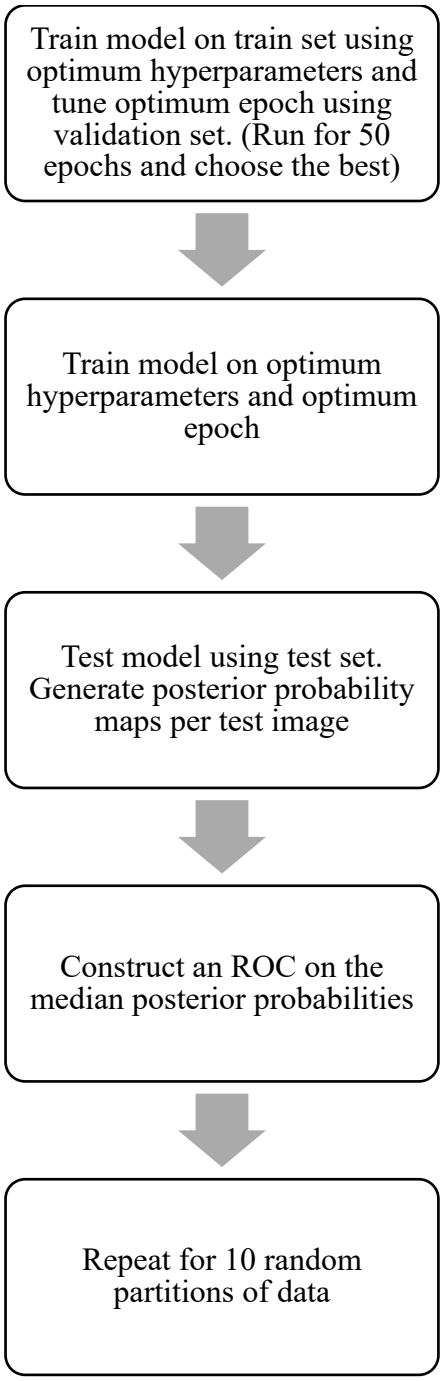
Chapter – 4

Pre-processing and model definition



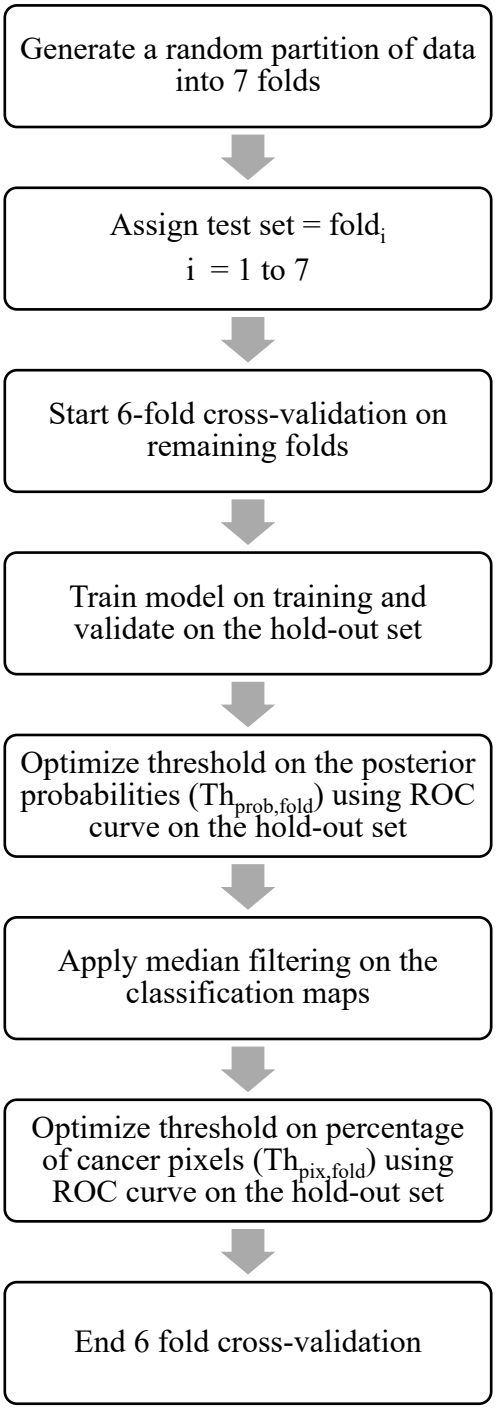
Workflow

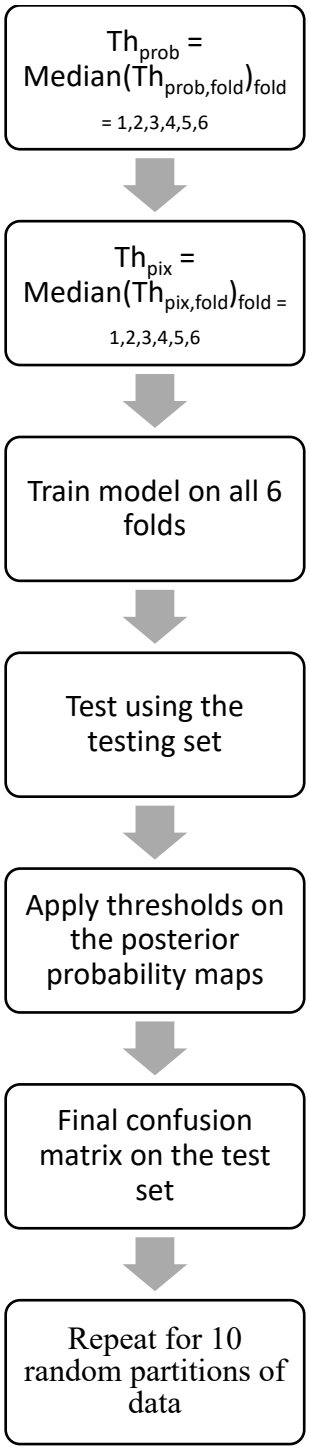

Generate a random partition of data into 5 folds (Keeping consistent proportions of benigns and malignants)
Assign test set = fold$_i$
i = 1 to 5
Randomly pick a validation set from the remaining folds
Remaining folds become training set
Train model on train set and use validation set to tune hyperparameters using Keras tuner


Train model on train set using optimum hyperparameters and tune optimum epoch using validation set. (Run for 50 epochs and choose the best)
Train model on optimum hyperparameters and optimum epoch
Test model using test set. Generate posterior probability maps per test image
Construct an ROC on the median posterior probabilities
Repeat for 10 random partitions of data

Chapter – 5

Workflow




$Th_{prob}$ = Median($Th_{prob,fold}$)$_{fold}$ = 1,2,3,4,5,6
$Th_{pix}$ = Median($Th_{pix,fold}$)$_{fold}$ = 1,2,3,4,5,6
Train model on all 6 folds
Test using the testing set
Apply thresholds on the posterior probability maps
Final confusion matrix on the test set
Repeat for 10 random partitions of data