

Methods and Applications in Fluorescence



TOPICAL REVIEW

OPEN ACCESS

RECEIVED
30 June 2023

REVISED
25 September 2023

ACCEPTED FOR PUBLICATION
6 December 2023

PUBLISHED
8 February 2024

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Applications of machine learning in time-domain fluorescence lifetime imaging: a review

Dorian Gouzou^{1,*} , Ali Taimori² , Tarek Haloubi² , Neil Finlayson³ , Qiang Wang⁴ , James R Hopgood²  and Marta Vallejo¹ 

¹ Dorian Gouzou and Marta Vallejo are with Institute of Signals, Sensors and Systems, School of Engineering and Physical Sciences, Heriot Watt University, Edinburgh, EH14 4AS, United Kingdom

² Tarek Haloubi, Ali Taimori, and James R. Hopgood are with Institute for Imaging, Data and Communication, School of Engineering, University of Edinburgh, Edinburgh, EH9 3FG, United Kingdom

³ Neil Finlayson is with Institute for Integrated Micro and Nano Systems, School of Engineering, University of Edinburgh, Edinburgh EH9 3FF, United Kingdom

⁴ Qiang Wang is with Centre for Inflammation Research, University of Edinburgh, Edinburgh, EH16 4TJ, United Kingdom

* Author to whom any correspondence should be addressed.

E-mail: dg67@hw.ac.uk, ataimori@ed.ac.uk, tarek.haloubi@ed.ac.uk, n.finlayson@ed.ac.uk, q.wang@ed.ac.uk, james.hopgood@ed.ac.uk and m.vallejo@hw.ac.uk

Keywords: machine learning, fluorescence lifetime imaging, FLIm, deep learning, biomedical engineering

Abstract

Many medical imaging modalities have benefited from recent advances in Machine Learning (ML), specifically in deep learning, such as neural networks. Computers can be trained to investigate and enhance medical imaging methods without using valuable human resources. In recent years, Fluorescence Lifetime Imaging (FLIm) has received increasing attention from the ML community. FLIm goes beyond conventional spectral imaging, providing additional lifetime information, and could lead to optical histopathology supporting real-time diagnostics. However, most current studies do not use the full potential of machine/deep learning models. As a developing image modality, FLIm data are not easily obtainable, which, coupled with an absence of standardisation, is pushing back the research to develop models which could advance automated diagnosis and help promote FLIm. In this paper, we describe recent developments that improve FLIm image quality, specifically time-domain systems, and we summarise sensing, signal-to-noise analysis and the advances in registration and low-level tracking. We review the two main applications of ML for FLIm: lifetime estimation and image analysis through classification and segmentation. We suggest a course of action to improve the quality of ML studies applied to FLIm. Our final goal is to promote FLIm and attract more ML practitioners to explore the potential of lifetime imaging.

1. Introduction

1.1. Fluorescent lifetime imaging

Fluorescence is a form of light emission produced from the shift to excited states happening subsequently to the absorption of photon energy from an external light source. The phenomenon exhibits emission and excitation spectra characterised by different metrics such as intensity and lifetime. Lifetime is the average time between excitation and return to the ground state. Fluorescence Lifetime Imaging (FLIm) is a valuable bio-imaging approach for investigating the behaviour and interaction of very small-scale objects, such as molecular and cellular microorganisms. In this

technique, a sample is first excited by the focused light of a laser. The electrons inside the sample get stimulated and move from a calm, ground state to an excited state. Then, the fluorophore molecules fluoresce. The fluorophore photons are captured by a detector and the temporal responses are captured across different optical wavelengths and used to create the FLIm images by characterising the fluorescence lifetime from their decay profile [1].

FLIm devices have important, diverse applications, including diagnosis and prognosis, where it has shown great success in identifying tumours and delimiting them, such as oral cancer [2, 3] or breast tumour [4, 5], a task for which FLIm is a candidate,

among other technologies, to replace or complement histopathologic analysis. More broadly, it is also utilised for drug monitoring, obtaining penetration of micro- and nano-particles information, visualising topical drug uptake *in vivo* [6–8], disease treatment [9, 10] or other biomedical tasks such as sperm assessment [11].

As opposed to fluorescence spectroscopy or other imaging techniques, FLIm does not produce an image directly. For each ‘pixel’, the lifetime is estimated from the characterised decay curve, often for multiple spectral channels, as discussed in section 2.6. In FLIm, lifetime information obtained at each pixel complements per-pixel measurements of fluorescence intensity, enabling differentiation of image regions which cannot be obtained from intensities. FLIm sensing can be carried out in both time [12] and frequency domains [13], which differ in terms of hardware and software aspects. Time-domain FLIm systems use pulsed lasers for excitation and exploit transient temporal responses to estimate the fluorescence lifetime with the aim of fast imaging, whereas frequency-domain FLIm utilises sinusoidal excitation and measures the phase shift of the wave [14] to determine the lifetime indirectly. This review narrows its focus more on developed time-domain systems.

1.2. Artificial intelligence and machine learning

Artificial Intelligence (AI) refers to the ensemble of methods aiming to make computers solve problems. It is divided into different categories such as natural language processing, evolutionary computation or Machine Learning (ML). ML encompasses a set of techniques designed to make a machine ‘learn’ from data how to perform one or multiple given tasks [15]. Deep Learning (DL) is a subset of ML which makes use of a large amount of data to train Artificial Neural Network (ANN) with multiple layers, alternating linear computation and non-linear activation function, the adjective deep denoting the difference with ANN used in ML which do not have intermediary layers.

ML and DL are very powerful tools for data analysis [16], and promoted the progress of many complex tasks such as text processing and translation [17, 18] or picture captioning [19] through multimodal learning. Another successful area of applicability is medical and biomedical imaging. Examples of such use are Magnetic Resonance Imaging (MRI) [20], Electroencephalography (EEG) [21], Optical Coherence Tomography (OCT) and OCT-Angiography [22, 23], biomedical optics modalities [24] or generally for task such as segmentation [25].

ML includes a wide variety of algorithms, also called ‘models’, trained in a standardised way. The most common and easiest training type is supervised learning, which includes two or three stages: (1) training, in which the model is exposed to the data to learn

its patterns; (2) testing, when the trained model is exposed to new, unseen data and is evaluated on how well it performs the assigned task; and (3) validation, sometimes not performed, which is used to select the best model, either among different ML modalities or by selecting the best hyper-parameters for a specific model. The evaluation methods are problem-dependent. On the other side, in unsupervised learning, the models make use of unlabelled data to make their own representation of them, either to group them by similarity or to mimic them [26].

Now, it is widely considered that ML-based techniques will be key to addressing many medical and biomedical challenges and improving diagnostic techniques. Specifically, FLIm benefits from progress in ML, whether by reducing the computation load needed to characterise lifetime or by automated processing of FLIm images which are not humanly understandable today. Our aim is to address this important subject and review researches that fall into this category in a structured manner as schematised in figure 1. We separated signal processing and raw data studies from image analysis, according to the evolution of the data through the multiple processing stages. Figure 2 also gives a general view to readership how output images of a FLIm system on a sample look like.

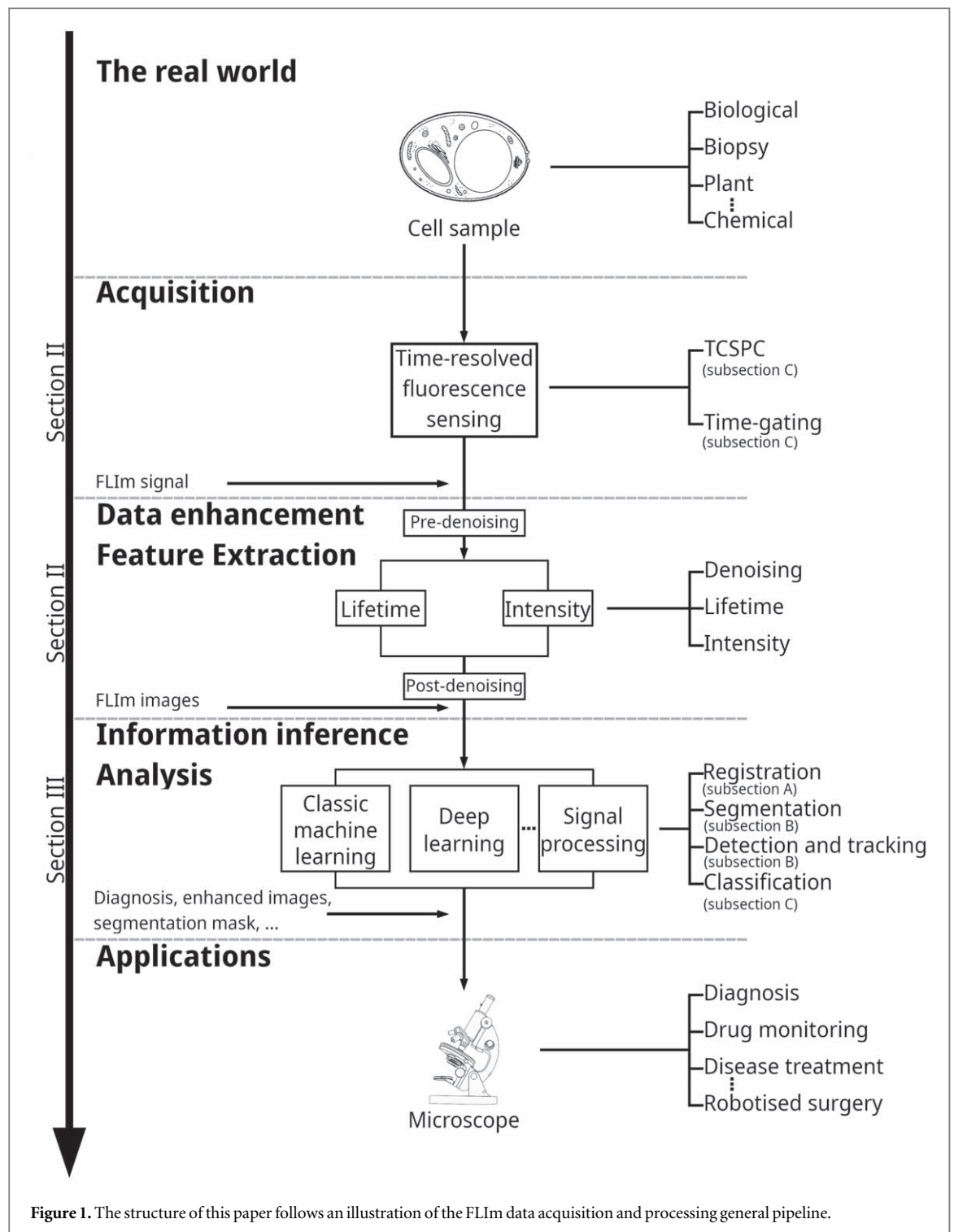
1.3. The purpose and synergy of this paper

In the literature of FLIm, several valuable review papers explain the underlying physical processes and applications of FLIm exist [9, 13, 27–32]. Recently, a few researches mentioned some applications of ML to FLIm [24, 33–35]. We aim to complement those studies by providing an up-to-date, detailed and structured review exclusively focused on FLIm.

This review highlights the importance of ML in medical research, summarises the current state of the ML applications in FLIm and focuses on recent advances made in FLIm since those reported by Marcu [36] and Becker [28] a decade ago. In this paper, we will review applications of regression and classification problems for lifetime estimation and mostly disease classification, including cancer. Regression uses criteria such as bias and variance for evaluating estimation efficiency, whereas classification employs metrics like sensitivity and accuracy to evaluate how well a model classifies the unknown element or differentiates the different classes under investigation.

1.4. Paper organisation

The rest of this paper is organised as indicated in figure 1. Section 2 addresses FLIm data acquisition and enhancement, as well as the FLIm feature extraction and ongoing studies of ML-based model for lifetime estimation. It presents FLIm to beginners in this field and introduces important resources. Section 3 reviews research in FLIm images analysis. It explains to ML



researchers what the different ML tasks are in FLIm, summarises current methodologies and provides advice for more advanced ML studies. Finally, section 4 concludes the paper by remarking on key challenges and opportunities in the field.

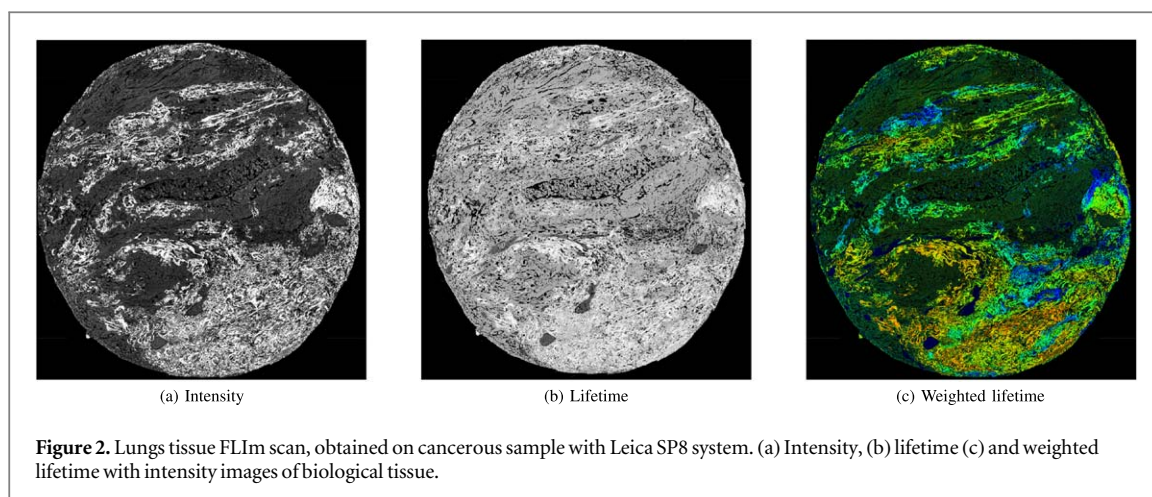
2. Fluorescence lifetime image acquisition

2.1. An early history of FLIm related research

The foundations and applications of biomedical photonics, fluorescence spectroscopy, and Time-Correlated

Single Photon Counting (TCSPC) supporting time-resolved fluorescence are described in seminal texts by Chance *et al* [37, 38], Lakowicz *et al* [1, 39].

Fluorescence spectroscopy and imaging often employ targeted fluorescent labels or probes called exogenous fluorophores. Another approach is the investigation of auto-fluorescence tissue properties through endogenous fluorophores. For example, the coenzymes Nicotinamide Adenine Dinucleotide (NADH) and Flavin Adenine Dinucleotide (FAD) enable label-free detection of metabolic changes and



interpretation of cellular metabolic states [40], and are extensively used in FLIm images analysis with ML. FLIm has also shown that it is able to detect, at pixel resolution, abnormal cellular metabolism through the level of activation of oxidative phosphorylation measured as an increase in the fraction of protein-bound Nicotinamide Adenine Dinucleotide (NAD), yielding important information about cell responses in metabolic diseases and cancer [41, 42].

In 1962, Chance *et al* reported a study where fluorescence emission was used to study the localized metabolic response of tissue to variation in oxygen concentration, interpreting the cellular metabolic state in terms of the relative amount of oxidised and reduced form of NADH [37]. However, the emission spectra of many tissue fluorophores overlap. FLIm is a technique which enables such overlapping fluorophores to be distinguished by their lifetime. Examples of such techniques are given by Lakowicz *et al* [39], who used the FLIm technique to exploit the high information content of Time-Resolved Fluorescence (TRF) in a 1992 study of free and protein-bound NADH metabolites. Subsequently, Skala *et al* [43] investigated early stages of cancer development through a combination of cellular redox ratio, NADH and FAD lifetime, and subcellular morphology *in vivo* imaging [44].

Naturally, early devices of FLIm employed traditional approaches of computing. However, with the current progress in computational power, a new generation of them is equipped with AI and ML.

2.2. Foundation

Confocal beam scanning is a microscopy technique first patented by Minsky [45] and later developed by Petrán *et al* [46], Davidovits and Egger [47] and Sheppard and Choudhury [48]. The technique uses object and image plane pinholes to reject unwanted scattered background light, building an image pixel-by-pixel with a raster scan typically using scanning galvanometer mirrors and is now used widely in FLIm applications. TCSPC-based FLIm utilising a laser

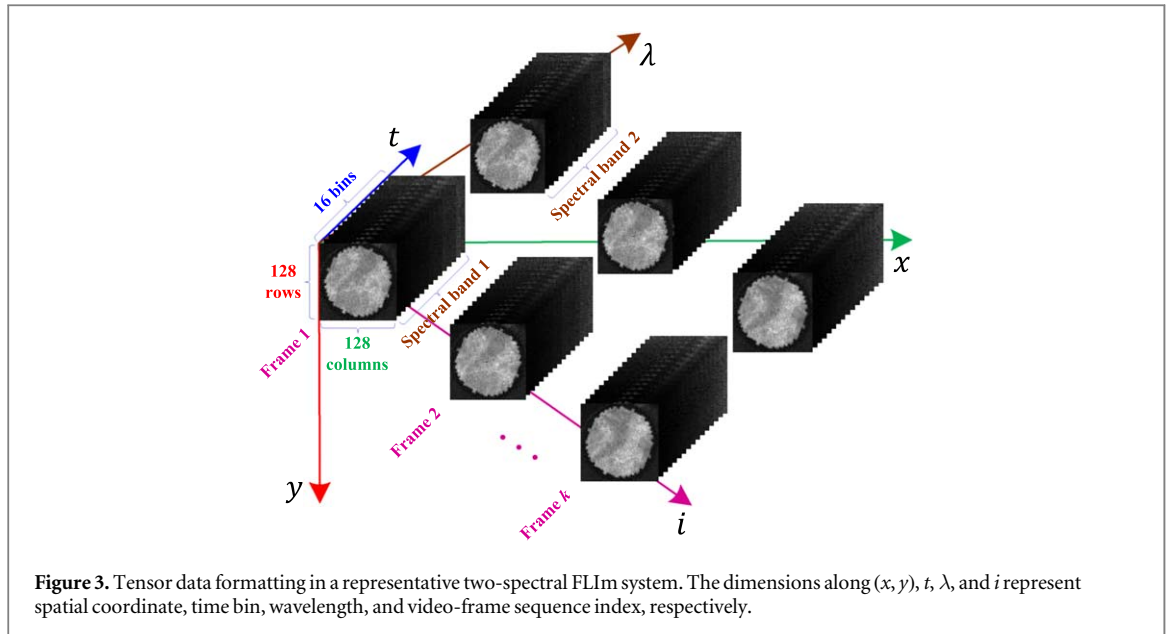
confocal scanning microscope was first described by Bugiel *et al* [49]. Wide-field FLIm in contrast is a technique where time-resolved signals in all pixels are acquired in parallel [50] yielding greater frame rates at the cost of increased background light levels.

A key goal for clinical research applications is *in vivo* FLIm, where video rates of frame acquisition are desirable to avoid motion artefacts. In 2005, Munro *et al* [51] demonstrated high-speed video-rate FLIm through a flexible endoscope, pointing the way to clinical applications of such techniques. Wide-field FLIm was exploited to achieve high acquisition rates based on gated optical image intensifier technology. Elson *et al* [52] investigated forward- and side-viewing FLIm fibre bundles comprised of 10,000 coherent fibres.

For fluorophores requiring excitation in the UV range, it is necessary to utilise two-photon excitation [53]. NADH for example absorbs light in the 335–350 nm range, with an emission peak around 440–470 nm [54].

TCSPC originated in 1960 with Bollinger and Thomas [55]. Multidimensional TCSPC signals as a function of wavelength, polarisation and other quantities became possible with multiple detector channels in 1993, and TCSPC acquisition times became 100 times faster. TCSPC became combined with microscopy [56]. The last two decades have seen major advances in detector technologies. Single-Photon Avalanche Diode (SPAD) arrays emerged in standard Complementary metal-oxide-semiconductor (CMOS) technology in 2003 [57] and time-resolved SPAD detection is now one of the most important and rapidly advancing technologies enabling advances in FLIm [52].

Comprehensive reviews of fluorescent lifetime techniques in medical applications with an emphasis on translational research potential supporting *in vivo* characterization and diagnosis of disease are provided by Marcu [36] and Alfonso-Garcia *et al* [58]. Other reviews include Suhling *et al* [29] and Liu *et al* [27]. Marcu provides a detailed overview of time-resolved fluorescence techniques in biomedical diagnostics and



the development of Time-Resolved Fluorescence Spectroscopy (TRFS) and FLIm instrumentation and methodologies. A key concern is evaluating whether intrinsic label-free endogenous fluorescence signals from biological tissues provide sufficient contrast for the diagnosis of cancers of the gastrointestinal tract, lung, head and neck, brain, skin and eye disease, and cardiovascular disease. Marcu lists 63 *in vivo* autofluorescence studies in humans, classified by organ investigated, method, excitation and emission ranges, power levels and numbers of patients.

Other significant reviews include those on FLIm, reviewed by Berezin and Achilefu [9], König [59], Datta *et al* [30] and Liu *et al* [27]; on wide-field FLIm Hirvonen and Suhling [50]; on the role of optical fibre technologies in clinical imaging of lung cancer Fernandes *et al* [40]; on fluorescence-guided surgery Stewart and Birch [60]; on CMOS SPAD time-resolved detection technologies, reviewed by Bruschini *et al* [61], Gyongy *et al* [62]; and, of most relevance to this review, on ML analysis of biomedical images Cao *et al* [63].

2.3. Fluorescence lifetime imaging sensors

SPAD arrays, developed in standard CMOS solid-state technology since 2003 [57], support a host of biophotonics applications [61]. SPAD detectors offer single photon spectroscopy and imaging capabilities with unparalleled photon counting and time-resolved performance when compared to CCDs/-sCMOS imagers. Among the advantages of CMOS SPAD arrays [61, 62] are highly parallel single-photon counting and photon timing, integration with standard digital blocks for data acquisition and/or processing, absence of read-out noise, very high frame rate binary implementations for real-time capture of fast transient phenomena and time-gating/binning to support Raman spectral acquisition in the presence of strong

fluorescence [64, 65]. A key advantage of CMOS integration is that smart sensor architectures become possible, which include on-chip, and indeed in-pixel, timestamping, histogramming and signal processing functionality.

SPAD devices leverage decades of investment in scalable, high-yield CMOS technologies and are exhibiting rapid improvements in key parameters such as Photon Detection Probability (PDP) and fill factor, dark count rates (DCR), spatial resolution (quarter megapixel, with megapixel arrays on the horizon), shrinkage of pixel size to well below $10 \mu\text{m}$ and excellent timing accuracy (typically 50-100 ps, with the best SPADs in the 20-30 ps range).

A CMOS SPAD line sensor for FLIm application was described by Erdogan *et al* [66]. This sensor supports 512 spectral channels, single photon counting and on-chip per-pixel TCSPC histogramming with up to 32 time bins. The sensor was subsequently used for FLIm [12] and combined fluorescence lifetime/Raman spectroscopy [65]. To visualise the hardware output format, figure 3 shows the arrangement of raw data in a FLIm system with the capability of generating multi-spectral video sequences. The data representation is considered as a 5D tensor arrangement consisting of a matrix of cubes as (x, y, t, λ, i) . The dimensions along (x, y, t, λ, i) denote spatial coordinate, bin time, wavelength and frame sequence index, respectively. In *ex vivo* microscopy applications, the sensor generates very large $x \times y \times t \times \lambda = 256 \times 256 \times 16 \times 512$ cube per frame, which facilitate detailed investigation of the dispersion of fluorescence lifetime in tissue and Raman and fluorescence spectroscopic correlations. Biophotonics applications include (endoscopic) FLIm, (multi-beam multiphoton) FLIm-Förster resonance energy transfer (FRET), SPIM-FCS, super-resolution microscopy, time-resolved Raman NIROT and positron emission tomography.

ML associated with SPADs is receiving increased attention [64, 67]. In [64], a fluorescence-suppressed Raman spectrometer utilising 16256 time-resolved CMOS SPADs was employed with K-means, an unsupervised clustering model to obtain high-quality images of human teeth. K-means is a low-complexity algorithm, easy to implement but which provides good results. It was used to identify regions in the images without clinical labelling. While it is effective and perfect for proof of concept work as this one, recent approaches in DL such as unsupervised Convolutional Neural Network (CNN) for segmentation, promise improvements in image and data quality [68].

2.4. Spatio-temporal-spectral sensing and early-stage processing

Similar techniques can be adopted in both frequency- and time-domain for measuring the physical parameter of fluorescence lifetime. E.g, the Fourier space approach of phasors [69, 70]. Focusing on time-domain systems, two basic forms of time-determined fluorescence (aka TRF) can be described, namely time-gating [71] and TCSPC [72]. From a data analysis perspective, FLIm systems designed by the above techniques mainly differ in terms of temporal resolution and video frame rate [14]. Generally, the temporal resolution (or the time bin resolution) in TCSPC is higher than time-gating [73]. However, the frame rate of the time-gating technique is typically higher than that of TCSPC. Other key attributes of TCSPC are low-noise and wide bandwidth characteristics [72], which make it utilisable in a broad range of applications [65].

In FLIm, a point of a target specimen is excited by a focused laser, yielding emission spectra extending typically within visible light range. The first emitted photon from an excited location that can reach the sensing equipment is detected, and then its arrival time is measured. Repeating the process for a given measurement period leads to a set of timestamps in each spectral channel. Before determining the lifetime, the observations must be pre-processed by histogramming of the timestamps. In this process, a number of bins or time gates are available with a given temporal resolution. Subsequently, timestamps falling into any individual bin interval are counted. Finally, an estimation of the distribution of arrived photons (called histogram of photon counts) is calculated. The whole step may be done on-chip in the detection unit [66] or on a destination computer in order to generate a histogram from photon counting. A scanning for all points of the sample by the above procedure provides spatial information.

Both time-gating and TCSPC techniques can output the histogram of photon counts as a function representing the fluorescence decay curve. This temporal sensing may be only a time series from a single point on a sample, e.g., a resultant reactional response

obtained from a chemical sample in spectroscopy applications [65], or pixel-by-pixel imaging of a biological tissue section [12]. Regardless of the details behind the electronics of the FLIm system, what is important for machine learners and signal processing experts is the formatting of its output data, where algorithms import the raw data for diverse purposes ranging from signal processing tasks to visual feature extraction.

2.5. Lifetime benchmark dataset

One important issue in FLIm is the availability of experimental datasets from known reference fluorophores. Benchmark datasets facilitate measuring the efficiency of a lifetime estimator in terms of mean, as a measure of accuracy or bias, standard deviation, as a measure of precision, and Signal-to-Noise Ratio (SNR), as a measure of signal strength. Chemical dyes or generally any samples with a plain property on image plane [74] can be appropriate options for a lifetime benchmark dataset. Fluorescein, Rhodamine B, Coumarin 6, Lucifer Yellow and 9-Cyanoanthracene are prominent fluorescent dyes [75, 76]. Partial information about the behaviour of fluorophores is already available, which helps in calibration. However, it should be noted that they are not sufficient for training (deep) supervised approaches, where multifluorophore data with a range of lifetime values are required. As will be discussed in the next section, learning-based methods often rely on synthetic data to solve the problem of lifetime estimation [77, 78].

2.6. Lifetime estimation

The histogram of photon counts is recorded for each pixel of a FLIm system. The fluorescence lifetime, as an important molecular bio-marker representing the decay rate of the histogram function, is estimated from the photon measurements. Figure 4 illustrates the process of lifetime estimation of a pixel. Estimated lifetimes over all pixels discriminate objects and regions of interest in molecular environments, where they cannot otherwise be easily distinguished on the basis of intensity of emitted photons. Based on the behaviour of fluorophores, the shape of the function may be mathematically modelled by a single- [74], double- [79], tri- [80], or generally multi-exponential decay curve. E.g., the function:

$$x[n] = A[\underbrace{\alpha}_{\neq \alpha_1} e^{-\frac{\Delta n}{\tau_1}} + \underbrace{(1 - \alpha)}_{\neq \alpha_2} e^{-\frac{\Delta n}{\tau_2}}], \forall n = 0, 1, \dots, N - 1 \quad (1)$$

represents a double- or bi-exponential model, in which $A \in \mathbb{R}^+$, $0 < \alpha < 1$, $\tau_1 \in \mathbb{R}^+$ and $\tau_2 \in \mathbb{R}^+$ represent initial amplitude, pre-exponential factor, short and long lifetimes, respectively. The parameters Δ and N denote the time bin width and the number of histogram bins, respectively. The symbol \mathbb{R}^+ represents the set of all positive real numbers. Setting $\alpha = 1$ in equation (1) reduces it to a single- or mono-exponential function. In practice, the measurements

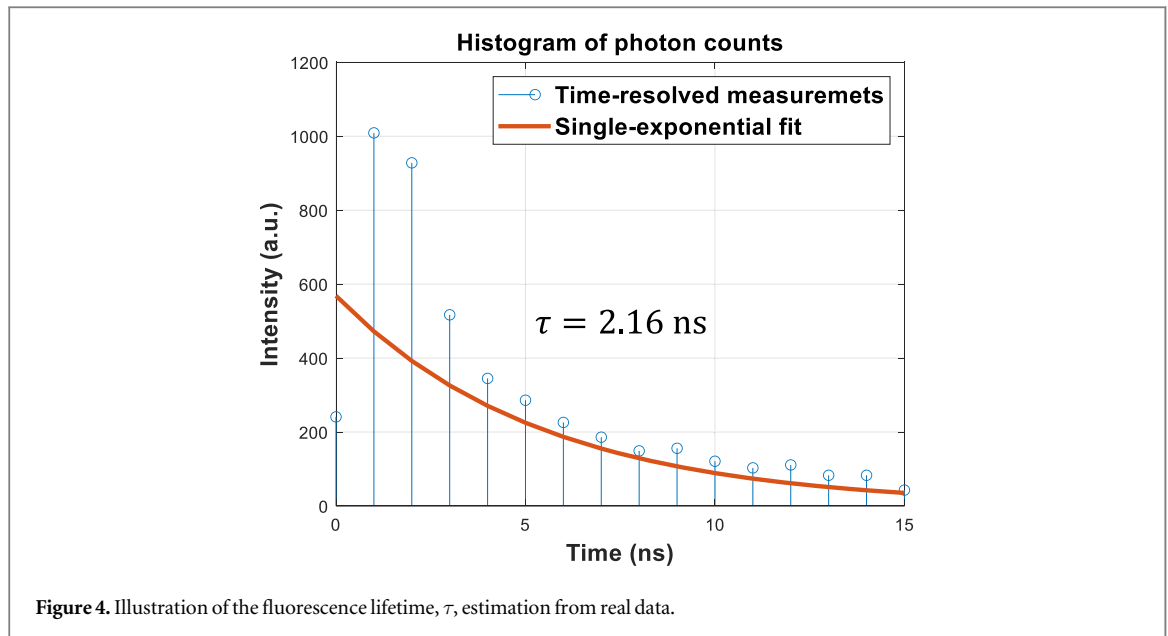


Figure 4. Illustration of the fluorescence lifetime, τ , estimation from real data.

are affected by noise from different sources as well as deformation in the histogram due to the convolution of a given model and the Instrument Response Function (IRF) [74]:

$$I[n] = [x[n]*h[n] + \eta[n]], \quad (2)$$

where the round operator of $[\cdot]$ delivers quantised photon counts, the symbol $*$ is convolution operator, and functions $I[n]$, $h[n]$ and $\eta[n]$ signify fluorescence intensity, IRF and additive noise term, respectively. At a given wavelength λ , the intensity in terms of photons per histogram in sensing or photons per pixel (p/p) in imaging can simply be measured by summing photons across all bins as:

$$I_\lambda \triangleq \sum_{n=0}^{N-1} I[n]. \quad (3)$$

These perturbations to the underlying signal make the problem of lifetime estimation a complicated task, especially in low photon count regimes with the intensity I_λ around 100 to 200 photons [78]. For the model presented in (1), it can be shown that the expected (mean) lifetime is:

$$\tau_{\text{mean}} = \frac{\alpha\tau_1^2 + (1 - \alpha)\tau_2^2}{\alpha\tau_1 + (1 - \alpha)\tau_2}. \quad (4)$$

A considerable number of researchers measures a special weighted averaging formula as $\tau_{\text{avg}} \triangleq \alpha\tau_1 + (1 - \alpha)\tau_2$ instead, which has been already flagged to be incorrect in Lakowicz's book [1]. The task of determining (4) can be translated back into a parameter estimation problem in regression or time series analysis. Different approaches have been proposed for lifetime estimation in the FLIm literature, including those enumerated by a recent review of [33] published in 2022. In this section, we provide a complementary critical review focusing more on detailed technical characteristics of developed AI- and ML-based approaches.

We divide the methods described in the literature into two broad categories of unsupervised and supervised methods. It should be noted that in information analysis/inference tasks of FLIm, any lifetime estimator from the above-mentioned types can be potentially employed as will be seen in section 3.3. Supervised lifetime estimators exploit learning data for pre-training, whereas unsupervised approaches do not require such data. Unsupervised methods can be grouped into two subcategories based on their nature, namely generic non parametric [81, 82] and statistical signal processing techniques [76, 83–91]. Generic processing approaches such as Rapid Lifetime Determination (RLD) [81] and RLD with overlapping windows [82] seek closed-form mathematical solutions to the fluorescence lifetime, whereas statistical ones such as Least Squares (LS), Maximum Likelihood Estimation (MLE) and Centre of Mass Method (CMM) are constructed upon estimation theory assuming a parametric model [92]. In this regard, optimised statistical estimators in LS [83, 84] and MLE [86–88] families are known as gold standard techniques of lifetime estimation. In Laguerre expansion-based lifetime estimation techniques [76, 91], the histogram of photon counts in each pixel is modelled by a linear combination of Laguerre basis functions. In fact, each of the two single-frequency exponential components in a bi-exponential model can be represented as Laguerre bases. Then, the problem of decay rate determination in the bi-exponential function is converted to statistically estimating corresponding Laguerre parameters. Statistical approaches in [89, 90] tried to jointly estimate the IRF of $h[n]$ and fluorescence lifetime components from time-resolved data using iterative methods of extended Kalman filter and expectation maximisation, respectively. Unsupervised approaches have the potential to be used in the hardware realisation of FLIm systems [33], so long as they support real-time

processing constraints such as existing approaches in RLD [74, 81, 82] or center of mass estimation [85]. However, a number of researchers have recently shifted the custom paradigms of lifetime estimation to learning-based mechanisms to specifically profit from the past decade of progress in DL methods, leading to the second supervised category [77–80, 93–96]. Supervised approaches target the function approximation capability of artificial Neural Network (NN) available in both subcategories of shallow and deep networks for fluorescence lifetime estimation. One of the outstanding deep learning models widely employed for estimating the lifetime [79, 80, 95, 96] is CNN. The architecture of a CNN consists of two main block of: (1) convolutional layers, as a representation learner; and, (2) Fully Connected (FC) decision layers, as a classifier/regressor. The representation block learns a set of dimensionality-reduced, influential features from raw input data in an automatic manner in connection with the classification/regression block. The classifier/regressor trains networks' weights based on desired parameters. Table 1 tabulates important characteristics of AI- and ML-based lifetime estimators from a set of selected references. Specifically, the table describes each approach in terms of utilised architecture, input/output data format, training/test dataset splitting, decay model, noise model, learning/test speed and performance metrics.

The seminal research of Wu *et al* [93] developed in 2016 was the first attempt towards supervised learning-based lifetime estimation. Their authors trained a shallow Multi-Layer Perceptron (MLP) with two hidden layers, where histograms of photon count having 57 bins are directly fed to the network as input features. The assumed fluorescence decay model is bi-exponential. In [79], a DL-based lifetime estimation was developed via a 3D CNN architecture. The authors considered the input of the network as a 3D tensor of (x, y, t) . Their network topology consisted of an initial pedestal branch for extracting temporal correlations and three subsequent parallel branches for spatial filtering and regression purposes. Outputs of the network encompass three 2D maps including images of amplitude and two lifetime components of a bi-exponential fluorescence decay model. Training data of the CNN was generated from synthesised temporal histograms inserted into white-pixel locations of a binarised version of the well-known MNIST character recognition dataset. Zickus *et al* [77] and Xiao *et al* [80] have suggested other FC NN and CNN architectures for lifetime estimation, where their specifications are summarised in table 1. Other authors used similar techniques [94–96]. E.g., a FC NN with the aim of reducing complexity than deep schemes and improving estimation efficiency (i.e., bias and variance) was developed in [94]. It extracts four features from each temporal histogram (including 2 phasor parameters) and approximates a bi-exponential model, hence constituting a 4-input, 11-hidden layer (6 neurons per

layer), and 3-output architecture. Similar CNN-based architectures in [95, 96] were proposed for reconstructing both intensity and lifetime images from a compressive sensing-based single-pixel FLIm acquisition system [97]. Chen *et al* [78] and Taimori *et al* [74] employed game theoretic approaches to deal with challenging scenarios of photon-starved (50 p/p) and Low Photon Count (LPC) regimes. The former exploits a Generative Adversarial Network (GAN) architecture at its heart to be able to generate high-quality temporal data by using the learning distribution of histogram of photon counts generated from a set of real samples via a generative model training. The latter proposed a Robust RLD algorithm that models the lifetime estimation problem in the presence of both noise and blurring perturbations as a signal recovery. It represents the histogram of photon counts as a multi-resolution framework, where perturbed measurements are denoised via adaptive smoothing mechanisms.

Here, as a critical conclusion, ML-based lifetime estimation versions substantially improved standard metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), bias and variance of the fluorescence lifetime estimates in comparison to custom baseline techniques like RLD [81, 82], CMM [85], LS [83, 84] and MLE [87, 88]. Among all, very promising results are seen for the game-theoretic approaches, especially on LPC regimes [74, 78] as a sign of further attention. Nonetheless, challenges yet exist for the realisation of state-of-the-art methods in industrial microscopes and endomicroscopes. For instance, handling of the problem on a very wide range of physical parameters (e.g., imaging samples ranging from sub-nanosecond to hundreds of nanosecond lifetimes) in a completely controlled manner, fully automating and hardware embedding are issues that will demand more research and development in future towards mature lifetime estimators.

2.7. Lifetime signal to noise ratio

In fluorescence, the lifetime is denoted by τ . The fluorescence lifetime error (standard deviation) $\Delta\tau$ and SNR as:

$$\text{Lifetime SNR} = \frac{\tau}{\Delta\tau} \quad (5)$$

ultimately determine the sensitivity and image contrast achievable in FLIm. To apprise SNR, the measure F -value $\triangleq \frac{\sqrt{N}\Delta\tau}{\tau}$ is also employed in the literature [98], in which the parameter N denotes the number of detected photons and F -value ≥ 1 . In 1992, Köllner and Wolfrum [99] addressed the important question of how many time-domain photon detections are required to perform measurements of fluorescent lifetime with a given accuracy. The required number of photons for a desired variance is:

Table 1. A summary overview of important characteristics of developed AI- and ML-based lifetime estimators.

Methodology	Architecture	Data format		Dataset		Decay	Noise model	Speed metric		Performance
		Input	Output	Training	Test			Learning	Run-time ^a	
MLP [93]	2 hidden layers	57-bin histograms	2 pre-exponential coefficients 2 lifetime terms	.21k histograms	Synthesised histograms Real images	Bi-exp	Poisson	4 h	180 ×	Success Rate = 99.93%
CNN [79]	Pedestal branch for extracting temporal correlations 3 subsequent parallel branches for spatial filtering and regressing	3D tensor (x, y, t)	Three 2D maps including images of amplitude 2 lifetime components	8k random data ^b	2k validation data Real images	Mono-exp Bi-exp	Poisson	Hours	~94k × ^c	MAE = .083 ns (LPC)
FCNN [77]	3 hidden layers with (100, 50, 25) nodes ReLU activation function	.2 bbin decays ^d	Lifetime	2M histograms	1M synthesised histograms Actual experimental data	Mono-exp	Poisson + Gauss	38 min	1k ×	MSE = .0053 ns ²
1DCNN [80]	A main branch including 2 residual networks 3 subsequent parallel CNNs	N-tuple 256 bins	3, N × 1 arrays	32k decays	8k decay curves	Bi-exp ^e	Poisson	.5 h	300 ×	F-val = (35.5, 5, 79.8) ^f
GAN [78]	G: 2 convolutional + pooling layers and 3 FC layers ^g D: 3 FC hidden layers E: 2 FC NNs + a concatenation layer + a hidden layer FCNN	G: 256 decay points, 256 IRF points D: 256 decay points E: 256 decay points, 256 IRF points	G: 256 outputs D: 1 neuron E: 3 neurons including amplitude and 2 lifetime components	.99k samples	Synthesised samples 2 experimental labeled sets of 256 × 256 or 512 × 512 images with 256 time bins	Bi-exp	Not reported	G + D = 6.1 h	258 ×	MSE = .21 ns ² (50 p/p)
Robust RLD [74]	Adaptive multi-bin parallel lifetime estimators A game-theoretic ‘amplitude-lifetime’ fuser	Photons’ histogram	Recovered decay	Unsupervised	Synthesised histograms Real 128 × 128 FLIm images with 16 bins	Mono-exp	Poisson + Gauss	0	~63 ×	Acc. ~28% + (LPC)

^a This metric shows the amount of run-time improvement than the standard LS.

^b The authors inserted a histogram of photon counts into the well-known MNIST character recognition database.

^c This approach is also about 30 times faster than SPCImage commercial software.

^d The authors also trained and tested another neural network for 30-bin decay curves. Decays in this research are represented as normalised forms.

^e For tri-exponential decay model, the research has also been extended to another specific architecture.

^f The triple in F-val is (τ₁, τ₂, α).

^g Here, the acronyms of G, D and E stand for Generator, Discriminator and Estimator models, respectively.

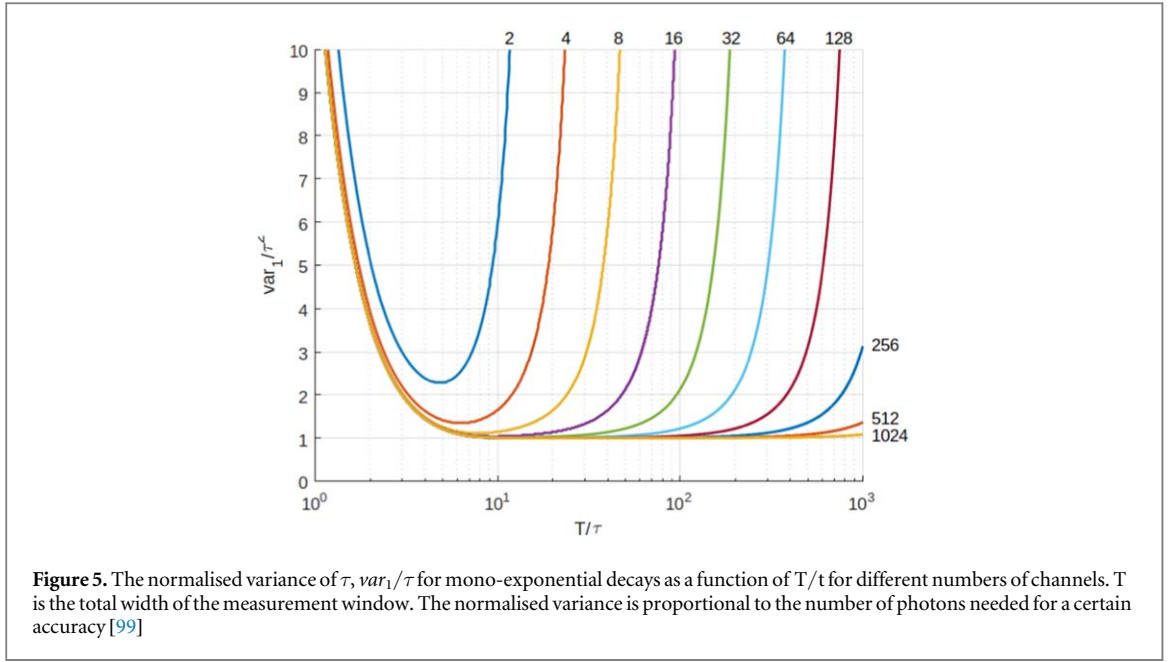


Figure 5. The normalised variance of τ , var_1/τ^2 for mono-exponential decays as a function of T/t for different numbers of channels. T is the total width of the measurement window. The normalised variance is proportional to the number of photons needed for a certain accuracy [99]

$$N \geq \frac{var_1(\tau)}{\text{desired } var(\tau)} \quad (6)$$

where $var_1(\tau)$ is the variance of fluorescence lifetime τ when the signal is only one photon. For a mono-exponential decay, the variance can be written in an explicit form as:

$$\begin{aligned} var_N(\tau, T, k) &= \frac{1}{N} \tau^2 \frac{k^2}{r^2} [1 - \exp(-r)] \\ &\times \left(\frac{\exp\left(\frac{r}{k}\right) [1 - \exp(-r)]}{\left[\exp\left(\frac{r}{k}\right) - 1\right]^2} - \frac{k^2}{\exp(r) - 1} \right)^{-1} \\ &= \frac{1}{N} var_1(\tau, T, k) \\ &= \frac{1}{N} \tau^2 var_1(r, k) \end{aligned} \quad (7)$$

where T is the length of the measurement window, k is the number of channels in the window and $r = T/\tau$ is the number of lifetimes in the measurement window.

The standard deviation of lifetime for mono-exponential decays can be written as:

$$\begin{aligned} \Delta\tau &= \sigma_N(\tau, T, k) \\ &= \sqrt{var_N(\tau, T, k)} \\ &= \sqrt{\frac{1}{N} \tau^2 var_1(r, k)} \end{aligned} \quad (8)$$

The fluorescence lifetime normalised error becomes [100]:

$$\frac{\Delta\tau}{\tau} = \frac{1}{\sqrt{N}} \sqrt{var_1(r, k)} \quad (9)$$

The fluorescence lifetime SNR is simply the inverse of equation (9). The Kollner-Wolfrum formulae (equation (6)) demonstrate that lifetime errors of less than 10% can be achieved with hundreds of photons or less and provide a useful model for tuning experimental parameters such as the number and optimal width of time channels, total acquisition time range and fluorophore lifetime contrast.

In both time gating and TCSPC, two channels/bins are often utilised and, as shown in figure 5, the

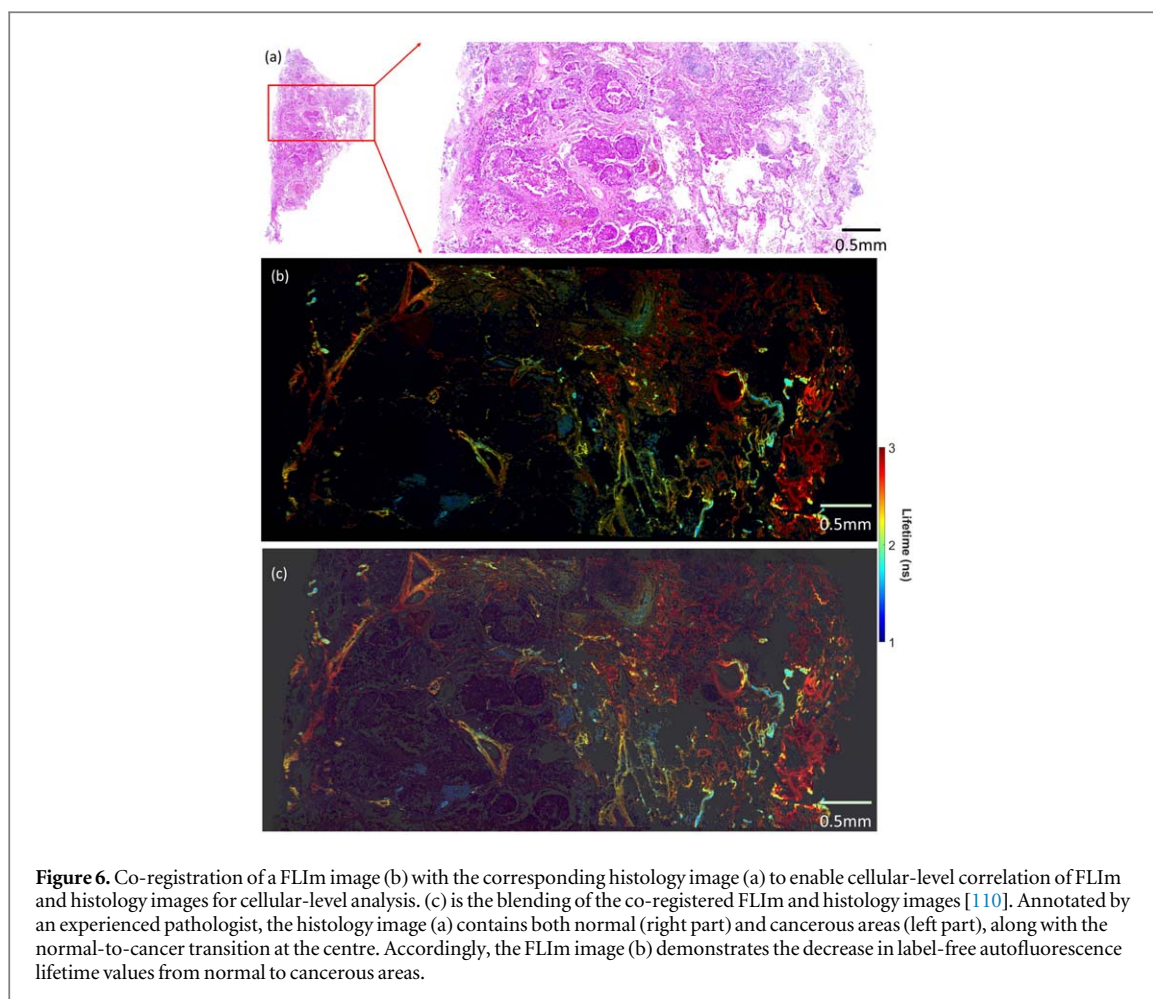
variance for this simplified case is seen to be greater and more sensitive to time-range than in cases where more channels are utilised. A lifetime error analysis for the two-channel case has been carried out by Ballew and Demas [81]. Philip and Carlsson [98] investigated the fluorescence lifetime imaging SNR in the context of frequency domain lifetime measurement methods and carried out comparisons with time domain techniques. On-chip fluorescence lifetime algorithms and SNR were investigated by Li *et al* [101].

From lifetime images and from repeated fluorescence lifetime SNR measurements, lifetime probability distribution functions can be derived. The degree of overlap in these distributions between different fluorophores ultimately determines the efficacy of lifetime as a signature of normal or diseased tissue states. When lifetime is used as a diagnostic test, the complete separation of two fluorescence lifetime distributions implies a perfectly discriminating test, while complete overlap implies no discrimination [102]. From the distributions, we can compute Receiver Operating Characteristic (ROC) which enables medical image data to classify patients as positive or negative with respect to any particular disease [103]. ROC analysis of diagnostic models has been utilised in ML-assisted fluorescence lifetime imaging endomicroscopy ex-vivo lung cancer studies by Wang *et al* [104] and in studies by Chen *et al* [105] aimed at distinguishing basal cell carcinoma from actinic keratosis and Bowen's disease.

3. Fluorescence lifetime image analyses

3.1. Fluorescence lifetime image registration

Image registration is an important task in medical image analysis. It facilitates the matching and alignment of images from different modalities or different



time periods in the same modality, aiming to provide the end-user with richer diagnostic information [106, 107]. Currently, good results can be achieved with the many registration algorithms proposed in the literature. However, very limited studies have looked at the registration problem in microscopic imaging or FLM, while most of the algorithms are considered to be very specific to certain clinical applications or datasets. As described in section 2, lifetime contrast is based on the averaged lifetime derived from statistical methods, for example, histogramming of lifetime images [30, 108]. Unfortunately, the statistical derivation is unable to reveal cellular-level characteristics of fluorescence phenomena under investigation. An insightful understanding of the investigated tissue requires the correlation at the pixel level of the FLM images with a reference image, such as a histology image, which can be achievable through co-registration. As a result, conclusive decisions can be achieved for various purposes, for example, cancer detection [30, 36], tumour margin detection [109], or cellular-level cancer characterisation [110]. Figure 6 shows an example of the co-registration of a FLM image with a Haematoxylin and Eosin (H&E)-stained histology image, where pixel lifetime can be correlated with tumour components in the histology image by detailed annotations. However, the co-registration remains

challenging due to the different nature of FLM and histology images [58].

Since FLM images contain both intensity and the corresponding lifetime images, both modalities could facilitate the co-registration. Intuitively, the intensity seems more suitable than the lifetime for the task as both intensity and histology images present fluorescence concentration distribution via optical scanning. In contrast, lifetime images are usually homogeneous due to the intensity independence of lifetime—raw lifetime images tend to have a flat nature. It is worth mentioning that a common practice to overcome the homogeneous presentation is to saturate a lifetime image with its corresponding intensity image so that both structural and spatial distribution in FLM images can be explicitly illustrated [12, 30, 108]. In addition, spectral FLM images at various emission wavelengths are often visually different since certain structural features may only emit at a particular range, which also deteriorates the effectiveness of the co-registration [12, 111]. Due to the complexity of the staining procedure, artefacts are often introduced to histology images, such as colour variations caused by the differences in staining and the scanners [112], distortion of tissue structure [113], or tissue contamination [114]. All these factors contribute to the challenges of co-registration.

Despite the importance of co-registration, little effort has been dedicated explicitly to it. Generally, co-registering FLIM and histological images can be performed on fixed or unfixed tissue. Since the underlying structure is unchanged, the co-registration of the images on fixed tissue is relatively straightforward. In this case, non-rigid registration could be applied directly. Bird *et al* [115] applied a landmark-based semi-automatic registration to spatially align spectral FLIM and H&E-stained histology images acquired on H&E-stained tissue for lung cancer diagnosis. Wang *et al* [110] used a widely-used regression-based approach to estimating the homography matrix for the co-registration of full-spectral FLIM and H&E-stained histology images to extract lifetime signatures of lung cancer. When it comes to unfixed tissue, the situation becomes more complicated as, in most cases, the spatial structure has changed, and advanced technologies may be required to deal with the distortion. For example, Unger *et al* [116] employed an indirect method to align *ex vivo* tissue with H&E-stained histology images. In their report, the histology images were first correlated with the corresponding white-light images of the tissue using a hybrid method containing landmark-based and deformable registration, and the relevant lifetime images were mapped to the histology images based on the white-light images.

As far as biomedical image registration is concerned, the success of DL technologies has significantly advanced registration performance, where both supervised and unsupervised techniques have achieved remarkable results for rigid and deformable registration [106, 107]. However, their applicability to FLIM image registration may not be straightforward. The reasons are multifold. FLIM and histology images usually contain sub-billion pixels. The direct application of the existing methods requires a significant amount of computational resources, which are not always available. As a result, patch-level registration may be a better solution, such as [110, 113]. Meanwhile, DL-based registration usually requires a significant number of images for training and validation, which is not always available, particularly for whole-slide images. In addition, obtaining ground truth in this domain remains challenging, which presents obstacles to the application of matured DL-based registration approaches to FLIM images. Nevertheless, some proven registration techniques for microscopic images could inspire the development of this area. Jiang *et al* [117] applied a hierarchical strategy to a kernel density estimator so that H&E and immunohistochemical (IHC)-stained WSI images can be robustly correlated. Ge *et al* [118] proposed an unsupervised DL model to align histological images generated using different staining techniques, where multi-scale structural features were retrieved to guide the model to generate an optimal deformation field.

3.2. Segmentation, detection and tracking in fluorescence imaging

Segmentation, detection, and tracking of biological and chemical substances are pivotal for a comprehensive semantic analysis of FLIM data. Such capabilities underpin critical applications like cancer analysis and drug-target engagement [119–121]. Examples include, segmenting, detecting, and tracking microscopic images of viruses (e.g., SARS-CoV and its variants [122]), bacteria like *E. coli* and *Staphylococcus aureus* [123], as well as various types of cells [124]. Essentially, the aim is to identify and quantify objects of interest to provide an enriched image semantic analysis.

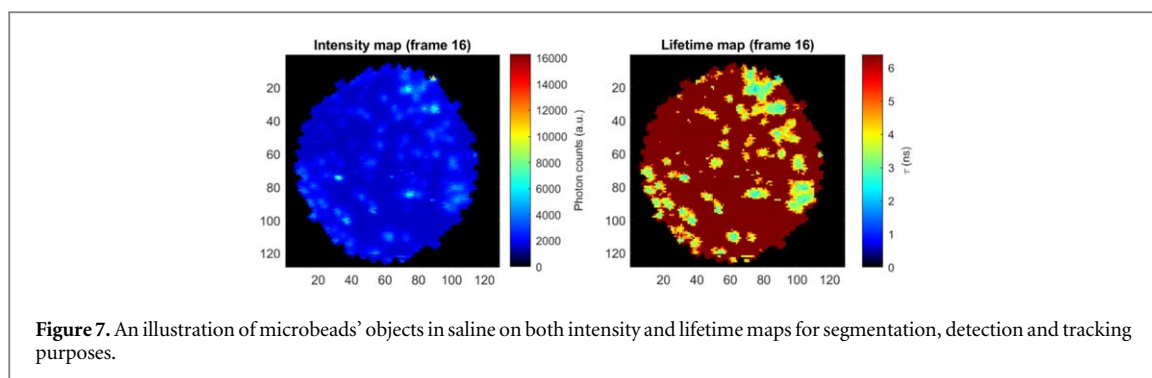
Examples of such methods within FLIM include the automatic segmentation of cellular FLIM images. This technique is inspired by a multiresolution community detection-based network segmentation approach. It focuses on distinguishing segments based on their average lifetimes relative to the background of the FLIM image [125], which underlines the role of precise segmentation in decoding cellular structures and interactions.

The journey from conventional image processing to machine and deep learning in image segmentation is noteworthy. A recent innovation combined machine learning-based interactive image segmentation with supervoxels and clustering. This fusion aids in automatically identifying similarly coloured images across vast datasets [126]. It addresses challenges like colour variability in biological and medical images, paving the way for more efficient interactive segmentation on larger datasets.

In terms of object classification, binary segmentation focuses on differentiating background from foreground at a pixel level, helping to identify the location and concentration of objects, whereas tracking interprets the objects' movement and path [127]. Historically, techniques used for segmentation, detection, and tracking can be categorized into three domains:

- (1) Traditional digital processing methods using Gaussian Mixture Model (GMM), such as Otsu's thresholding [128], and other methods such as mean shift [129], and the Kalman filter [130].
- (2) Feature engineering-based training approaches as seen in the works of Viola and Jones [131] and Kalal *et al.* [132].
- (3) Techniques centered around DL [133, 134].

In clinical microscopy imaging, precise object recognition can be intricate due to problem-specific factors. For instance, detecting specific bacteria [123] is often subject to expert opinion, making it more challenging than problems with clear visual cues. In the field of fluorescence imaging, many researchers have utilized computer vision techniques specifically for steady-state systems, as indicated by [135].



However, there is a growing emphasis on transitioning to the time-resolved capabilities provided by FLIm. Such a transition offers valuable insights into microorganisms, ranging from their fundamental structures to their ecological interactions. The essence of this idea is evident in the object mapping within FLIm systems:

$$f: (o_1, \dots, o_m) \mapsto (o'_1, \dots, o'_n),$$

where the object o_i , $\forall i = 1, \dots, m$ in an intensity map corresponds to the object o'_j , $\forall j = 1, \dots, n$ in its lifetime match by the mapper f . As a tangible illustration of the concept of the mapping, figure 7 shows, at a frame sequence, how polymer microbeads (as objects of interest) in saline appear on both intensity and lifetime maps. As a result, the segmentation, detection and tracking tasks can be tested on the complementary lifetime map [135].

Moreover, the manual analysis of imaging data is labor-intensive, subjective, and resource-heavy, making automation crucial. Koujan *et al* [154] offered an automated method to segment pulmonary Optical Endo-microscopy (OEM) imaging data based on their diagnostic importance, leveraging local binary patterns, the Support Vector Machine (SVM) classifier, and a one-versus-all strategy for improved accuracy. In bacterial detection, both supervised and unsupervised methods find applications. Research like [155] and [156] used supervised models for bacteria detection in microscopy images. In contrast, unsupervised methods, like the Bayesian approach for detecting microorganisms in pulmonary OEM imaging, treat bacterial objects as outliers added to actual pixel values disturbed by Gaussian noise [157]. Another approach divided images into overlapping patches, transforming bacteria detection into a minimization problem [158].

For other microorganisms' shapes can often be represented as 2D and 3D mathematical models. For instance, Zamani *et al* [124] introduced a mechanism for detecting human cells like HT-29 in the blood. They viewed cell detection as ellipse recovery, using a 2D CNN to estimate an elliptical shape's parameters.

Drug discovery is another area where the detection of biological phenomena using FLIm is deemed useful. One of the key elements across the drug discovery pipeline is the accurate quantification of protein-

protein interactions, also referred to in drug development as drug-target engagement [159]. There are different methods to measure these interactions, such as surface plasmon resonance, nuclear magnetic resonance spectroscopy, and many more [160]. However, these methods are limited and often require protein purification, which may alter the true functionality of proteins and sample preparation [160]. In recent years, the detection of FRET using FLIm technology has provided a robust imaging modality for drug discovery, enabling a reliable readout of sub-cellular information while significantly reducing photo-toxicity levels [159]. For example, Ochoa *et al* [161] reported the development of an Macroscopic Fluorescence Lifetime Imaging (MFLI) optical imaging system for the quantitative measurement of *in vivo* drug-target interaction. The system reported in the study acted as a direct lifetime sensor of FRET in live animals. Additionally, a recent study by Yang *et al* [162] highlighted the potential of live-cell image-based ML strategies in reducing variability in Pluripotent Stem Cell (PSC) differentiation systems, which can be instrumental in supporting recent FLIm and ML advancement in drug discovery, disease modeling, and regenerative medicine. The study emphasized the role of artificial intelligence in guiding and optimizing PSC differentiation, offering insights into the differentiation process for functional cell manufacturing in biomedical applications.

3.3. Fluorescence lifetime imaging knowledge extraction and categorisation

3.3.1. Machine learning uses cases in FLIm

Fluorescence-based medical imaging has proved to be effective for cancer discrimination for decades [36, 43, 163, 164]. However, Marcu [36] found in 2012 that the majority of those studies had been made on Fluorescence spectroscopy [165–168]. Meanwhile, research on FLIm was still relatively new at that time. However, ten years later, new studies have been published about FLIm and cancer [169–171], with the majority coming from a small number of research groups, thus limited in diversity.

ML has not been extensively used for images features extraction on FLIm data, such as disease detection or tissue segmentation. However, the number of

publications has been steadily increasing since 2018, as shown in table 2. From the literature, we can differentiate the following uses-cases: image-level classification, which assigns one label to an entire sample or region, and pixel-level classification, which gives a label to every pixel. The latter can then be used either for semantic segmentation, as each pixel is labeled, or for image-level classification by averaging the labels across all pixels. Both uses-cases can be applied to discriminate any number of classes such as cancer, pre-cancerous, adipose tissue, etc. The most common type of classification, binary classification, can be used, for example, to differentiate cancer (positive) from everything else (negative) without paying attention to specific details of the tissue. More recent studies have investigated ML cells classification, some of them being listed in table 2 [148–153]. They classify the type of cells or their states, such as activated and quiescent for T-cells.

In their review, [24] reported the absence of DL-based classification research on FLIm, but concluded that it would play a critical role in the future. This observation is reinforced when considering the growth of ML and DL techniques in the last decades, the accessibility of programming languages like Matlab [172] and Python [173], and frameworks such as Scikit-Learn [174].

However, as the learning curve flattens, research may be carried out without the proper methodology, and confusion can also be found in published results. Moreover, ML is also used as an analysis tool, to prove the effectiveness of FLIm data for a particular task, but fewer efforts are made after this to actually develop custom ML, state-of-the-art models.

3.3.2. Performance evaluation in ML tasks applied to FLIm

ML model evaluation is performed by comparing metrics generated from their outputs. Those metrics are chosen depending on the type of model and the specific of the problem being solved. Regression metrics have been discussed in section 2.6. Classification and segmentation use a subset of overlapping metrics, commonly used in a wide variety of domains such as information retrieval, statistics or detection theory. Those metrics are based on the four possible outcomes of a binary prediction, which features two labels, positive (P) and negative (N), as well as two possible predictions, positive (PP) and negative (PN): True Positive (TP) (P and PP), True Negative (TN) (N and PN), False Positive (FP) (N and PP), False Negative (FN) (P and PN), which can be extended to multi-class classification by choosing one class as positive and grouping all the others as negative. For example, consider the three possible results of a test: healthy, pre-cancerous, cancerous. By choosing cancerous as ‘positive’ and the rest as negative, the number of TP or TN can be deduced. Based on those outcomes, several

metrics can be derived, such as sensitivity ($\frac{TP}{P}$), specificity ($\frac{TN}{N}$), precision ($\frac{TP}{PP}$), accuracy ($\frac{TP+TN}{P+N}$), F1-score ($\frac{2TP}{2TP+FP+FN}$) and other [175].

In general, accuracy is considered a key metric for classification, when the different classes are balanced [176]. Otherwise, if one class is under-represented, the accuracy will be biased towards one or multiple classes. However, in some situations, such as medical diagnosis, the sensitivity and/or specificity may be more important because the detection probability of the model is evaluated [175]. The dice score, or F1-score, is an important metric for semantic segmentation [177], but specific metrics derived from the segmented object may be used. For example, if the objects can be counted or if they can be connected, such information can be transformed into metrics. Pixel-by-pixel semantic segmentation classifies one pixel at a time, without knowledge of the other pixels. This type of segmentation is close to classification, and thus similar metrics are often used. On the other hand, region-level segmentation, which takes an image as input and outputs an image, often uses the Jaccard index in addition to the dice score.

3.3.3. Machine learning modalities: model

Random Forest (RF) and SVM are the two most used models in the studies listed in 2. Extreme Learning Machine (ELM) has been used twice [144, 179]. DL implementations in FLIm range from MLP with dense layers, used alongside RF and SVM without distinction, to custom implementations of pre-trained CNN. However, DL has been scarcely used.

More complex ML architecture started to be created for FLIm. An autoencoder joint with a dense neural network has been developed by [137]. It works directly on the decay curve and does not use hand-crafted features (such as lifetime or phasor plot value). Wang *et al* [104] developed a custom CNN architecture by replacing some layers and building it on top of the classical ResNet [180] architecture.

CNNs are now widely used, and pre-trained networks are readily available online. Architectures such as ResNet, Inception [181] or EfficientNet [182, 183] can be found in many DL libraries and frameworks, and will provide researcher classification capacity far superior to simpler machine learning models. Through fine-tuning [184], those models can be easily trained without needing as many resources as being trained from scratch, and could provide better results than simpler ML models. Fully-convolutional NN such as UNet [25], or its state-of-the-art variants like UNet++ [185] and UNet3+ [186], should be preferred for segmentation tasks to pixel by pixel classification models. Those CNNs can take advantage of spatial feature information and are specifically designed for segmentation task, with a focus in medical imaging.

Table 2. Summary table of all ML paper reviewed in this section. The region used corresponds to the input fed to the model, either one pixel or a region (i.e. a group of pixels, like an image). For the pixel level, the classification result can be averaged across all the pixels of one sample. The majority of cell research uses the lifetime captured on one cell.

Author	Year	Medical effort	Detail	Region used	ML type	Detail	<i>in vivo</i>
Bianchetti <i>et al</i> [41]	2021	cancer	breast	pixel	clustering	Binary (1 versus N)	
Phipps <i>et al</i> [4]	2017				classification		
Unger <i>et al</i> [5]	2020				detection		
Jo <i>et al</i> [136]	2018	oral		pixel (average)	classification	Binary (1 versus N)	<i>in vivo</i>
Marsden <i>et al</i> [109]	2020			point	detection	Binary	<i>in vivo</i>
Caughlin <i>et al</i> [137]	2021			point (average)	classification	Binary (1 versus N)	<i>in vivo</i>
Duran <i>et al</i> [138]	2021			pixel (average)	detection	Binary (1 versus N)	<i>in vivo</i>
Cosci <i>et al</i> [139]	2016			(5) pixel (average)	detection	Binary	<i>in vivo</i>
Chen <i>et al</i> [105]	2019			subregion	classification	3 classes	
Yang <i>et al</i> [140]	2020	region		region	classification	Binary (1 versus N)	
Romano <i>et al</i> [141]	2020			region	detection	Binary (1 versus N)	<i>in vivo</i>
Vasanthakumari <i>et al</i> [142]	2022			region as feature	detection	Binary (N versus N)	<i>in vivo</i>
Wang <i>et al</i> [119]	2020	lung		region	detection	Binary	
Wang <i>et al</i> [104]	2021			region	detection	Binary	
Gu <i>et al</i> [143]	2014			region	detection	Binary (1 versus N)	
				(mean/std)	classification	Binary (N versus N)	
				detection	Binary (1 versus N)		
Gu <i>et al</i> [144]	2015			region	detection	Binary (1 versus N)	
Sahoo <i>et al</i> [145]	2018			region	classification	Binary	
Ji <i>et al</i> [146]	2022			region	classification	Binary (1 versus N)	
Weyers <i>et al</i> [3]	2022		oropharyngeal	point	detection	Binary	<i>in vivo</i>
Butte <i>et al</i> [147]	2011	cell	brain	single point	classification	4 classes	<i>in vivo</i>
Walsh <i>et al</i> [148]	2020		T-cell	subregion (cell)	classification	Binary	

Table 2. (Continued.)

Author	Year	Medical effort	Detail	Region used	ML type	Detail	<i>in vivo</i>
						4 classes	
Dunkers <i>et al</i> [149]	2021		S. mutans	pixel (average)	classification	3 classes	
Qian <i>et al</i> [150]	2021		Stem cell	subregion (cell)	classification	Binary	
Cardonna <i>et al</i> [151]	2022		T-cell & Cancer cell	subregion (cell)	classification	4 classes	
Neto <i>et al</i> [152]	2022		macrophage	subregion (cell)	classification	Binary	
Kröger <i>et al</i> [153]	2022		macrophage	subregion (cell)	classification	Binary (N versus N) Binary (1 versus N) Binary (1 versus N) Binary	<i>in vivo</i>
Marsden <i>et al</i> [10]	2021	parathyroid/thyroid		point point (average)	detection detection	Binary (1 versus N)	<i>in vivo</i>

For binary classification, N refer to multiple class grouped into one, while 1 refer to only one label. E.g: Phipps *et al* [4] do Binary (1 versus N) as they classify cancer tissue (1 label) vs other tissue (fibrous and adipose grouped together, so 2 label grouped as 1).

Classification with pixel as region used refer to pixel-by-pixel classification, used for segmentation. It is not listed as semantic segmentation, since only one pixel is used each time as input.

Generally speaking, DL models will take as input an entire multi-spectral FLIm image. In the case of intensity and lifetime, images can be either concatenated to be fed as one in a NN, or two networks can be used for feature extraction, before being concatenated and then fed into the classifier part. Alternatively, intensity and lifetime can be mixed, similarly to what is shown in 2.

3.3.4. Machine learning modalities: features

The FLIm features that were captured in each study are listed in table 4. They were either directly used as input or processed and/or went through feature selection. Principle component analysis (PCA) has sometimes been used to reduce the dimensionality of the images, projecting the data into a new space but losing the spatial information [119, 145]. Duran-Sierra *et al* [138] used sequential forward selection to find out which features would be the best for classification. Other methods have been used to evaluate the ‘weight’ or the impact of each feature, e.g removing each feature from the model’s input and observing the classification results [150]. However, the final model they presented still uses all collected features without excluding any. They were not the only ones to evaluate subsets of feature, usually handmade subsets such as ‘intensity only’ or ‘lifetime only’ [136, 138, 148, 152].

Ex vivo experimentation is more common because it is easier, especially for cell studies which can be cultured. For tissue, imaging it *in vivo* means bringing the hardware into a medical setting which generally implies significant restrictions that are harder to meet during an experimental study. However, a few groups have been working on *in vivo* applications (Jo [136, 137, 141, 142]), (Marcu [3, 10, 109]), Cosci *et al* [139] and Butte *et al* [147]. Except for Butte *et al* who worked on brain cancer, all the other diseases investigated targeted the skin or the head and neck region, such as oral cancer. The data gathering methods are split between non-invasive analysis of the tissue, e.g directly on the skin for skin cancer, and invasive method, with the FLIm data being captured during a biopsy before the tissue is removed. Only [109] used both *ex vivo* and *in vivo* samples and obtained better results *in vivo*, with 86% and 72% accuracy for *in vivo* region level and point level classification, against 78% and 67% for *ex vivo*. With only one study for comparison, it is impossible to draw any conclusion on the usability of ML for *in vivo* live recognition of tissue.

3.3.5. Challenges and issues in machine learning

As described earlier, a model should be trained and then tested on different data. Alternatively, if the number of samples is too low, Cross-Validation (CV) can be used, by partitioning one dataset in N parts, training a different model on $N - 1$ partition and testing it on the remaining one, multiple times [175]. CV is normally used for optimising the hyperparameters of a model but can be used for testing as we

described as well. However, testing methods are not meant to be compared and should not be considered hyperparameters. Yang *et al* [140] used bootstrapping, hold-out and K-fold CV (K-CV) and compared the accuracy obtained. Besides wrongfully comparing the methods, the difference in the value of their metrics is very wide, hinting at some issue with over-fitting or in the distribution of their test set.

In the same manner, the training and testing sizes are not meant to be hyper-parameters, as the model would be optimised to improve the testing set results, which is equivalent to over-fitting to it rather than simply use it for evaluation. For example, [148] used different threshold split and handpicked the best result, which can lead to confusion on how well their model is performing. Similarly, it is not clear which validation method was used by Sahoo *et al* [145] who seems to mix CV and hold-out.

Incorrect uses of vocabulary are generally harmless but, in some cases, can lead to some confusion. For example, [148] regularly exchanges ‘accuracy’ and ‘Area Under the Curve (AUC)’ when referring to the ROC curve. This error can also be found in other papers such as [5, 152] or [144]. While they are usually understandable with the context, they can create misunderstanding, especially if both metrics are used and their results differ.

In their paper, Sahoo *et al* [145] defined a custom formula ‘combining’ sensitivity and specificity, which led to results of 100% for each value, but an accuracy of only 84% which is not possible. Metrics such as those defined earlier are standards that are used by everyone and are not meant to be modified. to allow for a fair comparison.

3.3.6. Challenges and issues in FLIm

The multiple exterior factors influencing the intensity and lifetime value, such as biological difference (pH, oxygen level, type of tissue imaged, ...) or spectroscopy parameters (absorption and emission wavelength, numbers of photons, ...) make it hard to compare different FLIm studies. As cited in Subsection 2.4, different systems exist: time and frequency domain. Additionally, custom homemade FLIm devices are used by a large number of groups, such as [105, 109, 136] to only cite a few, increasing the disparity in the data collected.

When exciting tissues, researchers are looking to target particular enzymes or probes, such as FAD or NADH. The choice of the target may be impacted by the type of tissue or cells. For example, Walsh *et al* [148] decided to use NAD(P)H after evaluating a sample of T-cell. Those different enzymes do not have the same properties and may react differently depending on the environment but also the excitation wavelength, impacting their emission wavelength. Figure 8 shows a visual representation of the emission wavelength captured, associated with the fluorophores (when listed). Different sources are cited when it

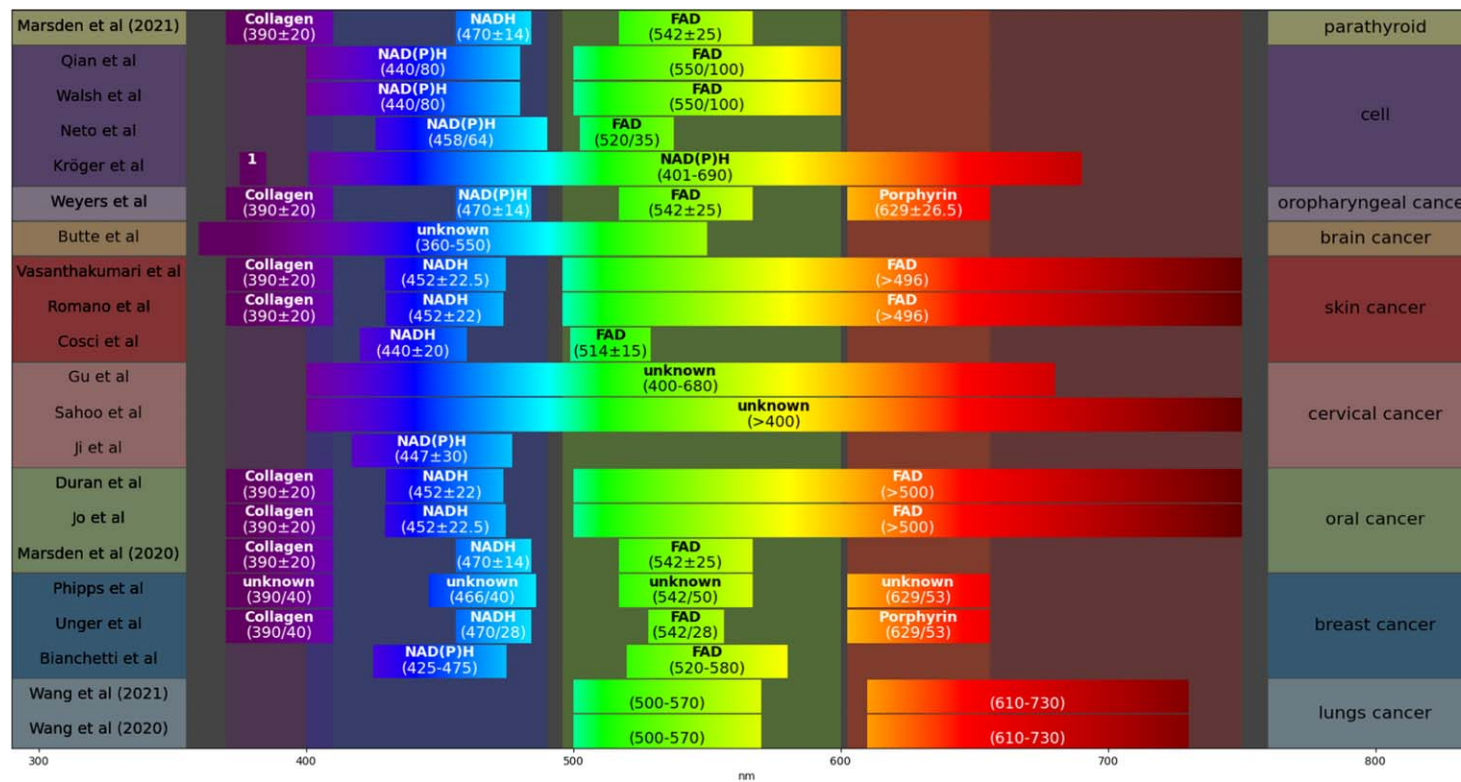


Figure 8. Visualisation of the channel bandpass/filter used on the emission wavelength. The target is listed when it has been mentioned. 1: The first filter from 375 nm to 385 nm was used to detect the second harmonic generation signal.

Table 3. Performance summary of the studies cited in figure 8, sorted chronologically. All models and validation methods have been listed, and the most important experiment’s results has been singled out for each study, with its specification in the last column, and the model used emboldened.

Author	Year	Validation	Model	ACC	SE	SP	AUC	Specification
Butte <i>et al</i> [147]	2011	LOOCV	LDA	71.43%	47.06%	94.64%		Temporal + Spectral SE/SP: HGG
Cosci <i>et al</i> [139]	2016	10-fold CV	KNN (K=5)	90%	92.8%	88.8%		SE/SP: healthy vs other
Phipps <i>et al</i> [4]	2017	LOPOCV	SVM (rbf)	97.8%	100%	96.90%		balanced cancer vs other
Jo <i>et al</i> [136]	2018	LOPOCV	QDA	89%	95%	86.79%	0.91	FLIm features
Unger <i>et al</i> [5]	2020	LOPOCV	RF		88.78%	93.14%	0.96	tumour vs no tumour
Marsden <i>et al</i> [109]	2020	LOOCV (tongue/tonsil only). Test on <i>other</i> with RF	RF , SVM (rbf), 1D-CNN		86%	87%	0.88	<i>in vivo</i> , region
Walsh <i>et al</i> [148]	2020	Split train/test	LogReg , RF, SVM				0.95	type and activation
Wang <i>et al</i> [119]	2020	LOPOCV (3 times) train=(90% / 10%) train/validation	2D-CNN 3D-CNN	86.5%	89.5%		0.858	3-channel DenseNet121
Romano <i>et al</i> [141]	2020	Split train/test 75% / 25%	LDA	73%	88%	67%	0.79	intensity and lifetime
Dunkers <i>et al</i> [149]	2021	Out of bag no validation	RF	95.91%				PBS buffer, lifetime and phasor variable
Wang <i>et al</i> [104]	2021	One patient out train=(90%10%) train/validation	Custom 3D-CNN	84.9%	80.95%		0.882	MSCD-ResNet50 preserved complexity
Qian <i>et al</i> [150]	2021	train/test set train=(80% / 20%)train/validation	LogReg , RF, SVM	>85%			0.9085	all variable logistic regression
Marsden <i>et al</i> [10]	2021	LOPOCV	NN, RF , SVM		100%	93%		region level
Duran <i>et al</i> [138]	2021	7-fold CV (train) Best model on test	NN, SVM, RF		78%	61%	0.81	ensemble
Weyers <i>et al</i> [3]	2022	Pre-trained then tested	RF		96%	89%	0.9	mean over patients
Neto <i>et al</i> [152]	2022	10-fold CV	RF				0.944	all 2p FLIm variables
Vasanthakumari <i>et al</i> [178]	2022	LOPOCV	QDA	88.33%	84.21%	90.24%		Phasor + intensity + lifetime variables
Ji <i>et al</i> [146]	2022	Train: 151 Cancer/CNI +217 normal Test: Images from 48 patients	K-means as classifier		90.90%	100%	0.95	$\tau_{avg} + \alpha_2$
Kröger <i>et al</i> [153]	2022	Split train/test 50% / 50% Repeat 10 000 times	Decision Tree		88% 82%	89% 90%		M1 MΦ M2 MΦ

HGG: High Grade Glioma; LDA=Linear discriminant analysis; RF=Random Forest; KNN=K-Nearest Neighbors; QDA=Quadratic Discriminant Analysis; SVM=Support Vector Machine; LogReg=Logistic Regression; CNN/NN=(Convolutional) Neural-Network; LOOCV/LOPOCV/CV=(Leave-One-(Patient)-Out) Cross-Validation; ACC=Accuracy; SE=Sensitivity; SP=Specificity; AUC=(Receiver operating characteristic) Area Under the Curve.

Table 4. Summary of the FLIm features and the data format used by the study examined in 8. For more information about the data, such as medical details, we refer the reader to the paper itself. The filter reference column describes the study cited to justify the channel filter/bandpass used, which can be visualised in figure 8.

Author	Filter reference	Lifetime	Features	data	Sample
Bianchetti <i>et al</i> [41]		Bi-Exponential	$I_{\text{NAD(P)H}}, \tau_{\text{NAD(P)H}}, I_{\text{FAD}}, \tau_{\text{FAD}}$	20 images 18-22 cells/images	512px × 512px
Phipps <i>et al</i> [4]	Yankelevich <i>et al</i> [187]	Laguerre	$\tau_{390}, \tau_{466}, \tau_{542}, \tau_{629}$	20 tissue 14 patients	112,468px
Unger <i>et al</i> [5]		Laguerre	Per channel: $\tau_{\text{avg}}, I_{\text{avg}}$, 12 Laguerre coefficients	18 patients & samples	323,026px after erosion
Marsden <i>et al</i> [109]		Laguerre	Decay for 1D-CNN. Per channel: $\tau_{\text{avg}}, I_{\text{avg}}$, 12 Laguerre coefficients	50 <i>in vivo</i> , 53 <i>ex vivo</i> (20 tonsil, 23 tongue, 10 <i>other</i>)	76,695px <i>in vivo</i> 67,893px <i>ex vivo</i>
Marsden <i>et al</i> [10]		Laguerre	$\tau_{390}, \tau_{470}, \tau_{542}$	15 patients (9 lymph, 15 adipose, 15 thyroid, 15 parathyroid)	29,393 pixels 41 runs
Weyers <i>et al</i> [3]		Laguerre	Per channel: $\tau_{\text{avg}}, I_{\text{ratios}}$, 12 Laguerre coefficients	55 patients train (38 tonsil, 17 tongue) 6 test (HNSCCUP)	2355 points sample (average)
Jo <i>et al</i> [136]	Cheng <i>et al</i> [188]	Bi-exponential	I_{452}/I_{500} ratio. Per channel: $\tau_{\text{avg}}, \tau_1, \tau_2, I, I_r$. For each feature: difference with the pixel value and image average. 6 features used only	73 patients (53 benign, 6 mDYS, 14 SCC)	
Duran <i>et al</i> [138]		Bi-exponential	Per channel: $I, \tau_{\text{avg}}, (\tau, \alpha)_1, \tau_2$ 3 intensity ratios (I_a/I_b) 3 sum intensity ratios ($(I_a+I_b)/I_c$)	68 train (34 healthy/cancer), 23 test	
Walsh <i>et al</i> [148]	Skala <i>et al</i> [43]	Bi-exponential	Cell size, Optical Redox Ratio, NAD(P)H (τ, α) ₁ , $\tau_2, \tau_{\text{avg}}$ FAD (τ, α) ₁ , $\tau_1, \tau_2, \tau_{\text{avg}}$	6 donors	8,355 cells
Qian <i>et al</i> [150]		Bi-exponential	Optical Redox Ratio, NAD(P)H (τ, α) ₁ , (τ, α) ₂ , τ_{avg} , I FAD (τ, α) ₁ , (τ, α) ₂ , τ_{avg} , I		11,218 cells (train) 19,245 cells (test)
Wang <i>et al</i> [119]		RLD	τ, I	≈70,000 images	128px × 128px
Wang <i>et al</i> [104]		RLD	τ, I	24,554 images	128px × 128px
Neto <i>et al</i> [152]	huang <i>et al</i> [189]	Bi-exponential	Optical redox Ratio, $\tau_1, (\tau, \alpha)_2, \tau_{\text{avg}}$	6 donors	1,153 cells
Dunkers <i>et al</i> [149]	no filter	Bi-exponential	$I, \tau_1, \tau_2, p_1, (G, S)$ (Phasor)	19,182px/399 obj SP, 21,381px/205 obj HK, 3,324px/33 FP	43,887px 631 objects 3,324px/class when balanced
Cosci <i>et al</i> [139]		Bi-exponential	NAD(P)H (τ, α) ₁ , (τ, α) ₂ FAD τ, α) ₁ , (τ, α) ₂	101 measurement (55 healthy, 30 mDYS, 15 mrDYS)	5px (measured point)
Butte <i>et al</i> [147]		Laguerre	$\tau_{380, 440, 440, 550}, \text{AVG}(\tau_{440+450+460})/\text{AVG}(\tau_{540+550}), \text{AVG}(\tau_{380+390})/\text{AVG}(\tau_{440+450+460}), I_{370}/I_{380}, I_{370}/I_{390}, I_{390}/I_{440}, I_{440}/I_{370}, I_{440}/I_{380}, I_{440}/I_{450}, I_{450}/I_{370}, I_{450}/I_{550}, I_{460}/I_{370}, I_{460}/I_{390}, I_{460}/I_{450}, I_{460}/I_{540}, I_{460}/I_{550}, I_{540}/I_{440}, I_{550}/I_{380}, \text{LEC-0}_{60}, \text{LEC-0}_{450}, \text{LEC-0}_{460}, \text{LEC-0}_{550}, \text{LEC-1}_{370}, \text{LEC-1}_{390}, \text{LEC-1}_{440}, \text{LEC-1}_{450}, \text{LEC-1}_{540}$	71 regions 35 normal cortex, 12 normal white matter, 7 low grade glioma, 5 anaplastic oligodendroglioma, 1 glioma with radiation necrosis	Single point

Table 4. (Continued.)

Author	Filter reference	Lifetime	Features	data	Sample
Romano <i>et al</i> [141]		Bi-exponential	Per channel: $I, I_{\text{norm}}, (\tau, \alpha)_1, \tau_2, \tau_{\text{avg}}$	76 samples 38 each class	140px × 140px
Vasanthakumari <i>et al</i> [142]		Bi-exponential	Per channel: $(\tau, \alpha)_1, \tau_2, \tau_{\text{avg}}, I, I_{390/452}, I_{390/500}, I_{452/500}$. Per frequency: distance, spread, angle, symmetry	30 patients, 60 samples 41 benign, 19 malignant	140px × 140px
Kröger <i>et al</i> [153]		Bi-exponential	$(\tau, \alpha)_1, (\tau, \alpha)_2, \tau_{\text{avg}}, \tau_2/\tau_1, \alpha_1/\alpha_2, (\alpha_1 - \alpha_2)/(\alpha_1 + \alpha_2)$	25 patients	399 cells
Ji <i>et al</i> [146]		Bi-exponential	$\alpha_2, \tau_{\text{avg}}$	71 patients 11 cancers, 7 CNI 18 Benign, 23 Normal 12 Follow-up	256 px × 256 px

comes to the filter used, all listed for each group in table 4. In the majority, groups tend to cite papers they have written previously, such as Marcu group [4, 5, 109] citing Yankelevich *et al* [187], Skala group [148, 150] citing Skala *et al* [43] or Jo *et al* [136] citing Cheng *et al* [188] to justify the channel used.

Moreover, even if they are looking at the same FLIm information, researchers may not use the same features as input for their models. They may use the intensity, the lifetime, the average lifetime, or other specific variables, such as the Laguerre variable, or the Shannon entropy, as seen in table 4. Thus it makes it harder to compare the results of two different studies even when they are from the same group.

3.3.7. Challenges and issues with metrics

Metrics are used to evaluate a model, to highlight its performance but also its flaws. Their choice, as described earlier in section 3.3.2, should be tailored to the problem tackled. For the common case of cancer detection seen in table 2, with cancer considered ‘positive’, a false negative have a greater impact than a false positive, as missing a cancer case during diagnosis carries more risks than diagnosing it to an healthy person. Therefore, a metric such as sensitivity needs to be used. However, it can not be used by itself, as classifying every sample as positive would lead to a perfect sensitivity, but a very inaccurate model. It should be paired with another metric such as specificity, to evaluate how well the model can recognise the disease. The distribution of the class should also be considered to not bias the metrics [175]. For this example, Unger *et al* [5] used the Matthew Correlation Coefficient (MCC). It has previously been described as a metric which properly takes care of unbalanced dataset [190], as the one Unger *et al* has, but this has been challenged recently [191]. Since Unger *et al* did not provide other metrics than MCC and ROC AUC, and no other studies in table 2, it is impossible to compare their results. This would be fine if they are the only one featuring an imbalanced dataset, but if other studies choose metrics that fit only with balanced data, then two issues are created: improper choices of metrics not reflecting the real performances of the model, and the impossibility of comparing the results with studies using proper metrics. The ROC AUC is also often used, when applicable, but, as with all metrics, it should not be used alone as it measures the model’s capacity to rank the samples rather than its capacity to classify them. It can be used to evaluate the trade-off between specificity and sensitivity. While imperfect, accuracy is still a pretty commonly used metric, especially in research to improve ML models, where many datasets are balanced so that the metric is still relevant for problem evaluation, Therefore, it should also be included when possible. Regarding segmentation, metrics commonly used, such as the dice score and Jaccard index, should be preferred to accuracy, which can be easily biased by an imbalance

in the labels. However, those metrics are dependent on the choice of the ‘positive’ class and thus need to be properly set to target cancer or other diseases when using binary classification.

3.3.8. Challenges and issues with information sharing and reproducibility

Apart from the lack of standardization and common elements between different research ML and FLIm research, another issue is the limited information about the ML methodology which hinder the reproducibility of the work. It is sometimes hard to understand which features were used or how they were obtained. Information about the data format is not always explicitly provided, which also applies to information about the dataset, as seen in tables 3 and table 4. Even when included, the information will often be hard to understand for someone not knowledgeable about FLIm, and inter-disciplinary researcher may struggle to properly extract information from those studies. Information such as the number of samples, the studied area or the data format, and more general details provided in tables 2–4, should be clearly stated and understandable for everyone.

4. Conclusion, challenges and opportunities

In the past few years, the use of FLIm applications in ML has grown exponentially. It appears that DL is on track to becoming a crucial component of lifetime estimation pipelines, thanks to the significant improvements it offers in both time savings and performance. Through successful proof of concept experiments, FLIm image classification and segmentation methods have demonstrated their effectiveness. As a result, we are now witnessing the publication of the first robust models, which further validate the potential of these methods [104, 137]. Many advancements which could benefit FLIm research still require attention. The main limiting factor is public data availability, which restricts both current FLIm research and the arrival of new ML researchers in the FLIm community. Datta *et al* [35] highlighted in their review the importance of dissemination to the community and standardisation of FLIm tools and methodology. We suggest two approaches for the current FLIm researchers to improve the quality of the ML research using FLIm data.

- Improvement and clarification of the methodology used, listing all parameters of interest in a clear and easy-to-read manner (tables 2–4 and figure 8 can be used as inspiration). Additionally, studies should include standardised metrics adapted to the problem as well as those included in similar papers, for comparison purposes.

- The creation of open-source datasets, extensively described and easily usable. A step further would be the creation of an ML contest featuring such a dataset to promote the field.

The benefits of the first step would be to aim for improved readability of publication, easier reproducibility, and the facilitation of easier comparison between studies. Whereas the creation of open source datasets could improve lifetime estimation, FLIm analysis as well as be used on both tasks at the same time, in a similar fashion as what Caughlin *et al* [137] did by training a joint autoencoder on the raw decay-curve and a classifier on the output of the encoder to classify the images created. The benefit of open-source datasets can be observed in other medical modalities. The creation of the ROSE dataset [192] and the OCTA-500 [193, 194] have pushed higher the number of DL-based OCTA vessel segmentation papers, with many studies published using only open source datasets [195–197]. Additionally, as mentioned, an ML contest could be created featuring such datasets. The Camelyon dataset (a breast cancer histopathology whole-slide images dataset) [198] challenge attracted a wide number of ML researchers, generating an important research output and providing a dataset for pre-training medical models. Kaggle⁵ is a popular website for hosting datasets and competitions, which could also be used for hosting such a contest.

However, open-source datasets will not solve FLIm disparity between setups and experiments. A universal FLIm classifier does not appear to be possible, because of the difference between tissue from different organs, biological environments, fluorophores, exogenous and endogenous fluorescence and other factors leading to disparities in FLIm data. Applications so far have focused on singular tasks such as discriminating between two types of skin cancer. Standardising to some degree FLIm methodologies (on the spectral channel used or the fluorophores targeted, for example) would bring the distribution of the generated images closer, to which DL is sensitive. As highlighted by [58], lifetime can be affected by the experimental setup and other biological factors, including inter-patient variability. This creates challenges but also opportunities for the implementation of transfer learning, which has been proven efficient for cancer in other modalities [199–201] by pre-training DL models on similar medical data from different modalities or pathologies. As FLIm data are limited and not easy to obtain, being able to use lung FLIm images to enhance cervical cancer classification in FLIm images, for example, would be very helpful but may prove to be highly difficult, if possible at all, because of the different obstacles we cited. If so, the dissemination of publicly available datasets would be even more important to allow the use of DL. Ideally,

public datasets should share similar configurations and parameters and be properly annotated, explained and detailed.

Nonetheless, to close the bridge in between the state-of-the-art in FLIm research and DL, we discuss here general improvements that can be applied to a majority of the studies discussed in this review. DL models are now readily available online in public repositories, or in multiple DL libraries such as PyTorch [202] or Tensorflow [203] and should be used in research. They can be trained from scratch, or used with pre-trained weights, usually obtained from Imagenet [204], and then fine-tuned. Such architectures or frameworks include ResNet [180] and its variants [205] for classification or regression, UNet [25] and other fully-convolutional NN for segmentation, and similarly UNet, or UNet-based frameworks such as VoxelMorph, as well as SuperPoint [206] or SuperGlue [207] for registration. Classifications models can be directly applied to replace ML models such as RF or SVM, while segmentation models such as UNet are better adapted for segmentation than pixel-by-pixel classification. The majority of models are trained on RGB images, and may need their input layer to be changed for FLIm images, which possesses different colour channels.

In more depth, multiple techniques can be implemented to make DL models better. Residual connections, introduced by ResNet, are now commonly used in the majority of DL models regardless of their applications. Classical convolutional block, along with convolutional layer, will be made with batch normalisation [208] or layer normalisation [209] and activation function such as Relu [210] or Gelu [211]. However, diverse improvements can be made to those blocks, such as squeeze-and-excitation block [212] or split-attention [213], both for channel attention to recalibrate features and show channel features relationships. Other channels and spatial attention methods can be used, including convolutional block attention [214] or dual attention block [215]. As it can be observed, attention is greatly used to improve DL models, especially since the introduction of transformers [17], which led to the development of visual transformers [216], notably the Swin transformers [217]. Those mechanisms of self-attention should be considered when implementing DL models for FLIm, especially channel-wise attention, since FLIm possesses complicated inter-channel relationships. Attention mechanism would benefit FLIm for spatial dependency, spectral/channel dependency, as well as time-dependency. Another application to attention mechanism would be use of pair of images, such as intensity and lifetime, which could be mixed together in more advanced way than concatenating or weighting them together. In regards to state-of-the-art implementations of DL, ConvNeXts [218] present similar performance to the previously mentioned visual transformers model while using only pure

⁵ [kaggle.com/](https://www.kaggle.com/)

CNNs, effectively updating the standard ResNet to today's best practice. Additionally, transformers have displayed impressive results on time-series data, and could be applied to FLIm and lifetime data, in the same way as recurrent NN. We also want to highlight once more the research by [137] who used an autoencoder to generate features from the raw FLIm data, discarding the lifetime or phasor variables. This approach was possible because of the joint training with a subsequent task, here cancer classification, relying on the automatically generated features. This is akin to using classification models such as ResNet, which creates their own features representation before the final layer, and are often used this way in pre-training for feature extraction [219].

We suggest additional techniques and effort going through the training and to compensate for the lack of data in FLIm. Data augmentation can be used to increase the number of samples to train on, applying different modifications, from simple flipping or rotation for images, to more complicated approaches for different types of data [220, 221]. However, we do not suggest using the state-of-the-art data augmentation used in images classifications directly without adapting them specifically to FLIm. Examples such as Mixup [222] or Cutmix [223] may create images largely out of the expected data distribution. Few-shot learning is used to train models on a small amount of data by comparing the sample tested with other existing samples of the same class [224]. It often takes advantage of transfer learning [225], mostly through pre-training on big datasets such as Imagenet. Transfer learning is useful to train a model to recognise high level features. For example, a model may learn to recognise the shape of cells while being trained on breast histology images, which it will be able to apply to FLIm images.

As we explained, additional modalities are used to provide ground truth and annotation, such as histopathology images. If FLIm usage were to increase in clinical settings, experts would need to be trained to interpret those images as they are trained to use the current gold standards. However, this lack of direct labelling makes it more difficult for humans to verify and trust the model result.

Complimentary image modalities can offer additional information. Fluorescence and lifetime signals are slowly varying functions of wavelength/wavenumber in comparison with Raman signals. The higher information content of Raman signals yields highly specific signatures of the molecules being probed, which opens up new opportunities for ML applications [58]. Raman signals, however, are generally much weaker than fluorescence, require higher laser powers or longer exposure times and can be obscured by fluorescence and noise. Histogramming TCSPC sensors enable the simultaneous acquisition of fluorescence and Raman signals [65], offering complementary spectroscopic analysis prospects [226]. Large and growing libraries of known Raman signals

are also available for which signal similarity algorithms can be used to identify the molecule being probed [227, 228]. Together with FLIm, it could improve many medical and biomedical tasks. Additionally, time-correlated fluorescence technology could also be improved by mixing features other than the lifetime. ML algorithms may be able to pick up on hidden links between features that we do not see. For example, phasor plot variables [69] have been used together with lifetime by [149], showing better results than with phasor or lifetime features alone.

The future of FLIm imaging is intrinsically linked to the progress of both ML and hardware. As the technology advances the overall performance of FLIm devices will improve, making them more reliable and trustworthy medical technology. Performance improvements, such as speed and sensitivity, will not only enhance the quality of data collected but also aid the performance of ML algorithms applied to it. These advancements will significantly contribute to the development of *in vivo*, real-time, non-invasive FLIm imaging, which can be a game-changer in detecting various diseases, including cancer.

As ML algorithms become more sophisticated, they can help to improve the accuracy of FLIm data analysis and provide more precise diagnostic results. Additionally, hardware improvements can reduce image noise, allowing for better signal-to-noise ratios, and increasing the sensitivity of FLIm devices, which would be particularly useful in detecting early-stage diseases.

Ultimately, it is crucial to continue the investigation in both ML and hardware advancements to paving the way for the widespread use of FLIm technology in the future, transforming medical diagnosis and treatment, especially in the early detection of cancer and other diseases.

Acknowledgments

The authors have declared that no conflicting interests exist. We would like to thank Gareth O. S. Williams for the medical images used in this review. We also would like to thank Mehmet Demirel for his help and feedback during the review of this document. Additionally, we thank Chengjia Wang, for his feedback and his help when revising this paper. Qiang Wang is funded by Cancer Research UK's Early Detection and Diagnosis Primer Award (EDDPA-May22/100054), Wellcome iTPA Translational Innovation Competition (PIII053) and Accelerator Award (PIII140, cofunded with Cancer Research Horizons), and the NVIDIA Academic Hardware Grant Program. Marta Vallejo and Neil Finlayson are funded by the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) Interdisciplinary Research Collaboration (grant number EP/R005257/1—Proteus). Ali Taimori, Neil Finlayson, Tarek Haloubi, and James Hopgood are supported by EPSRC through the

Healthcare Impact Partnerships, Grant Ref EP/S025987/1, 'Next-generation sensing for human *in vivo* pharmacology- Accelerating drug development in inflammatory diseases'.


Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Dorian Gouzou  <https://orcid.org/0000-0003-4865-7784>

Ali Taimori  <https://orcid.org/0000-0001-9550-7434>

Tarek Haloubi  <https://orcid.org/0000-0003-4111-1252>

Neil Finlayson  <https://orcid.org/0000-0003-1598-9563>

Qiang Wang  <https://orcid.org/0000-0002-1665-7408>

James R Hopgood  <https://orcid.org/0000-0002-3029-2425>

Marta Vallejo  <https://orcid.org/0000-0001-9957-954X>

References

- [1] Lakowicz J R 2006 *Principles of Fluorescence Spectroscopy* ed J R Lakowicz (Springer)
- [2] Weyers B W *et al* 2019 Fluorescence lifetime imaging for intraoperative cancer delineation in transoral robotic surgery *Translational biophotonics* **1** e201900017
- [3] Weyers B W *et al* 2022 Intraoperative delineation of p16. oropharyngeal carcinoma of unknown primary origin with fluorescence lifetime imaging: Preliminary report *Head & Neck* **44** 1765–76
- [4] Phipps J E *et al* 2017 Automated detection of breast cancer in resected specimens with fluorescence lifetime imaging *Physics in Medicine & Biology* **63** 015003
- [5] Unger J *et al* 2020 Real-time diagnosis and visualization of tumor margins in excised breast specimens using fluorescence lifetime imaging and machine learning *Biomedical optics express* **11** 1216–30
- [6] Jeong S *et al* 2018 Visualization of drug distribution of a topical minocycline gel in human facial skin *Biomedical Optics Express* **9** 3434–48
- [7] Pena A M *et al* 2020 Imaging and quantifying drug delivery in skin-part 2: Fluorescence and vibrational spectroscopic imaging methods *Adv. Drug Delivery Rev.* **153** 147–68
- [8] Alex A *et al* 2018 In situ biodistribution and residency of a topical anti-inflammatory using fluorescence lifetime imaging microscopy *British Journal of Dermatology* **179** 1342–50
- [9] Berezin M Y and Achilefu S 2010 Fluorescence lifetime measurements and biological imaging *Chem. Rev.* **110** 2641–84
- [10] Marsden M *et al* 2021 Intraoperative mapping of parathyroid glands using fluorescence lifetime imaging *Journal of Surgical Research* **265** 42–8
- [11] Vishnyakova P *et al* 2023 Fluorescence lifetime imaging microscopy as an instrument for human sperm assessment *Biochem. Biophys. Res. Commun.* **645** 10–16
- [12] Williams G O *et al* 2021 Full spectrum fluorescence lifetime imaging with 0.5 nm spectral and 50 ps temporal resolution *Nat. Commun.* **12** 1–9
- [13] Poudel C, Mela I and Kaminski C F 2020 High-throughput, multi-parametric, and correlative fluorescence lifetime imaging *Methods and Applications in Fluorescence* **8** 024005
- [14] Liu C *et al* 2013 Timing and operating mode design for time-gated fluorescence lifetime imaging microscopy *The Scientific World Journal* **2013** 801901
- [15] Jordan M I and Mitchell T M 2015 Machine learning: Trends, perspectives, and prospects *Science* **349** 255–60
- [16] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
- [17] Vaswani A *et al* 2017 Attention is all you need *Advances in neural information processing systems* **30** 5998–6008
- [18] Brown T B *et al* 2020 *Language models are few-shot learners* **33** 1877–1901
- [19] Radford A *et al* 2021 *Learning transferable visual models from natural language supervision* **139** 8748–63
- [20] Lundervold A S and Lundervold A 2019 An overview of deep learning in medical imaging focusing on mri *Zeitschrift für Medizinische Physik* **29** 102–27 Special Issue: Deep Learning in Medical Physics.
- [21] Roy Y *et al* 2019 Deep learning-based electroencephalography analysis: a systematic review *J. Neural Eng.* **16** 051001
- [22] Ran A R *et al* 2021 Deep learning in glaucoma with optical coherence tomography: a review *Eye* **35** 188–201
- [23] Hormel T T *et al* 2021 Artificial intelligence in oct angiography *Progress in Retinal and Eye Research* **85** 100965
- [24] Tian L *et al* 2021 Deep learning in biomedical optics *Lasers in Surgery and Medicine* **53** 748–775
- [25] Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* **18** (Berlin: Springer) pp 234–41
- [26] Greener J G *et al* 2022 A guide to machine learning for biologists *Nat. Rev. Mol. Cell Biol.* **23** 40–55
- [27] Liu X *et al* 2019 Fast fluorescence lifetime imaging techniques: A review on challenge and development *Journal of Innovative Optical Health Sciences* **12** 1930003
- [28] Becker W 2012 Fluorescence lifetime imaging-techniques and applications *J. Microsc.* **247** 119–36
- [29] Suhling K *et al* 2015 Fluorescence lifetime imaging (flim): Basic concepts and some recent developments *Medical Photonics* **27** 3–40
- [30] Datta R *et al* 2020 Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications *J. Biomed. Opt.* **25** 1
- [31] Borst J W and Visser A J 2010 Fluorescence lifetime imaging microscopy in life sciences *Meas. Sci. Technol.* **21** 102002
- [32] Noomnarm U and Clegg R M 2009 Fluorescence lifetimes: fundamentals and interpretations *Photosynth. Res.* **101** 181–94
- [33] Li Y *et al* 2022 Lifetime determination algorithms for time-domain fluorescence lifetime imaging: A review *Fluorescence Imaging-Recent Advances and Applications*
- [34] Mannam V *et al* 2020 Machine learning for faster and smarter fluorescence lifetime imaging microscopy *Journal of Physics: Photonics* **2** 042005
- [35] Datta R *et al* 2021 Recent innovations in fluorescence lifetime imaging microscopy for biology and medicine *J. Biomed. Opt.* **26** 070603
- [36] Marcu L 2012 Fluorescence lifetime techniques in medical applications *Ann. Biomed. Eng.* **40** 304–31
- [37] Chance B *et al* 1962 Intracellular Oxidation-Reduction States in Vivo *Science* **137** 499–508
- [38] Chance B and Williams G R 2006 *The Respiratory Chain and Oxidative Phosphorylation* **17** 65–134
- [39] Lakowicz J R *et al* 1992 Fluorescence lifetime imaging of free and protein-bound NADH *Proc. Natl Acad. Sci.* **89** 1271–5
- [40] Fernandes S *et al* 2021 Solitary pulmonary nodule imaging approaches and the role of optical fibre-based technologies *European Respiratory Journal* **57** 2002537
- [41] Bianchetti G *et al* 2021 Label-free metabolic clustering through unsupervised pixel classification of multiparametric fluorescent images *Anal. Chim. Acta* **1148** 238173

- [42] Walsh A J *et al* 2013 Optical Metabolic Imaging Identifies Glycolytic Levels, Subtypes, and Early-Treatment Response in Breast Cancer *Cancer Research* **73** 6164–74
- [43] Skala M C *et al* 2007 *In vivo* multiphoton microscopy of nadh and fad redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia *Proc. Natl Acad. Sci.* **104** 19494–9
- [44] Chakraborty S *et al* 2016 Quantification of the Metabolic State in Cell-Model of Parkinson's Disease by Fluorescence Lifetime Imaging Microscopy *Sci. Rep.* **6** 19145
- [45] Minsky M 1957 Us patent# 3 013 467, microscopy apparatus *The original patent for confocal microscopy by Marvin Minsky*
- [46] Petráň M *et al* 1968 Tandem-Scanning Reflected-Light Microscope *J. Opt. Soc. Am.* **58** 661
- [47] Davidovits P and Egger M D 1969 Scanning Laser Microscope *Nature* **223** 831
- [48] Sheppard C and Choudhury A 1977 Image Formation in the Scanning Microscope *Optica Acta: International Journal of Optics* **24** 1051–73
- [49] Bugiel I, König K and Wabnitz H 1989 Investigation of cells by fluorescence laser scanning microscopy with subnanosecond time resolution **3** 47–53
- [50] Hirvonen L M and Suhling K 2017 Wide-field TCSPC: methods and applications *Meas. Sci. Technol.* **28** 012003
- [51] Munro I *et al* 2005 Toward the clinical application of time-domain fluorescence lifetime imaging *J. Biomed. Opt.* **10** 051403
- [52] Elson D S, Jo J A and Marcu L 2007 Miniaturized side-viewing imaging probe for fluorescence lifetime imaging (FLIM): validation with fluorescence dyes, tissue structural proteins and tissue specimens *New J. Phys.* **9** 127
- [53] Ranawat H, Pal S and Mazumder N 2019 Recent trends in two-photon auto-fluorescence lifetime imaging (2p-flim) and its biomedical applications *Biomedical Engineering Letters* **9** 293–310
- [54] Rehman A U *et al* 2017 Fluorescence quenching of free and bound NADH in HeLa cells determined by hyperspectral imaging and unmixing of cell autofluorescence *Biomedical Optics Express* **8** 1488
- [55] Bollinger L M and Thomas G E 1961 Measurement of the Time Dependence of Scintillation Intensity by a Delayed-Coincidence Method *Rev. Sci. Instrum.* **32** 1044–50
- [56] Becker W 2015 *Advanced Time-Correlated Single Photon Counting Applications Springer Series in Chemical Physics* ed W Becker (Springer International Publishing) vol 111
- [57] Rochas A *et al* 2003 Single photon detector fabricated in a complementary metal-oxide-semiconductor high-voltage technology *Rev. Sci. Instrum.* **74** 3263–70
- [58] Alfonso-Garcia A *et al* 2021 Mesoscopic fluorescence lifetime imaging: Fundamental principles, clinical applications and future directions *J. Biophotonics* **14** 1–27
- [59] König K 2018 1 brief history of fluorescence lifetime imaging *Multiphoton Microscopy and Fluorescence Lifetime Imaging* 3–16
- [60] Stewart H L and Birch D J S 2021 Fluorescence Guided Surgery *Methods and Applications in Fluorescence* **9** 042002
- [61] Bruschini C *et al* 2019 Single-photon avalanche diode imagers in biophotonics: review and outlook *Light: Science & Applications* **8** 87
- [62] Gyongy I, Dutton N A W and Henderson R K 2021 Direct Time-of-Flight Single-Photon Imaging *IEEE Trans. Electron Devices* **69** 1–12
- [63] Cao C *et al* 2018 Deep Learning and Its Applications in Biomedicine *Genomics, Proteomics & Bioinformatics* **16** 17–32
- [64] Kekkonen J *et al* 2019 Chemical imaging of human teeth by a time-resolved Raman spectrometer based on a CMOS single-photon avalanche diode line sensor *Analyst* **144** 6089–97
- [65] Finlayson N *et al* 2021 Time-correlated single photon raman spectroscopy at megahertz repetition rates *Opt. Lett.* **46** 4104–7
- [66] Erdogan A T *et al* 2019 A CMOS SPAD line sensor with per-pixel histogramming TDC for time-resolved multispectral imaging *IEEE J. Solid-State Circuits* **54** 1705–19
- [67] Kapitany V *et al* 2020 Data fusion for high resolution fluorescence lifetime imaging using deep learning *Computational Optical Sensing and Imaging* (Optical Society of America) CW1B–4
- [68] Dalca A V *et al* 2019 Unsupervised deep learning for bayesian brain mri segmentation *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 356–65
- [69] Digman M A *et al* 2008 The phasor approach to fluorescence lifetime imaging analysis *Biophys. J.* **94** L14–6
- [70] Ranjit S *et al* 2018 Fit-free analysis of fluorescence lifetime imaging data using the phasor approach *Nat. Protoc.* **13** 1979–2004
- [71] Silva S F, Domingues J P and Morgado A M 2018 Accurate rapid lifetime determination on time-gated flim microscopy with optical sectioning *Journal of Healthcare Engineering* **2018** 7
- [72] Wei L, Yan W and Ho D 2017 Recent advances in fluorescence lifetime analytical microsystems: Contact optics and CMOS time-resolved electronics *Sensors* **17** 2800
- [73] Rousset F *et al* 2018 Time-resolved multispectral imaging based on an adaptive single-pixel camera *Opt. Express* **26** 10550–8
- [74] Taimori A *et al* 2022 Fast and robust single-exponential decay recovery from noisy fluorescence lifetime imaging *IEEE Trans. Biomed. Eng.* **69** 3703–16
- [75] Kristoffersen A S *et al* 2014 Testing fluorescence lifetime standards using two-photon excitation and time-domain instrumentation: Rhodamine B, coumarin 6 and lucifer yellow *Journal of fluorescence* **24** 1015–24
- [76] Pande P and Jo J A 2010 Automated analysis of fluorescence lifetime imaging microscopy (flim) data based on the laguerre deconvolution method *IEEE Trans. Biomed. Eng.* **58** 172–81
- [77] Zickus V *et al* 2020 Fluorescence lifetime imaging with a megapixel spad camera and neural network lifetime estimation *Sci. Rep.* **10** 1–10
- [78] Chen Y I *et al* 2022 Generative adversarial network enables rapid and robust fluorescence lifetime image analysis in live cells *Communications Biology* **5** 1–11
- [79] Smith J T *et al* 2019 Fast fit-free analysis of fluorescence lifetime imaging via deep learning *Proc. Natl Acad. Sci.* **116** 24019–30
- [80] Xiao D, Chen Y and Li D D U 2021 One-dimensional deep learning architecture for fast fluorescence lifetime imaging *IEEE J. Sel. Top. Quantum Electron.* **27** 1–10
- [81] Ballew R M and Demas J 1989 An error analysis of the rapid lifetime determination method for the evaluation of single exponential decays *Anal. Chem.* **61** 30–3
- [82] Sharman K K *et al* 1999 Error analysis of the rapid lifetime determination method for double-exponential decays and new windowing schemes *Anal. Chem.* **71** 947–52
- [83] Sorensen D C 1982 Newton's method with a model trust region modification *SIAM J. Numer. Anal.* **19** 409–26
- [84] Conn A R, Gould N I and Toint P L 2000 *Trust region methods* (Philadelphia, PA: SIAM)
- [85] Li D U *et al* 2011 Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm *J. Biomed. Opt.* **16** 096012
- [86] Chessel A *et al* 2013 A maximum likelihood method for lifetime estimation in photon counting-based fluorescence lifetime imaging microscopy *IEEE 21st Eur Signal Process Conf* **2013** 1–5
- [87] Hall P and Selinger B 1981 Better estimates of exponential decay parameters *The Journal of Physical Chemistry* **85** 2941–6
- [88] Laurence T A and Chromy B A 2010 Efficient maximum likelihood estimator fitting of histograms *Nat. Methods* **7** 338–9
- [89] Gao K and Li D D U 2017 Estimating fluorescence lifetimes using extended Kalman filter *Electron. Lett* **53** 1027–9
- [90] Gao K and Li D D U 2018 Estimating fluorescence lifetimes using the expectation-maximisation algorithm *Electron. Lett* **54** 14–6
- [91] Jo J A, Fang Q and Marcu L 2005 Ultrafast method for the analysis of fluorescence lifetime imaging microscopy data

- based on the laguerre expansion technique *IEEE J. Sel. Top. Quantum Electron.* **11** 835–45
- [92] Kay S M 1993 *Fundamentals of statistical signal processing: Estimation theory* vol I (Englewood Cliffs, NJ: Prentice-Hall)
- [93] Wu G *et al* 2016 Artificial neural network approaches for fluorescence lifetime imaging techniques *Opt. Lett.* **41** 2561–4
- [94] Héliot L and Leray A 2021 Simple phasor-based deep neural network for fluorescence lifetime imaging microscopy *Sci. Rep.* **11** 1–13
- [95] Yao R *et al* 2019 Net-flics: fast quantitative wide-field fluorescence lifetime imaging with compressed sensing—a deep learning approach *Light: Science & Applications* **8** 1–7
- [96] Ochoa M *et al* 2020 High compression deep learning based single-pixel hyperspectral macroscopic fluorescence lifetime imaging in vivo *Biomedical Optics Express* **11** 5401–24
- [97] Pian Q *et al* 2017 Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging *Nat. Photonics* **11** 411–4
- [98] Philip J and Carlsson K 2003 Theoretical investigation of the signal-to-noise ratio in fluorescence lifetime imaging *JOSA A* **20** 368–79
- [99] Köllner M and Wolfrum J 1992 How many photons are necessary for fluorescence-lifetime measurements? *Chem. Phys. Lett.* **200** 199–204
- [100] Trinel D *et al* Upgrading time domain FLIM using an adaptive Monte Carlo data inflation algorithm *Cytometry Part A* **July** 528–37
- [101] Li D *et al* 2008 On-chip, time-correlated, fluorescence lifetime extraction algorithms and error analysis *Journal of the Optical Society of America A* **25** 1190
- [102] Hajian-Tilaki K 2013 Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation *Caspian journal of internal medicine* **4** 627–35
- [103] Metz C E 2008 ROC analysis in medical imaging: a tutorial review of the literature *Radiological Physics and Technology* **1** 2–12
- [104] Wang Q, Hopgood J R and Vallejo M 2021 Fluorescence lifetime imaging endomicroscopy based ex-vivo lung cancer prediction using multi-scale concatenated-dilation convolutional neural networks *Medical Imaging 2021: Computer-Aided Diagnosis* (International Society for Optics and Photonics) vol 11597, p 1159720
- [105] Chen B *et al* 2019 Support vector machine classification of nonmelanoma skin lesions based on fluorescence lifetime imaging microscopy *Anal. Chem.* **91** 10640–7
- [106] Haskins G, Kruger U and Yan P 2020 Deep learning in medical image registration: A survey *Mach. Vis. Appl.* **31** 1–18
- [107] Cao X *et al* 2020 Chapter 14—image registration using machine and deep learning *Handbook of Medical Image Computing and Computer Assisted Intervention The Elsevier and MICCAI Society Book Series* ed S K Zhou, D Rueckert and G Fichtinger (New York: Academic) pp 319–42
- [108] Wang Y *et al* 2016 Rapid, label-free, and highly sensitive detection of cervical cancer with fluorescence lifetime imaging microscopy *IEEE J. Sel. Top. Quantum Electron.* **22** 228–34
- [109] Marsden M *et al* 2020 Intraoperative margin assessment in oral and oropharyngeal cancer using label-free fluorescence lifetime imaging and machine learning *IEEE Trans. Biomed. Eng.* **68** 857–68
- [110] Wang Q *et al* 2022 Deep learning-assisted co-registration of full-spectral autofluorescence lifetime microscopic images with h&e-stained histology images *Communication Biology* **5** 8–12
- [111] Becker W, Bergmann A and Biskup C 2007 Multispectral fluorescence lifetime imaging by tcspc *Microsc. Res. Tech.* **70** 403–9
- [112] Madabhushi A and Lee G 2016 Image analysis and machine learning in digital pathology: Challenges and opportunities *Med. Image Anal.* **33** 170–5 20th Anniversary of the Medical Image Analysis journal (MedIA)
- [113] Rivenson Y *et al* 2019 Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning *Nature Biomedical Engineering* **3** 466–77
- [114] Taqi S A *et al* 2018 A review of artifacts in histopathology *Journal of oral and maxillofacial pathology: JOMFP* **22** 279
- [115] Bird B *et al* 2012 Infrared spectral histopathology (shp): a novel diagnostic tool for the accurate classification of lung cancer *Lab. Invest.* **92** 1358–73
- [116] Unger J *et al* 2018 Method for accurate registration of tissue autofluorescence imaging data with corresponding histology: a means for enhanced tumor margin assessment *J. Biomed. Opt.* **23** 11
- [117] Jiang J *et al* 2019 Robust hierarchical density estimation and regression for re-stained histological whole slide image co-registration *PLOS ONE* **14** 5–9
- [118] Ge L *et al* 2022 Unsupervised histological image registration using structural feature guided convolutional neural network *IEEE Trans. Med. Imaging* **41** 2414–31
- [119] Wang Q *et al* 2020 Deep learning in ex-vivo lung cancer discrimination using fluorescence lifetime endomicroscopic images 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (Piscataway, NJ: IEEE) pp 1891–4
- [120] Bu L, Shen B and Cheng Z 2014 Fluorescent imaging of cancerous tissues for targeted surgery *Adv. Drug Delivery Rev.* **76** 21–38
- [121] Zhao X *et al* 2018 A deep learning model integrating fcnnns and crfs for brain tumor segmentation *Med. Image Anal.* **43** 98–111
- [122] Fozouni P *et al* 2021 Amplification-free detection of sars-cov-2 with crispr-cas13a and mobile phone microscopy *Cell* **184** 323–33
- [123] Megia-Fernandez A *et al* 2021 Red-shifted environmental fluorophores and their use for the detection of gram-negative bacteria *Chemosensors* **9** 117
- [124] Zamani H *et al* 2021 Elliptical shape recovery from blurred pixels using deep learning *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Piscataway, NJ: IEEE) pp 2775–9
- [125] HU D 2014 Automatic segmentation of fluorescence lifetime microscopy images of cells using multiresolution community detection—a first study *J. Microsc.* **253** 54–64
- [126] Vallmitjana A, Torrado B and Gratton E 2021 Phasor-based image segmentation: machine learning clustering techniques *Biomedical Optics Express* **12** 3410
- [127] Gonzales R C and Woods R 2006 *Digital Image Processing (3rd Edition)* (USA: Prentice-Hall, Inc.) 013168728X (<https://doi.org/10.5555/1076432>)
- [128] Otsu N 1979 A threshold selection method from gray-level histograms *IEEE transactions on systems, man, and cybernetics* **9** 62–6
- [129] Comaniciu D and Meer P 2002 Mean shift: A robust approach toward feature space analysis *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 603–19
- [130] Kalman R E 1960 A new approach to linear filtering and prediction problems *ASME. J. Basic Eng* **82** 35–45
- [131] Viola P and Jones M J 2004 Robust real-time face detection *Int. J. Comput. Vision* **57** 137–54
- [132] Kalal Z, Mikolajczyk K and Matas J 2011 Tracking-learning-detection *IEEE Trans. Pattern Anal. Mach. Intell.* **34** 1409–22
- [133] Zhao Z Q *et al* 2019 Object detection with deep learning: A review *IEEE Trans Neural Netw. Learn. Syst.* **30** 3212–32
- [134] Mannam V *et al* 2021 *Convolutional Neural Network Denoising in Fluorescence Lifetime Imaging Microscopy (FLIM) Multiphoton Microscopy in the Biomedical Sciences XXI* 116481C (SPIE)
- [135] Mukherjee S *et al* 2023 Domain adapted multi-task learning for segmenting amoeboid cells in microscopy *IEEE Trans. Med. Imaging* **42** 42–45
- [136] Jo J A *et al* 2018 Endogenous fluorescence lifetime imaging (flim) endoscopy for early detection of oral cancer and dysplasia 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (Piscataway, NJ: IEEE) pp 3009–12
- [137] Caughlin K *et al* 2021 End-to-end neural network for feature extraction and cancer diagnosis of in vivo fluorescence

- lifetime images of oral lesions 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (Piscataway, NJ: IEEE) pp 3894–7
- [138] Duran-Sierra E *et al* 2021 Machine-learning assisted discrimination of precancerous and cancerous from healthy oral tissue based on multispectral autofluorescence lifetime imaging endoscopy *Cancers* **13** 4751
- [139] Cosci A *et al* 2016 Time-resolved fluorescence spectroscopy for clinical diagnosis of actinic cheilitis *Biomedical Optics Express* **7** 4210–9
- [140] Yang Q *et al* 2020 Classification of skin cancer based on fluorescence lifetime imaging and machine learning *Optics in Health Care and Biomedical Optics X* **11553** 115531Y International Society for Optics and Photonics
- [141] Romano R A *et al* 2020 Multispectral autofluorescence dermoscopy for skin lesion assessment *Photodiagn. Photodyn. Ther.* **30** 101704
- [142] Vasanthakumari P *et al* 2022 Discrimination of cancerous from benign pigmented skin lesions based on multispectral autofluorescence lifetime imaging dermoscopy and machine learning *J. Biomed. Opt.* **27** 066002
- [143] Gu J 2014 Investigation of fluorescence lifetime imaging for quantitative cervical cancer screening and diagnosis *PhD dissertation* Nanyang Technological University
- [144] Gu J *et al* 2015 Enhancement of early cervical cancer diagnosis with epithelial layer analysis of fluorescence lifetime images *PLoS One* **10** e0125706
- [145] Sahoo G R *et al* 2018 Improving diagnosis of cervical precancer: Combination of pca and svm applied on fluorescence lifetime images *Photonics* (Multidisciplinary Digital Publishing Institute) vol 5, p 57
- [146] Ji M *et al* 2022 Early detection of cervical cancer by fluorescence lifetime imaging microscopy combined with unsupervised machine learning *Int. J. Mol. Sci.* **23** 11476
- [147] Butte P V *et al* 2011 Fluorescence lifetime spectroscopy for guided therapy of brain tumors *Neuroimage* **54** S125–35
- [148] Walsh A J *et al* 2021 Classification of t-cell activation via autofluorescence lifetime imaging *Nature biomedical engineering* **5** 77–88
- [149] Dunkers J P *et al* 2021 Toward absolute viability measurements for bacteria *J. Biophotonics* **14** e202100175
- [150] Qian T *et al* 2021 Label-free imaging for quality control of cardiomyocyte differentiation *Nat. Commun.* **12** 1–11
- [151] Cardona E N and Walsh A J 2022 Identification of rare cell populations in autofluorescence lifetime image data *Cytometry Part A* **101** 497–506
- [152] Neto N G *et al* 2022 Non-invasive classification of macrophage polarisation by 2p-flim and machine learning *Elife* **11** e77373
- [153] Kröger M *et al* 2022 Label-free imaging of m1 and m2 macrophage phenotypes in the human dermis in vivo using two-photon excited flim *Elife* **11** e72819
- [154] Koujan M R *et al* 2018 Multi-class classification of pulmonary endomicroscopic images *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (Washington, DC, USA)* (<https://doi.org/10.1109/ISBI.2018.8363874>)
- [155] Lenseigne B *et al* 2007 Support vector machines for automatic detection of tuberculosis bacteria in confocal microscopy images *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (Arlington, VA, USA)* **85–88**
- [156] Seth S *et al* 2018 Estimating bacterial and cellular load in fcfm imaging *Journal of Imaging* **4** 11
- [157] Eldaly A K *et al* 2019 Bayesian bacterial detection using irregularly sampled optical endomicroscopy images *Medical Image Analysis* **57** 18–31
- [158] Eldaly A K *et al* 2019 Patch-based sparse representation for bacterial detection *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (Venice, Italy)* **657–61**
- [159] Rudkouskaya A *et al* 2020 *Macroscopic Fluorescence Lifetime-based Förster Resonance Energy Transfer Imaging for Quantitative Ligand-receptor Binding* (Elsevier)
- [160] Zhou M, Li Q and Wang R 2016 Current experimental methods for characterizing protein-protein interactions *ChemMedChem* **11** 4
- [161] Ochoa M *et al* 2022 Macroscopic Fluorescence Lifetime Imaging for Monitoring of Drug–Target Engagement *Biomedical Engineering Technologies: Volume 2* ed A Rasooly, H Baker and M R Ossandon (Springer US) pp 837–56
- [162] Yang X *et al* 2023 A live-cell image-based machine learning strategy for reducing variability in psc differentiation systems *Cell Discovery* **9** 6
- [163] Chen H M *et al* 2005 Time-resolved autofluorescence spectroscopy for classifying normal and premalignant oral tissues *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery* **37** 37–45
- [164] Pfefer T J *et al* 2003 Temporally and spectrally resolved fluorescence spectroscopy for the detection of high grade dysplasia in barrett’s esophagus *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery* **32** 10–6
- [165] Brancalion L *et al* 2001 *In vivo* fluorescence spectroscopy of nonmelanoma skin cancer *Photochem. Photobiol.* **73** 178–83
- [166] Richards-Kortum R and Sevick-Muraca E 1996 Quantitative optical spectroscopy for tissue diagnosis *Annu. Rev. Phys. Chem.* **47** 555–606
- [167] Alfano R *et al* 1987 Fluorescence spectra from cancerous and normal human breast and lung tissues *IEEE J. Quantum Electron.* **23** 1806–11
- [168] Levitt J M *et al* 2011 Automated biochemical, morphological, and organizational assessment of precancerous changes from endogenous two-photon fluorescence images *PLoS One* **6** e24765
- [169] Liu L *et al* 2019 Fluorescence lifetime imaging microscopy and its applications in skin cancer diagnosis *Journal of Innovative Optical Health Sciences* **12** 1930004
- [170] Pires L *et al* 2014 Time-resolved fluorescence lifetime for cutaneous melanoma detection *Biomedical optics express* **5** 3080–9
- [171] Wang M *et al* 2017 Rapid diagnosis and intraoperative margin assessment of human lung cancer with fluorescence lifetime imaging microscopy. *bba clin. bba clin.* **8** 7–13
- [172] T. M. Inc., Matlab, 2022. [Online]. Available: <https://mathworks.com>
- [173] Python Core Team, Python: A dynamic, open source programming language. Python Software Foundation., 2022. [Online]. Available: <https://python.org/>
- [174] Pedregosa F *et al* 2011 Scikit-learn: Machine learning in Python *Journal of Machine Learning Research* **12** 2825–30
- [175] Flach P 2012 *Machine learning: the art and science of algorithms that make sense of data* (Cambridge university press)
- [176] Grandini M, Bagli E and Visani G 2020 Metrics for multi-class classification: an overview arXiv:2008.05756
- [177] Eelbode T *et al* 2020 Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index *IEEE Trans. Med. Imaging* **39** 3679–90
- [178] Vasanthakumari P *et al* 2020 Classification of skin-cancer lesions based on fluorescence lifetime imaging *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging* vol 11 317 (International Society for Optics and Photonics) 113170Z
- [179] Gu J *et al* 2014 Quantitative diagnosis of cervical neoplasia using fluorescence lifetime imaging on haematoxylin and eosin stained tissue sections *J. Biophotonics* **7** 483–91
- [180] He K *et al* 2016 Deep residual learning for image recognition *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR* 770–8
- [181] Szegedy C *et al* 2016 Rethinking the inception architecture for computer vision *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR* 2818–26
- [182] Tan M and Le Q 2019 Efficientnet: Rethinking model scaling for convolutional neural networks *International Conference on Machine Learning, PMLR ICML* 6105–14
- [183] Tan M and Le Q 2021 Efficientnetv2: Smaller models and faster training *International Conference on Machine Learning, PMLR ICML* 10096–106
- [184] Tajbakhsh N *et al* 2016 Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging* **35** 1299–312

- [185] Zhou Z *et al* 2018 Unet++: A nested u-net architecture for medical image segmentation *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4* (Springer) pp 3–11
- [186] Huang H *et al* 2020 Unet 3+: A full-scale connected unet for medical image segmentation *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE ICASSP 1055–9
- [187] Yankelevich D R *et al* 2014 Design and evaluation of a device for fast multispectral time-resolved fluorescence spectroscopy and imaging *Rev. Sci. Instrum.* **85** 034303
- [188] Cheng S *et al* 2014 Handheld multispectral fluorescence lifetime imaging system for in vivo applications *Biomedical optics express* **5** 921–31
- [189] Huang S, Heikal A A and Webb W W 2002 Two-photon fluorescence spectroscopy and microscopy of nad (p) h and flavoprotein *Biophys. J.* **82** 2811–25
- [190] Chicco D 2017 Ten quick tips for machine learning in computational biology *BioData mining* **10** 35
- [191] Zhu Q 2020 On the performance of matthews correlation coefficient (mcc) for imbalanced dataset *Pattern Recognit. Lett.* **136** 71–80
- [192] Ma Y *et al* 2020 Rose: a retinal oct-angiography vessel segmentation dataset and new model *IEEE Trans. Med. Imaging* **40** 928–39
- [193] Li M *et al* 2020 Image projection network: 3d to 2d image segmentation in octa images *IEEE Trans. Med. Imaging* **39** 3343–54
- [194] Li M *et al* 2020 OCTA-500: A Retinal Dataset for Optical Coherence Tomography Angiography Study arXiv:2012.07261
- [195] Hu K *et al* 2022 Joint-seg: Treat foveal avascular zone and retinal vessel segmentation in octa images as a joint task *IEEE Trans. Instrum. Meas.* **71** 1–13
- [196] Peng L *et al* 2021 Fargo: A joint framework for faz and rv segmentation from octa images *Ophthalmic Medical Image Analysis: 8th International Workshop, OMIA 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 8 OMIA - MICCAI* (Springer) pp 42–51
- [197] Liu X *et al* 2022 Vct-net: An octa retinal vessel segmentation network based on convolution and transformer *2022 IEEE International Conference on Image Processing (ICIP)*, IEEE ICIP 2656–60
- [198] Litjens G *et al* 2018 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset *GigaScience* **7** giy065
- [199] Hanczar B, Bourgeois V and Zehraoui F 2022 Assessment of deep learning and transfer learning for cancer prediction based on gene expression data *BMC Bioinf.* **23** 1–23
- [200] Sevakula R K *et al* 2018 Transfer learning for molecular cancer classification using deep neural networks *IEEE/ACM Trans. Comput. Biol. Bioinf.* **16** 2089–100
- [201] Hadad O *et al* 2017 Classification of breast lesions using cross-modal deep learning *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, IEEE ISBI 109–12
- [202] Paszke A *et al* 2019 Pytorch: An imperative style, high-performance deep learning library *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems NIPS*
- [203] Abadi M *et al* 2016 Tensorflow: Large-scale machine learning on heterogeneous distributed systems <https://arxiv.org/abs/1603.04467>
- [204] Deng J *et al* 2009 Imagenet: A large-scale hierarchical image database *2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee CVPR* 248–55
- [205] He T *et al* 2019 Bag of tricks for image classification with convolutional neural networks *Proceedings of the IEEE/CVF Conference on computer vision and Pattern Recognition CVPR* 558–67
- [206] DeTone D, Malisiewicz T and Rabinovich A 2018 Superpoint: Self-supervised interest point detection and description *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition workshops CVPR* 224–36
- [207] Sarlin P E *et al* 2020 Superglue: Learning feature matching with graph neural networks *Proceedings of the IEEE/CVF Conference on Computer vision and pattern Recognition CVPR* 4938–47
- [208] Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift *International conference on machine learning, pmlr ICML* 448–56
- [209] Ba J L, Kiros J R and Hinton G E 2016 Layer normalization arXiv:1607.06450
- [210] Nair V and Hinton G E 2010 Rectified linear units improve restricted boltzmann machines *Proceedings of the 27th international conference on machine learning (ICML-10) ICML* 807–14
- [211] Hendrycks D and Gimpel K 2016 Gaussian error linear units (gelus) arXiv:1606.08415
- [212] Hu J, Shen L and Sun G 2018 Squeeze-and-excitation networks *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **2018** 7132–41
- [213] Zhang H *et al* 2022 Resnest: Split-attention networks *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR* 2022, 2736–46
- [214] Woo S *et al* 2018 Cbam: Convolutional block attention module *Proceedings of the European conference on computer vision (ECCV) ECCV* 2018, 3–19
- [215] Fu J *et al* 2019 Dual attention network for scene segmentation *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR* 2019, 3146–54
- [216] Dosovitskiy A *et al* 2020 An image is worth 16x16 words: Transformers for image recognition at scale *9th International Conference on Learning Representations ICLR* arXiv:2010.11929
- [217] Liu Z *et al* 2021 Swin transformer: Hierarchical vision transformer using shifted windows *Proceedings of the IEEE/CVF international conference on computer vision ICCV* 2021, 10012–22
- [218] Liu Z *et al* 2022 A convnet for the 2020s *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR* 2022, 11976–86
- [219] Deniz E *et al* 2018 Transfer learning based histopathologic image classification for breast cancer detection *Health information science and systems* **6** 1–7
- [220] Chlap P *et al* 2021 A review of medical image data augmentation techniques for deep learning applications *Journal of Medical Imaging and Radiation Oncology* **65** 545–63
- [221] Shin H C *et al* 2018 Medical image synthesis for data augmentation and anonymization using generative adversarial networks *Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3 SASHIMI - MICCAI* (Springer) 1–11
- [222] Zhang H *et al* 2017 Mixup: Beyond empirical risk minimization arXiv:1710.09412
- [223] Yun S *et al* 2019 Cutmix: Regularization strategy to train strong classifiers with localizable features *Proceedings of the IEEE/CVF international conference on computer vision ICCV (Seoul, Korea)* **6023–32**
- [224] Wang Y *et al* 2020 Generalizing from a few examples: A survey on few-shot learning *ACM computing surveys (csur)* **53** 1–34
- [225] Alzubaidi L *et al* 2020 Towards a better understanding of transfer learning for medical imaging: a case study *Applied Sciences* **10** 4523
- [226] Bec J *et al* 2020 Investigating origins of flim contrast in atherosclerotic lesions using combined flim-raman spectroscopy *Frontiers in Cardiovascular Medicine* **7** 122
- [227] Luo C *et al* 2018 Cosine normalization: Using cosine similarity instead of dot product in neural networks *International Conference on Artificial Neural Networks ICANN* (Springer) pp 382–91
- [228] Hu Q *et al* 2022 Biological characteristics of cell similarity measure *Advanced Intelligent Systems* **4** 2100093