

Fluorescence Lifetime Endomicroscopic Image-based ex-vivo Human Lung Cancer Differentiation Using Machine Learning

Qiang Wang ¹, Marta Vallejo ¹, and James Hopgood ¹

¹Affiliation not available

October 30, 2023

Abstract

Over 20,000 fluorescence lifetime images from 10 patients were collected using a fibre-based custom fluorescence lifetime imaging endomicroscopy (FLIM) system. During the data collection, various measuring conditions were applied, including exposure time, optical wavelength, and lifetime extraction approaches to obtain diverse results rich in spatial and spectral resolution. The data for further processing was chosen with exposure time of 6 and 20 ns, excitation bands of 490-570 and 594-764 nm, and RLD. In addition, there are some images with sizes different than 128x128. In order to avoid any artificial errors on the lifetime images during the processing, only the lifetime images with 128x128 resolution were selected. After the selection, there were 10,155 and 11,363 frames of cancer and normal tissues respectively, and each frame contained one intensity and one corresponding lifetime image.

Fluorescence Lifetime Endomicroscopic Image-based ex-vivo Human Lung Cancer Differentiation Using Machine Learning

Qiang Wang¹, Marta Vallejo^{2*}, James R. Hopgood^{3*}

1 Centre for Inflammation Research, University of Edinburgh, UK

2 School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK*

3 School of Engineering, Institute for Digital Communications, University of Edinburgh, UK*

Abstract

Cancerous and normal tissue are different in their fluorescent lifetime. However, conventional approaches to derive the lifetime are based mainly on statistical information with auxiliary information, such as pathology images, or machine learning technologies with handcrafted features. This study reports preliminary results of using machine learning methods for the automatic discrimination of human lung cancer with fluorescence lifetime imaging endomicroscopy (FLIM). Over 20,000 fluorescence lifetime images from 10 patients were collected using a fibre-based custom FLIM system. Using the pixel values of the lifetime images as input, four different machine learning technologies, namely, K-nearest neighbour (KNN), support-vector classifier (SVC), neural network (NN), and random forest (RF), were applied for the classification. The training was performed on nine sets of the data using 10-folder cross-validation to obtain the optimal parameters, and testing was performed on the remaining dataset using the optimised parameters. By repeating the procedure three times, the area under the receiver operating characteristic (ROC) curve (AUC) scores were averaged on individual results from three repetitions. Overall, all methods except KNN achieved acceptable scores, while SVC obtained the highest AUC, slightly better than NN and RF.

1. Introduction

Machine learning (ML) has achieved remarkable success in detecting cancer of different organs with the images at macro and/or micro level from various medical instruments, e.g. computed tomography and microscopy [1]. Particularly, fluorescence lifetime techniques have been successfully applied for various medical applications to diagnose disease and cancer [2–4]. However, there are a few papers in the literature focusing on the integration of ML techniques with fluorescence lifetime images for cancer classification. Lin et al. [5] combined principal component analysis (PCA) with a support-vector classifier (SVC)¹ to classify the spectral data from autofluorescence spectroscopy to distinguish nasopharyngeal carcinomas from normal tissues, yielding over 94% diagnostic accuracy. Majumder et al. [6, 7] proposed PCA with SVC and its variants to estimate oral cancer with laser-induced fluorescence spectroscopy, which achieved over 85% scores on independent testing samples. However, all those studies were based on fluorescence intensity images, rather than the lifetime. Chen et al. [8] reported the discrimination of three different types of skin cancer

¹ Support-vector classifier or SVC is used as an alternative name of support-vector machine (SVM)

using SVC and fluorescence lifetime imaging endomicroscopy (FLIM) data. Their approach, however, was based on handcrafted features derived from the lifetime reconstruction results. Interestingly, all these studies used SVC with handcrafted features from fluorescence data, implying the prevalence and robustness of SVC. So far, little effort has been made on automatic discrimination of human disease/cancer directly on FLIM images by machine learning.

In this study, we address two fundamental problems: whether fluorescence lifetime data could be directly used for lung cancer classification, and whether MLs are feasible to be employed as automatic detection tools in this context. In this regard, we apply four prevalent ML algorithms, namely K-nearest neighbour (KNN), SVC, neural network (NN), and random forest (RF), to over 20,000 fluorescence lifetime images collected by a fibre-based custom fluorescence lifetime imaging endomicroscopy, with various user-specified configurations. The models are trained and validated using 10-folder cross-validation on the datasets from nine patients and tested on the data from the remaining patient. This procedure is repeated three times to obtain three different sets of metrics. The final AUC score is derived by averaging the obtained metrics to quantify the performance of the models. Scikit-learn [15] is utilised to perform ML-related processing in the present investigation.

The rest of the report is arranged as follows. Section 2 reviews the MLs techniques used in the evaluation. Section 3 introduces the methodology of data collection and image preprocessing. Results are presented in Section 4, followed by the conclusions in Section 5.

2. Background

2.1. *Machine Learning Techniques*

A KNN classifier is a type of instance-based learning, which does not find out an internal relationship (model) among the given variables, but simply stores the instances of the data, and predicts a query sample according to its nearest neighbours with a certain metric, e.g. Euclidean distance [10, 11]. Due to its conceptual and computational simplicity, KNN requires the least time and resources for classification but sacrificing prediction accuracy.

As a non-linear discriminative classifier, the SVC was firstly introduced by Boser et al. [12] and has been widely applied for classifying multivariate data since then. The algorithm outputs an optimal hyperplane that maximises the margin between the data belonging to different classes and, therefore, separates the data with the decision boundaries. A significant characteristic of SVC is that it supports different functions, called kernels, to solve the problems with different complexity. For example, non-linearly separated data can be classified by being transformed into a higher-dimension space, so that a linear boundary can be found.

A NN classifier is a computational model inspired by how biological systems process information. Typically, a NN classifier is a linear combination of several non-linear activation functions, which transfers a number of given examples (input layer) into a certain number of outputs (output layer) associated with the number of classes to be separated [10, 11]. The stacked activation functions, along with relevant parameters, between input and output

layers are called hidden layers. Nowadays, NNs have been widely applied to various areas. They achieved huge success thanks to the fast development of computational power, the increment in the number of hidden layers, and the development of diverse architectures, which are referred as deep neural networks [13]. In this study, a feed-forward neural network with two hidden layers is used.

As an ensemble learning approach, the RF classifier combines decision tree predictors to vote for the most popular class for a given example [14]. A major advantage is that, by randomly sampling many independent decision trees, their individual correlation is minimised, and hence the prediction by each tree is relatively uncorrelated, making the overall prediction less bias.

3. Methods

3.1. Image filtering

During the data collection, various measuring conditions were applied, including exposure time, optical wavelength, and lifetime extraction approaches to obtain diverse results rich in spatial and spectral resolution. However, this may introduce some variance in lifetime estimation. For example, lifetime values estimated by rapid lifetime determination (RLD) and the centre-of-mass method (CMM) are significantly different. Therefore, the data for further processing was chosen with exposure time of 6 and 20 ns, excitation bands of 490-570 and 594-764 nm, and RLD. In addition, there are some images with sizes different than 128x128. In order to avoid any artificial errors on the lifetime images during the processing, only the lifetime images with 128x128 resolution were selected. After the selection, there were 10,155 and 11,363 frames of cancer and normal tissues respectively, and each frame contained one intensity and one corresponding lifetime image.

3.2. Image pre-processing

The overall schema of the pre-processing includes thresholding, normalisation, and refactoring. As mentioned previously, each frame contains an intensity and lifetime image. Let $F_a^i = \{f_{x,y}^a | x, y \in \{1, 2, \dots, N\}\}$ denote i th frame, a is either it or lf representing the intensity or lifetime image of the frame, $f_{x,y}^a$ is the pixel at x and y location, and N is the size of the images assuming that all images are square. Accordingly, the thresholding is defined as:

$$D_{it}^i = \begin{cases} 0 & f_{x,y}^{it} < threshold \\ f_{x,y}^{it} & f_{x,y}^{it} \geq threshold \end{cases} \quad (1)$$

$$D_{lf}^i = \begin{cases} 0 & f_{x,y}^{lf} < threshold \\ f_{x,y}^{lf} & f_{x,y}^{lf} \geq threshold \end{cases} \quad (2)$$

where D_a^i represents the denoised intensity and lifetime images. Afterwards, the thresholded images are normalised using the dark background and lightfield data collected during the experiments. The normalisation is governed by:

$$N_a^i(x, y) = \frac{GB(D_a^i(x, y) - D_e(x, y))}{GB(L_e(x, y) - D_e(x, y))} \quad (3)$$

where D_e and L_e are the dark background and lightfield images relating to the input images D_a^i , N_a^i is the normalised intensity and lifetime images, and GB is a 3x3 Gaussian smoothing filter defined by [9]:

$$GB(x, y) = 1/(2\pi\sigma^2)e^{-(x^2+y^2)/(2\sigma^2)} \quad (4)$$

where σ is the standard deviation of the Gaussian distribution which is empirically set to two in this study. OpenCV library [9] is utilised during the image pre-processing. Eventually, FLIM images are row-wise flattened to 1-dimensional vector as the input to the MLs.

3.3. Dimensionality reduction

The dimensions of the lifetime image are 128x128, and the pixel values are utilised for classification, but they are sparse and contain many zero values, including boundary and zero lifetime pixels. Dimensionality reduction is required so that the most important features are retained, in this case, the most important features refer to pixels that “contribute” the most. The employed technique is PCA, which in principle projects the high-dimension input data into a lower-dimension space so that the variance of the data can be best explained [10, 11]. In this study, the percentage of variance explained by the PCA-selected components is set to 95%. After the reduction, the number of features decreases from the number of pixel values in the flattened lifetime image (16,384), to 2100.

3.4. Training and validation of machine learning algorithms

There are, in total, 10,155 and 11,363 frames of cancer and normal tissue respectively, and each frame contains a fluorescence intensity and the corresponding lifetime image. For the training, 10-folder cross-validation is applied to optimise the hyperparameters of the ML methods using the data collected on nine patients. After the fine-tuning of the hyperparameters with grid search, the values that achieved the best scores are obtained as listed in Table 1 and will be used for the independent evaluation. Predictions are derived and aggregated to calculate average accuracy and AUC.

Table 1: Hyperparameters of the MLs and fine-tuned results

ML	Hyperparameters to be tuned	Best value		
		Test set 1	Test set 2	Test set 3
KNN	Number of neighbours (1, 2, 3, 4, 5, 6)	3	3	3
SVC	Regularisation parameter (0.001, 0.01, 0.1, 1, 10, 100)	10	10	10
	Gamma (0.001, 0.01, 0.1, 1, 10, 100)	10	10	10
NN	Learning rate (1, 0.1, 0.01, 0.001, 0.0001)	0.0001	0.0001	0.0001
	Hidden layers with units ((64, 32), (128, 64), (256, 64))	(256, 64)	(256, 64)	(256, 64)
RF	Max features (16, 32, 64, 128, 256)	256	256	256
	Number of estimators (100, 500, 1000, 2000, 3000)	3000	3000	3000

4. Results

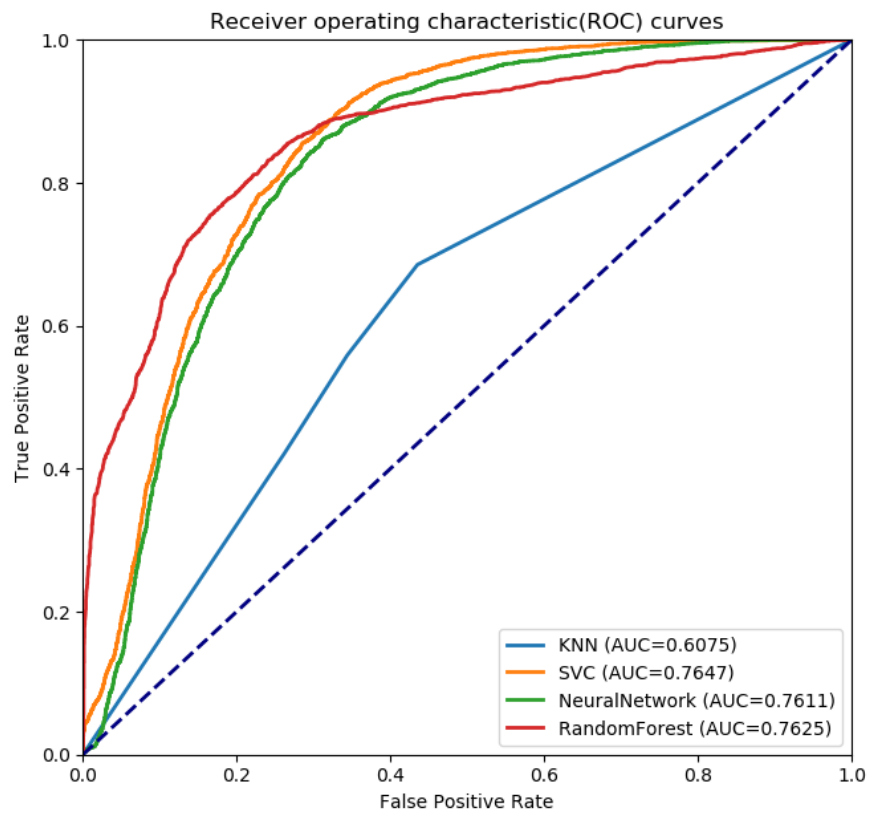


Figure 1: ROC curves and AUC scores reached by the ML techniques on the testing datasets

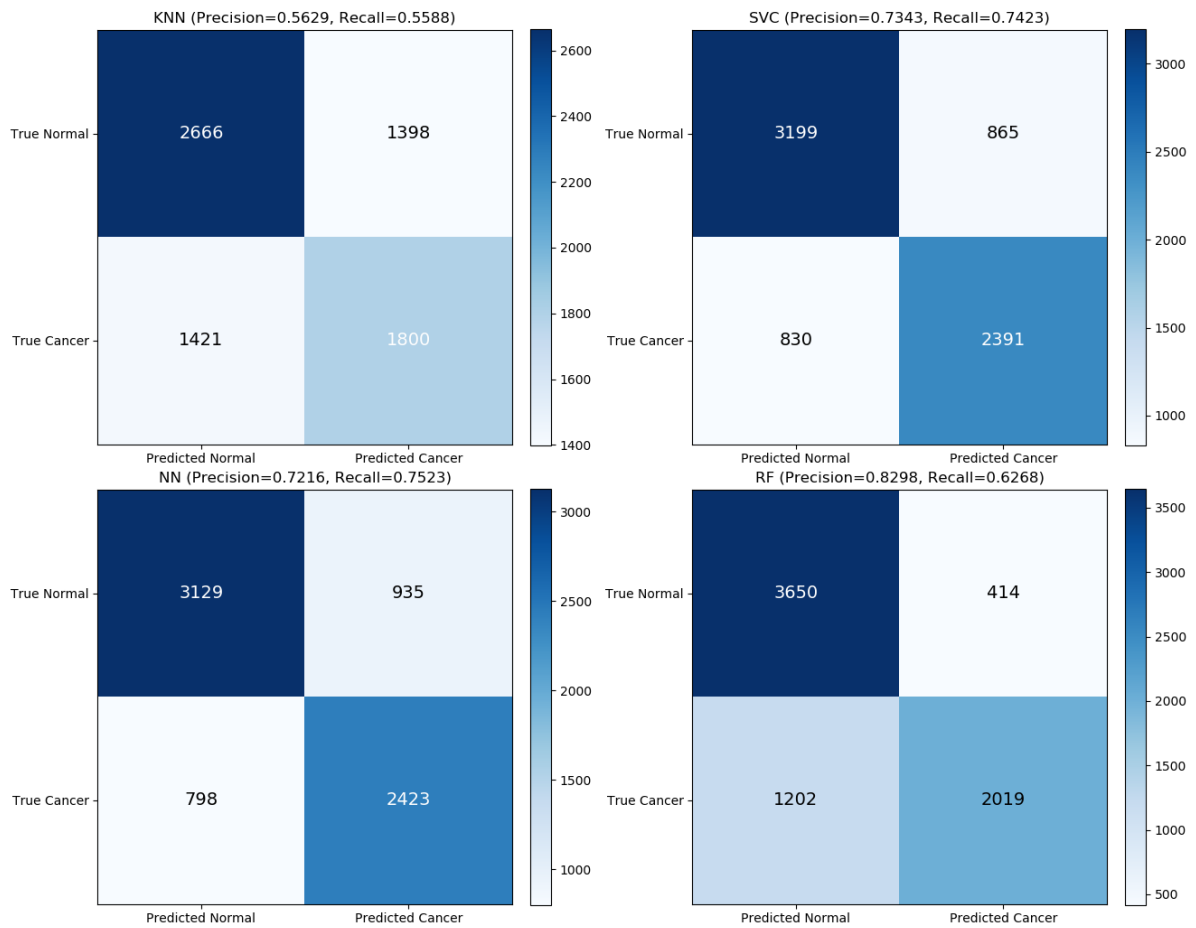
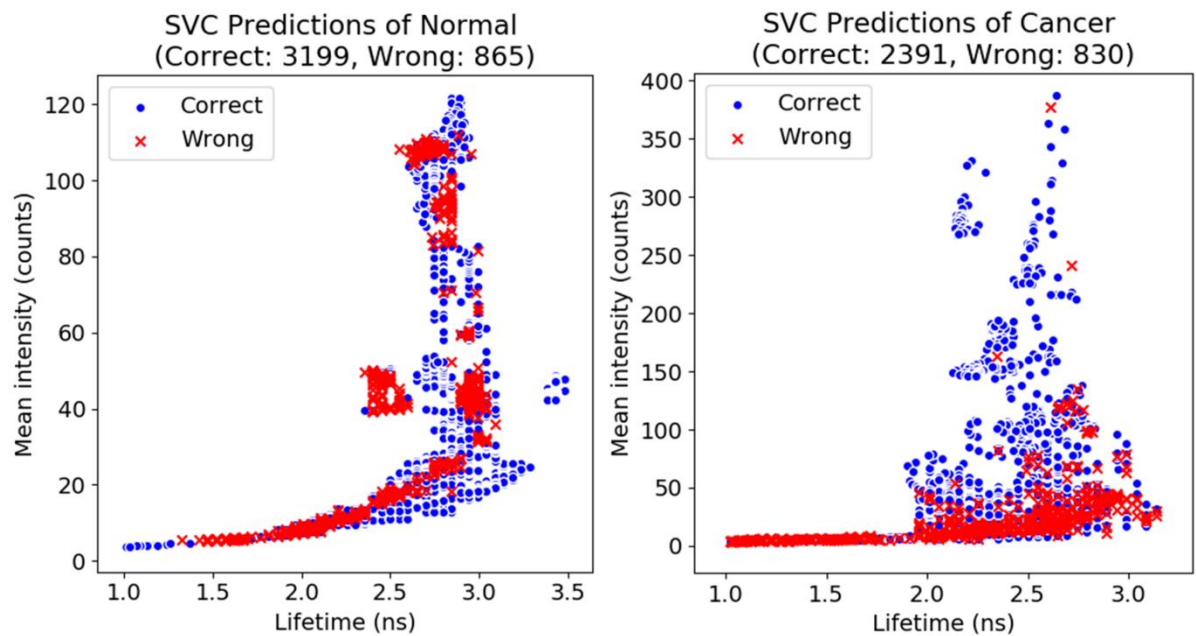
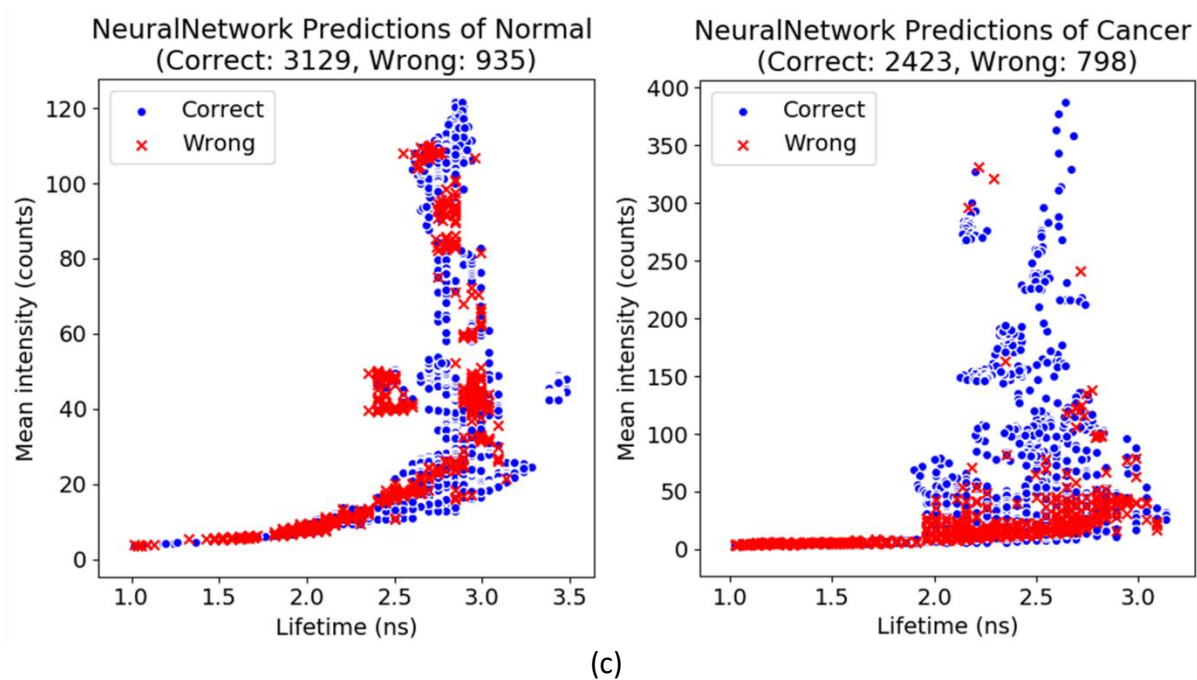
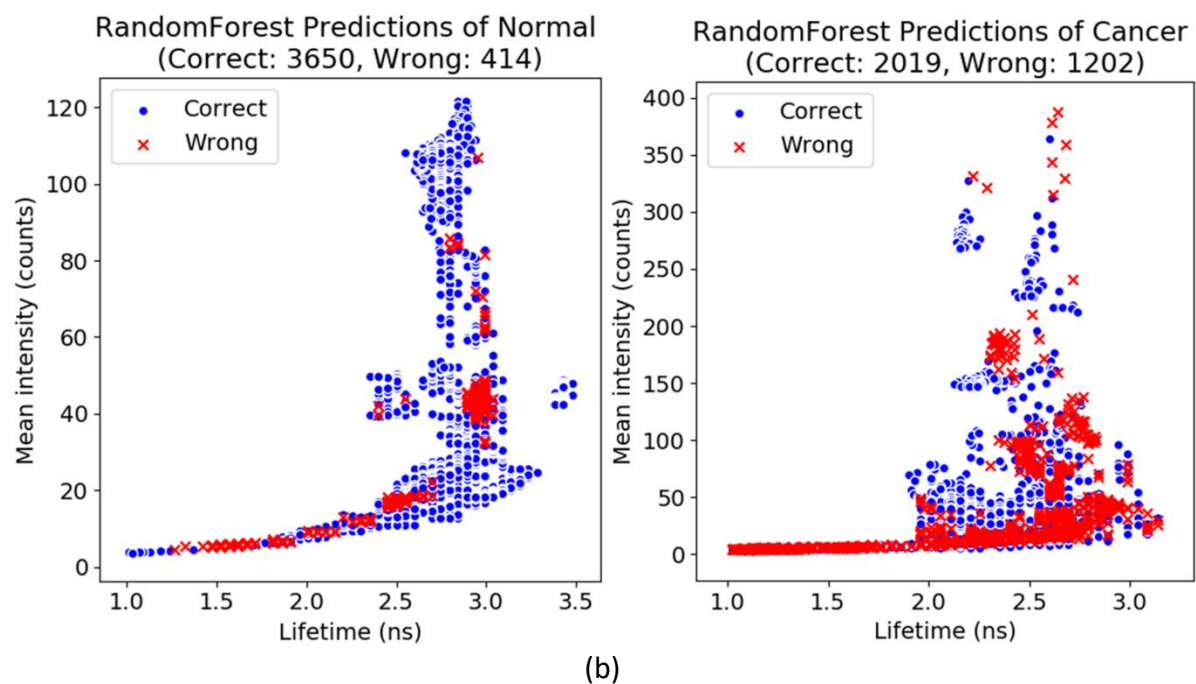


Figure 2: Confusion matrix of the predictions of the testing datasets by the MLs



(a)



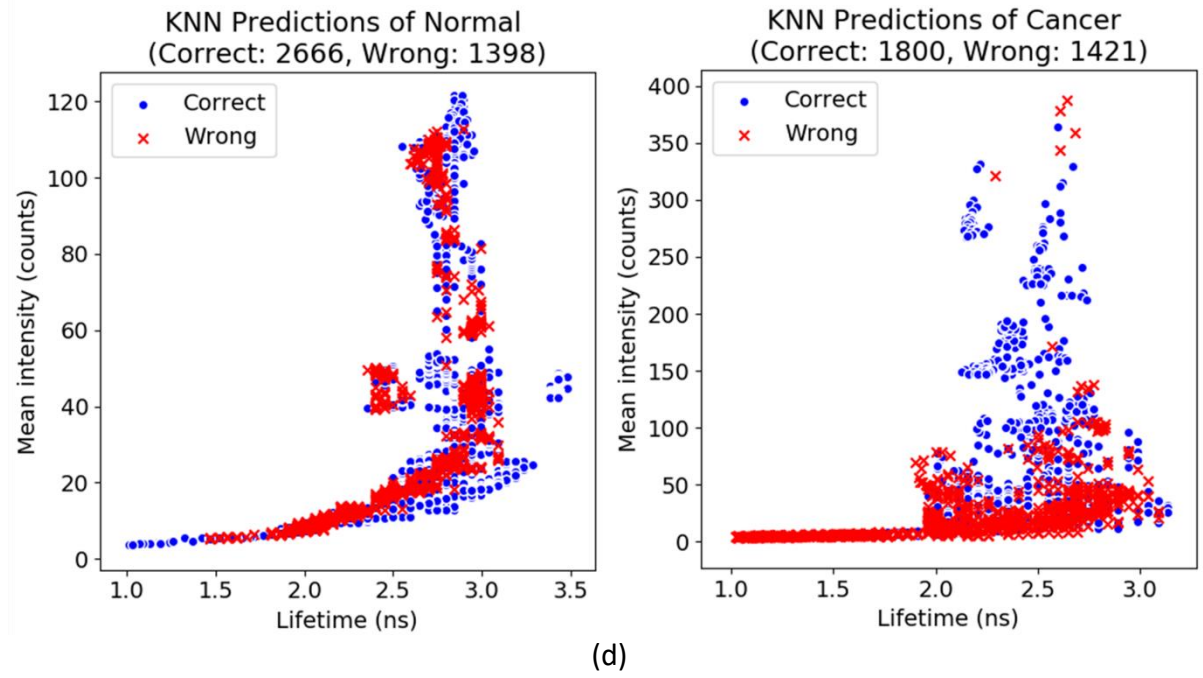


Figure 3: Cancerous and normal tissue predictions by SVC (a), RF (b), NN (c), and KNN (d)

Figure 1 presents the ROC curves and the corresponding AUC scores achieved by the ML approaches. Let TP, FP, TN and FN denote true positive, false positive, true negative, and false negative, respectively. Precision, and recall can be defined as:

$$\text{Precision} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Recall} = \frac{FP}{TP+FN} \quad (6)$$

Figure 2 illustrates the confusion matrices of the predictions on the testing dataset generated by the MLs, along with the calculated precision and recall. Similar to the AUC scores, KNN performs the worst in terms of deriving precision and recall. SVC and NN have comparable performance for precision and recall, since the incorrect predictions (TN and FP) are close to each other. As far as RF is concerned, it is good at predicting normal tissue with only 414 false predictions, and thus achieves the highest precision among the four ML techniques. However, RF struggles to predict cancerous tissue.

The predictions on the testing sets gathered by the MLs are illustrated in Figure 3, using averaged lifetime against mean intensity. Figure 3 depicts that the estimations generated by SVC and NN are very close for both cancer and normal tissues, whereas RF is more confident in estimating normal tissues rather than the cancerous, which is also reflected by the confusion matrix in Figure 2.

5. Conclusion

In this study, we reported the feasibility investigation of applying ML techniques to fluorescence to distinguish *ex-vivo* normal and cancerous human lung with lifetime images. Through the experimental results, we can conclude that the most complex ML algorithms are able to be utilised for FLIM-based lung cancer classification. In addition, pixel lifetime in FLIM

images, rather than professional engineered features based on the images, can be applied for the classification.

Despite the encouraging results, there are still some aspects need to be addressed in the future. The input for the MLs was flattened from 2-dimensional lifetime to 1-dimensional vector, and thus the correlations among adjacent pixels were lost. Accordingly, the classification at the image level, e.g. using deep convolutional neural networks [13], could be a promising direction to further improve the prediction accuracy. In addition, enhancing the quality of the images with higher resolution and more spectral information would be helpful to make the predictions more robust.

Ethics statement

All experiments using ex-vivo human lung tissue were performed following approval by the appropriate regional Research Ethics Committee (REC), NHS Lothian (references 13/ES/0126 and 16/LO/1883), and all subjects gave written informed consent.

Acknowledgements

We would like to thank the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) Interdisciplinary Research Collaboration (grant number EP/K03197X/1 and EP/R005257/1) for funding this work.

We acknowledge Dr Neil Finlayson, Dr Gareth O. S. Williams, and Dr Elvira Williams who designed and developed the hardware and software of the FLIM system. We also thanks Mrs Susan Fernandes for preparing the sample tissue and collecting the data, and Dr Ahsan Akram and Professor Kevin Dhaliwal for providing clinical-relevant information. We also appreciate Dr Catharine Ann Dhaliwal for annotating the histological images.

We sincerely thank the relatives of the lung donors and all those involved in enabling the use of human organs in research, made possible through charitable funding from Warren's Wish.

References

1. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med. Image Anal.* Elsevier B.V.; 2017; 42: 60–88.
2. Marcu L. Fluorescence lifetime techniques in medical applications. *Ann. Biomed. Eng.* 2012; 40: 304–331.
3. Berezin MY, Achilefu S. Fluorescence lifetime measurements and biological imaging. *Chem. Rev.* 2010; 110: 2641–2684.
4. Perperidis A, Dhaliwal K, McLaughlin S, Vercauteren T. Image computing for fibre-bundle endomicroscopy: A review. 2018.
5. Lin W, Yuan X, Yuen P, Wei WI, Sham J, Shi P, Qu J. Classification of in vivo autofluorescence spectra using support vector machines. *J. Biomed. Opt.* 2004; 9: 180.
6. Majumder SK, Ghosh N, Gupta PK. Support vector machine for optical diagnosis of

- cancer. *J. Biomed. Opt.* 2005; 10: 024034.
7. Majumder SK, Ghosh N, Gupta PK. Relevance vector machine for optical diagnosis of cancer. *Lasers Surg. Med.* 2005; 36: 323–333.
 8. Chen B, Lu Y, Pan W, Xiong J, Yang Z, Yan W, Liu L, Qu J. Support Vector Machine Classification of Nonmelanoma Skin Lesions Based on Fluorescence Lifetime Imaging Microscopy. *Anal. Chem.* 2019; : acs.analchem.9b01866
 9. Bradski G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* 2000; .
 10. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.
 11. Kevin P Murphy. Machine learning: a probabilistic perspective. Cambridge: MIT Press; 2012.
 12. Boser, Bernhard E. and Guyon, Isabelle M. and Vapnik VN. A training algorithm for optimal margin classifiers. *Proc. Fifth Annu. Work. Comput. Learn. Theory* Pittsburgh: ACM; 1992. p. 144–152.
 13. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521: 436–444.
 14. Breiman L. Random forests. *Mach. Learn.* 2001; : 5–32.
 15. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot Éd. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 2011; 12: 2825–2830.