Deep Learning in ex-vivo Lung Cancer Discrimination using Fluorescence Lifetime Endomicroscopic Images

Deep Learning in *ex-vivo* Lung Cancer Discrimination using Fluorescence Lifetime Endomicroscopic Images

Qiang Wang[1], James R. Hopgood[2], Neil Finlayson[3], Gareth O. S. Williams[1], Susan Fernandes[1], Elvira Williams[1],Ahsan Akram[1], Kevin Dhaliwal[1], Marta Vallejo[4]

***Abstract*— Fluorescence lifetime is effective in discriminating cancerous tissue from normal tissue, but conventional discrimination methods are primarily based on statistical approaches in collaboration with *prior* knowledge. This paper investigates the application of deep convolutional neural networks (CNNs) for automatic differentiation of *ex-vivo* human lung cancer via fluorescence lifetime imaging. Around 70,000 fluorescence images from *ex-vivo* lung tissue of 14 patients were collected by a custom fibre-based fluorescence lifetime imaging endomicroscope. Five state-of-the-art CNN models, namely ResNet, ResNeXt, Inception, Xception, and DenseNet, were trained and tested to derive quantitative results using accuracy, precision, recall, and the area under receiver operating characteristic curve (AUC) as the metrics. The CNNs were firstly evaluated on lifetime images. Since fluorescence lifetime is independent of intensity, further experiments were conducted by stacking intensity and lifetime images together as the input to the CNNs. As the original CNNs were implemented for RGB images, two strategies were applied. One was retaining the CNNs by putting intensity and lifetime images in two different channels and leaving the remaining channel blank. The other was adapting the CNNs for two-channel input. Quantitative results demonstrate that the selected CNNs are considerably superior to conventional machine learning algorithms. Combining intensity and lifetime images introduces noticeable performance gain compared with using lifetime images alone. In addition, the CNNs with intensity-lifetime RGB image is comparable to the modified two-channel CNNs with intensity-lifetime two-channel input for accuracy and AUC, but significantly better for precision and recall.**

I. INTRODUCTION

Fluorescence microscopy techniques have broad applications for the analysis of tissue specimens since they are extremely sensitive and able to deliver rich information about biochemical interactions at the molecular level. Along with this intensity and spectral information, a unique characteristic of fluorescence is its lifetime, i.e. the average time taken for the decay from an excited state to the ground state. The decay varies with the biological environment, e.g. pH, but is usually independent of fluorophore concentration [1]. As a result, fluorescence lifetime images are visually more homogeneous than intensity images. Visual interpretation of fluorescence lifetime endomicroscopic (FLIM) images is extremely challenging even for experienced experts. Conventional methods mainly fall into the application of statistical technologies along with auxiliary information. A typical scenario uses averaged lifetime by histogramming lifetime images to distinguish normal/cancerous tissue based on the corresponding histological images. Due to the independence of lifetime from intensity, imaging a sample can result in images different in intensity but similar in lifetime. In addition, lifetime of healthy tissue can be either shorter or longer than unhealthy tissue. For example, Jo et al. [2] reported that the lifetime for oral cancer is shorter than for non-cancerous tissue. In contrast, McGinty et al. found that tumours have a longer lifetime [3]. Furthermore, hardware configurations and reconstruction methods, such as excitation wavelength and reconstruction algorithms, also introduce differences in derived lifetime.

Machine learning (ML), particularly deep learning (DL), has revolutionized our world in many aspects. However, few papers have addressed the application of ML to FLIM images for automatic differentiation of normal/cancerous tissue. The primary reason may be the shortage of real data, the challenge of accurate interpretation of the data, and the requirement of prior knowledge. Nevertheless, Gu *et al.* [4] utilized a traditional feedforward neural network (NN) classifier and extreme learning for training to detect cervical neoplasia using FLIM on H&E-stained samples. Chen *et al.* [5] tried an indirect strategy for skin lesion detection. In their approach, artificial features extracted from lifetime reconstruction results were utilized for automatic detection by a Support-Vector Machine (SVM). Jo *et al.* [2] used quadratic discriminant analysis to classify malignant and benign lesions for oral cancer with six FLIM-based features. The authors have applied four different ML technologies, namely, K-nearest neighbour (KNN), SVM, random forest (RF), and neural network (NN) to FLIM images [6]. Unlike other approaches that use expert-engineered features, pixel values of lifetime images were used as the input into the four ML algorithms. Surprisingly, little research has been conducted on the applications of DL to automatic differentiation of human cancer with FLIM.

This study aims to investigate the feasibility of applying CNNs for automatic classification of *ex-vivo* human lung cancer with fluorescence lifetime images, with about 70,000 FLIM images collected from normal and cancerous tissue

This work is supported by the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) Interdisciplinary Research Collaboration (grant number EP/K03197X/1 and EP/R005257/1)

[1]Q. Wang, G. O. S. Williams, S. Fernandes, E. Williams, A. Akram, and K. Dhaliwal are with the Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK

[2]J. R. Hopgood is with the Institute for Digital Communications, School of Engineering, University of Edinburgh, Edinburgh, UK

[3]N. Finlayson is with the Institute for Integrated Micro and Nano Systems, School of Engineering, University of Edinburgh, Edinburgh, UK

[4]M. Vallejo (corresponding author) is with the School of Engineering and Physical Science, Heriot-Watt University, Edinburgh, UK (email: m.vallejo@hw.ac.uk)

of 14 patients by a custom fibre-based FLIM system [7] [1]. For feasibility investigation, five well-recognized CNN models, namely ResNet, ResNeXt, Inception, Xception, and DenseNet, are chosen and compared with conventional ML approaches.

II. Methodology

A. Data collection

The custom FLIM imaging system recorded sequences of lifetime images with a resolution of 128×128 pixels, at a frame rate of 9 frames a second, from two user-specified spectral bands, which were aggregated across a line sensor of single photon detectors. Each frame, therefore, contains four images, yielding an intensity and corresponding lifetime images for each of the two spectral bands. Lifetime values were calculated using the rapid lifetime determination method (RLD) [8].

For each *ex-vivo* tissue experiment, two freshly excised unfixed lung tissue samples from each patient obtained from cardiothoracic surgery: one normal and one cancerous. For each piece of tissue, multiple measurements were conducted at different physical points by direct contact of the fibre with the tissue. Two different pixel exposure times of 6 and 20 μs were used to either enable high frame rates (6 μs) or increased accuracy in lifetime determination (20 μs). It is worth noting that once the configurations are fixed, the regions of interest (ROIs) of the images were the same. Around 70,000 fluorescence lifetime images were collected from 14 pairs of lung normal/cancerous tissues. Fig. 1 depicts an example of the experiment workflow, where normal and cancerous lung tissues were fixed on a cork board, and the resultant 128x128 images of autofluorescence intensity and lifetime were generated with an exposure time 6 μs, a spectral band of 498-570 nm, and RLD approach. These settings were chosen representing intended characteristic conditions for future clinical trials. Histology images for the tissues were also presented for comparison purposes.

B. Image pre-processing

Unlike the fluorescence intensity images, fluorescence lifetime images are more homogeneous since they are independent of fluorophore concentration [1]. To derive plausible lifetime from intensity using RLD method, optimal signal-noise ratio (SNR) of the intensity measurement needs to be obtained. In this study, we used a threshold value $\sqrt{\hat{I}}$ related to SNR, where $\hat{I}$ is the mean of the measured intensity. A fluorescence intensity greater than the SNR is required to perform a lifetime calculation of acceptable accuracy.

Let $I^I = \{i^I_{x,y} | i^I_{x,y} \geq 0 \text{ and } x, y \in [0, M]\}$ denote a intensity image I^I with size of $M \times M$, and $I^L = \{i^L_{x,y} | i^L_{x,y} \geq 0 \text{ and } x, y \in [0, M]\}$ denote the corresponding lifetime image I^L with size of $M \times M$. A simple filter can be defined as:

$$\hat{i}^I_{x,y} = \begin{cases} 0 & i^I_{x,y} \leq \sqrt{\hat{I}} \\ i^I_{x,y} & \text{otherwise} \end{cases} \quad (1)$$

$$\hat{i}^L_{x,y} = \begin{cases} 0 & i^I_{x,y} \leq \sqrt{\hat{I}} \\ i^L_{x,y} & \text{otherwise} \end{cases} \quad (2)$$

Afterwards, the intensity images are normalized with dark background D and lightfield images L. The normalization is adapted from [9] as:

$$\bar{I}^I = \frac{G^*(\hat{I}^I - D)}{G^*(L - D)} \quad (3)$$

where $\bar{I}^I$ is the intensity image and G^* is a convolutional Gaussian smoothing filter with a 3×3 kernel defined in [10]. Later, a histogram-based contrast enhancement approach [10] is applied to the normalised images to further improve the quality of the images. Eventually, the normalized intensity image is used as a mask on the thresholded lifetime image to yield the pre-processed intensity and corresponding lifetime images.

C. Convolutional neural networks

Contemporary architectures, like the Inception [11], ResNet and its variants [12], [13], Xception [14], and DenseNet [15], have advanced the state-of-the-art classification performance significantly. In this study, we use ResNet50, ResNeXt50, Inception V3, Xception, and DenseNet121, implemented by *Keras*.

Lifetime images, shown in the middle of Fig. 2, are the input of the models. In addition, we also stack intensity and its corresponding lifetime images to reflect the independence of fluorescence lifetime from its intensity and make full use of the available images. Since the original implementation of the selected models requires RGB images as input, intensity and lifetime images are put into different channels of an RGB image, where the intensity and lifetime images are normalized to [0.0, 1.0]. Fig. 2 demonstrates an example of the stacking. Furthermore, we adapt the CNNs to be suitable for two-channel input, in order to investigate the impact of keeping a blank channel in the input images on the classification.

D. Evaluation Metrics

In order to quantify the performance of the CNNs on the inputs of lifetime images, two- and three-channel lifetime and intensity images, we use three metrics for the evaluation, including accuracy, precision, and recall. Let TP, FP, TN and FN denote true positive, false positive, true negative, and false negative, respectively. Accuracy, precision, and recall can be defined as:

$$\text{Accuracy} = \frac{\text{TP + TN}}{\text{TP+FP+TN+FN}} \quad (4)$$

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

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

[1]The datasets will be available on https://proteus.ac.uk/technology/signal-processing/.

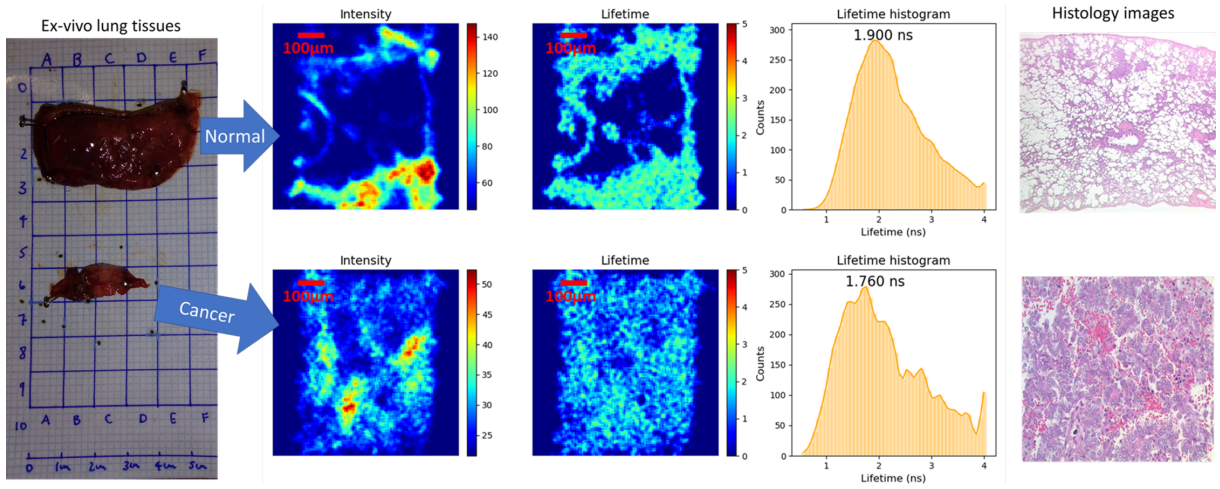

Fig. 1. Typical prepared sample with a normal tissue sample (top of the first column) and a cancerous sample (bottom of the first column). An example of fluorescence intensity (second column) and lifetime (third column) images collected by the system on a normal and cancerous tissue (first column), along with the lifetime histogram (fourth column) and the histology images (fifth column).

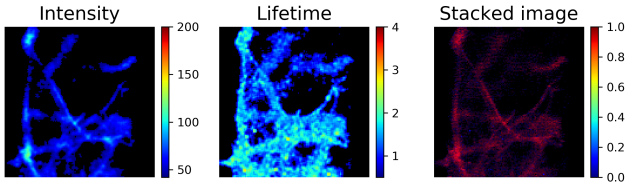

Fig. 2. Intensity (left), lifetime (middle), and the stacked (right) images.

In addition, we also use the area under receiver operating characteristic (ROC) curve (AUC) as the fourth metric. A ROC curve uses two parameters, true positive rate (TPR) and false positive rate (FPR), to demonstrate the performance of a classification model at different classification thresholds. TPR is calculated by (6) as recall, and FPR is defined as:

$$\text{FPR} = \frac{\text{FP}}{\text{TP+FN}} \tag{7}$$

AUC measures the area underneath the ROC curve.

III. RESULTS

We used one-out-of-all testing on three independent datasets from three patients. That is, 13 patients' images were utilized for training and validation and the remaining dataset from one patient for testing. By repeating the procedure three times and each time on a different patient. 10% of training data from 13 patients were split out for validation. All models were trained using Adam optimizer [16] with the parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate was initialized to 0.001, and divided by 10 at epoch 50 and 75. All models were trained and validated with 100 epochs of batch size 32. We use cross-entropy as the loss function. Weights are initialized using He's method [17]. In addition, we also employ weight decay equal to 1e-4. For data augmentation, we utilize the same strategy reported in [12], [15]. All training and testing were performed using NVidia V100 GPU provided by EPCC Cirrus[2] and JADE[3].

TABLE I

PERFORMANCE COMPARISON. THE BEST SCORES FOR EACH CATEGORY ARE IN BLUE, AND THE OVERALL BEST SCORES ARE IN RED.

Conventional ML approaches [6]				
	Accuracy	Precision	Recall	AUC
KNN	0.613	0.563	0.559	0.608
SVC	0.767	0.734	0.742	0.765
NN	0.763	0.722	0.752	0.761
RF	0.778	0.83	0.627	0.763
CNNs on lifetime images				
	Accuracy	Precision	Recall	AUC
ResNet50	0.828	0.83	0.893	0.813
ResNeXt50	0.785	0.824	0.812	0.779
DenseNet121	0.795	0.831	0.823	0.788
Inception	0.794	0.838	0.812	0.79
Xception	0.769	0.854	0.739	0.777
CNNs on intensity+lifetime 3-channel input				
	Accuracy	Precision	Recall	AUC
ResNet50	0.816	0.804	0.943	0.767
ResNeXt50	0.839	0.839	0.902	0.824
DenseNet121	0.865	0.88	0.895	0.858
Inception	0.852	0.876	0.874	0.846
Xception	0.817	0.895	0.785	0.825
Modified CNNs on intensity+lifetime 2-channel input				
	Accuracy	Precision	Recall	AUC
ResNet50	0.822	0.761	0.856	0.826
ResNeXt50	0.848	0.816	0.806	0.841
DenseNet121	0.821	0.774	0.787	0.815
Inception	0.855	0.813	0.832	0.851
Xception	0.843	0.798	0.818	0.839

Table I lists the quantitative results, including accuracy, precision, recall, and AUC, from both four conventional ML algorithms from [6], i.e. K-nearest neighbour, support-vector classifier, neural network, and random forest, and the selected five contemporary CNNs, including ResNet50, ResNeXt50, Inception V3, Xception, and DenseNet121.

[2]http://www.cirrus.ac.uk
[3]http://www.jade.ac.uk

Overall, the selected CNNs surpassed the ML approaches for almost every metric. The results reflect that considering two-dimensional spatial information in lifetime images by convolutions in CNNs achieves better performance than one-dimensional pixel values in the ML algorithms. It is worth noting that the amount of data used in [6] is less than those in this study, which might introduce some uncertainty of the classification results.

As far as the CNNs are concerned, different input formats also influences the resultant quantities. In general, using lifetime images only as input is inferior to combining intensity and lifetime images together. This is primarily due to the introduction of an extra dimension, i.e. the intensity information. However, the precision and recall scores given by lifetime only are better than two-channel input, but they are still not as good as three-channel input. When it comes to the combination of intensity and lifetime images, the CNNs with the three-channel input are comparable to the modified CNNs with the two-channel input in accuracy and AUC. In contrast, the precision and recall scores by the former are considerably better than the latter. This might be due to that the CNNs were designed for three-channel RGB images, whereas the modified ones were not fully optimized for two-channel input. It might also be because that the third zero-value channel may be marginally beneficial by adding extra sparsity in the network which could be more suitable for our data.

IV. CONCLUSIONS

In this work, we investigated the feasibility of applying deep convolutional neural netwoks to fluorescence lifetime images for *ex-vivo* human lung cancer classification. Through the experimental results, we addressed three critical questions which have not been answered previously for DL in FLIM-based human lung cancer classification:

- The selected CNNs outperform the ML algorithms on FLIM images;
- The CNNs with lifetime images only are inferior to those with the combination of intensity and lifetime images, although the former are slightly better than the latter with two-channel intensity and lifetime input in precision and recall; and
- The CNNs with three-channel intensity and lifetime images are superior to those with two-channel input, even though they are comparable in accuracy and AUC.

Despite the encouraging results, there are several aspects worth of further investigation. First is to enhance the quality of the input images so that the classification outcomes could be improved. Since the reconstructed FLIM images are rich in multi-dimensional information, such as excitation wavelength, how to integrate multi-dimensional information into the CNNs is another challenge. Finally, due to the unique characteristics of fluorescence lifetime and its images, further research should be conducted on the development of special architectures suitable for such images, although general-purpose CNNs have achieved remarkable outcomes.

ACKNOWLEDGEMENTS

We appreciate Dr Catharine Ann Dhaliwal for annotating the histological images. This work used the Cirrus UK National Tier-2 HPC Service at EPCC funded by the University of Edinburgh and EPSRC (EP/P020267/1). This project also made use of time on Tier 2 HPC facility JADE, funded by EPSRC (EP/P020275/1).

REFERENCES

[1] Laura Marcu, "Fluorescence lifetime techniques in medical applications," *Annals of biomedical engineering*, vol. 40, no. 2, pp. 304–331, 2012.

[2] J. A. Jo, S. Cheng, R. Cuenca-Martinez, E. Duran-Sierra, B. Malik, B. Ahmed, K. Maitland, Y. L. Cheng, J. Wright, and T. Reese, "Endogenous fluorescence lifetime imaging (flim) endoscopy for early detection of oral cancer and dysplasia*," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2018, pp. 3009–3012.

[3] James McGinty, Neil P Galletly, Chris Dunsby, Ian Munro, Daniel S Elson, Jose Requejo-Isidro, Patrizia Cohen, Raida Ahmad, Amanda Forsyth, Andrew V Thillainayagam, et al., "Wide-field fluorescence lifetime imaging of cancer," *Biomedical optics express*, vol. 1, no. 2, pp. 627–640, 2010.

[4] Jun Gu, Chit Yaw Fu, Beng Koon Ng, Sirajudeen Gulam Razul, and Soo Kim Lim, "Quantitative diagnosis of cervical neoplasia using fluorescence lifetime imaging on haematoxylin and eosin stained tissue sections," *Journal of Biophotonics*, vol. 7, no. 7, pp. 483–491, 2014.

[5] Bingling Chen, Yuan Lu, Wenhui Pan, Jia Xiong, Zhigang Yang, Wei Yan, Liwei Liu, and Junle Qu, "Support Vector Machine Classification of Nonmelanoma Skin Lesions Based on Fluorescence Lifetime Imaging Microscopy," *Analytical Chemistry*, vol. 91, no. 20, pp. 10640–10647, 2019.

[6] Qiang Wang, Marta Vallejo, and James Hopgood, "Fluorescence Lifetime Endomicroscopic Image-based ex-vivo Human Lung Cancer Differentiation Using Machine Learning," *TechRxiv Preprint*, Jan. 2020.

[7] Ahmet T Erdogan, Richard Walker, Neil Finlayson, Nikola Krstajić, Gareth Williams, John Girkin, and Robert Henderson, "A cmos spad line sensor with per-pixel histogramming tdc for time-resolved multispectral imaging," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 6, pp. 1705–1719, 2019.

[8] Richard M Ballew and JN Demas, "An error analysis of the rapid lifetime determination method for the evaluation of single exponential decays," *Analytical Chemistry*, vol. 61, no. 1, pp. 30–33, 1989.

[9] Tim N. Ford, Daryl Lim, and Jerome Mertz, "Fast optically sectioned fluorescence HiLo endomicroscopy," *Journal of Biomedical Optics*, vol. 17, no. 2, pp. 021105, 2012.

[10] Milan Sonka, Vaclav Hlavac, and Roger Boyle, *Image processing, analysis, and machine vision*, Cengage Learning, 2014.

[11] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[13] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[14] François Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[15] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2261–2269.

[16] Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization," in *ICLR*, 2015.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.