

## Clasificación Automática de Imágenes m-FLIM por Redes Convolucionales

F. J. Sánchez-González<sup>1</sup>, M. O. Rodríguez-López<sup>1</sup>, D. U. Campos-Delgado<sup>1</sup> y Javier A. Jo<sup>2</sup>

<sup>1</sup>Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México

<sup>2</sup>Departamento de Ing. Biomédica, Texas A&M University, EUA

**Resumen**— En este trabajo se explora el uso de imágenes multi-espectrales microscópicas de tiempo de vida fluorescente (m-FLIM), pertenecientes a segmentos de arteria coronaria postmortem con arteriosclerosis, para detectar tres comportamientos en los componentes orgánicos: alto contenido de colágeno (HC), alto contenido de lípidos (HL) y bajo contenido de colágeno y lípidos (LCL). El objetivo de este trabajo es la clasificación automática de las muestras m-FLIM mediante una red neuronal convolucional cuyos parámetros se ajustan a través de la información proporcionada por las muestras puras de estos componentes bajo una filosofía supervisada. Enseguida se realiza un estudio comparativo con un algoritmo basado en separación lineal y optimización cuadrática para la clasificación de las muestras. Los resultados muestran que la red neuronal convolucional puede generar una clasificación precisa y con baja complejidad.

**Palabras clave**— tiempo de vida de fluorescencia, redes convolucionales, clasificación, extracción de rasgos

### I. INTRODUCCIÓN

Los avances recientes en la instrumentación electro-óptica, han contribuido a posicionar a las técnicas ópticas como herramientas viables para el diagnóstico temprano de diversas patologías, esto debido a que presentan numerosas ventajas, entre ellas una alta especificidad, buena resolución; sin mencionar que muchas de ellas se aplican de forma no invasiva, brindando mayor seguridad y comodidad al usuario. Sin embargo, es necesario seleccionar cuidadosamente la técnica a utilizar según el tipo de estudio y la naturaleza de lo que se pretende conocer, ya que de ello dependerán parámetros como la especificidad y el nivel de resolución obtenido. Actualmente existen diversas herramientas para abordar el diagnóstico médico mediante técnicas ópticas como la tomografía de coherencia óptica (OCT), la termografía, la espectroscopia de resonancia magnética, la espectroscopia por infrarrojo cercano, espectroscopia Raman, entre otras.

Un método que tiene el potencial de cuantificar los componentes químicos presentes en una muestra son las imágenes multi-espectrales microscópicas de tiempo de vida fluorescente (m-FLIM) [1]. A pesar de que mediante esta técnica es posible identificar los diferentes componentes de una muestra, resulta complejo en ciertos escenarios realizar

la clasificación de muestras que resultan de la combinación de varios componentes. Para lograr identificar la presencia de los componentes en las mezclas, se usan técnicas como la deconvolución para eliminar el efecto de la excitación laser y estimar el tiempo de vida promedio [2], sin embargo, este tipo de técnicas resultan complejas, sin mencionar que es necesario realizar una extracción de rasgos o características que sean discriminantes entre las clases para lograr una buena clasificación.

Una de las técnicas con avances recientes y prometedores en clasificación automática es aprendizaje profundo (*deep learning*), en particular las redes neuronales convolucionales (CNN), cuyo uso en señales se ha explorado recientemente [3],[4] con resultados altamente satisfactorios. Las CNN son un tipo de redes neuronales profundas con la particularidad de que, en lugar de utilizar capas ocultas totalmente conectadas introducen estructuras especiales en la red: las capas de convolución y de *pooling*, estas dos capas funcionan como un extractor de rasgos en el proceso de aprendizaje.

En este contexto, esta investigación estudia la clasificación automática de muestras m-FLIM a través de CNN. En este trabajo se utiliza una base de datos constituida por segmentos de arteria coronaria postmortem con arteriosclerosis [1], para identificar tres estados asociados con esta patología: alto contenido de colágeno (HC), alto contenido de lípidos (HL) y bajo contenido de colágeno y lípidos (LCL). Como punto de comparación se evalúa la clasificación con un algoritmo basado en separación lineal y optimización cuadrática [7], donde el enfoque basado en CNN permite una clasificación con alta precisión y complejidad reducida.

### II. METODOLOGÍA

#### A. Base de Datos

En este trabajo se utilizó la base de datos m-FLIM [1], que consta de 78 muestras de emisión de fluorescencia obtenidas a partir de segmentos de arteria coronaria de tejido postmortem con arteriosclerosis. La respuesta fluorescente se capturó en tres longitudes de onda con el fin de identificar tres compuestos orgánicos:  $390 \pm 20$  nm para el colágeno,  $452 \pm 22.5$  nm para la elastina, y  $550 \pm 20$  nm para los lípidos.

Cada muestra posee una dimensión de  $60 \times 60 \times 700$ , donde las primeras dos dimensiones definen el plano

espacial, y pueden ser consideradas como píxeles; mientras que la última dimensión, establece la respuesta temporal. En esta dimensión, el vector temporal de emisión de fluorescencia se genera a través de concatenar las mediciones de las tres bandas asociadas a cada uno de los componentes orgánicos mencionados anteriormente.

Las muestras fueron clasificadas histopatológicamente en tres grupos: HC, HL y LCL. Las primeras 26 muestras son de carácter puro, es decir solo uno de los tres componentes orgánicos se manifiesta mayoritariamente en la muestra de tejido; de esta forma se cuenta con cinco muestras de HC, cinco de HL y 16 de LCL. Las 52 muestras restantes poseen mezclas de los tres componentes.

### B. Diseño de la red convolucional

Para el diseño de la CNN se utilizó el software MATLAB®, y el toolbox de redes neuronales que permite crear, entrenar y simular este tipo de redes [5]. Al ser necesario realizar una clasificación pixel a pixel de las muestras en las que se encuentran mezclados los tres componentes (HC, HL, LCL), es necesario trabajar con las señales individuales de espectro de emisión de fluorescencia asociadas a cada pixel dentro de la imagen. De esta forma es posible generar regiones dentro de la imagen que se asocien con una clase específica. Al observar los datos desde esta nueva perspectiva, se logra la ventaja de tener más información para realizar el proceso de entrenamiento.

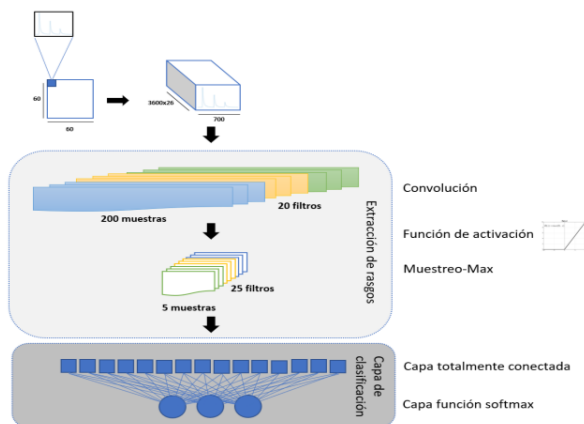


Figura 1. Arquitectura propuesta de la red convolucional.

El uso de redes convolucionales para la clasificación de señales unidimensionales aún se encuentra poco explorado, por esta misma razón las funciones diseñadas en MATLAB para trabajar con esta técnica asumen que se trabajará con imágenes. Con este fin se optó por formar una estructura de datos de tres dimensiones: (i) ancho del campo de vista, (ii) largo del campo de vista, y (iv) número de muestras temporales; donde el campo de vista define las dimensiones espaciales de la muestra.

La arquitectura de la CNN para la clasificación de mezclas fue establecida después de realizar una validación cruzada de cinco iteraciones (ver Figura 1) [6], donde se midió el desempeño de tres arquitecturas de CNN diferentes (ver Tabla 1). En esta validación se entrenó la CNN con las muestras puras, utilizando el 80% de los datos como conjunto de entrenamiento, y el 20% restante como conjunto de prueba, alternando la información de entrenamiento y prueba de manera aleatoria, con la finalidad de tener un resultado independiente de los datos. Estas pruebas de validación también fueron utilizadas para afinar dos hiperparámetros de la CNN, la tasa de aprendizaje (LR) y el número de épocas de validación. La LR especifica el peso del gradiente en el proceso de actualización, influyendo en el valor de los pesos de la CNN; el número de épocas hace alusión al número de veces que en el proceso de entrenamiento se ha cubierto la totalidad de los datos marcando el fin de un época, y el comienzo de otra. Una vez entrenada la CNN en la etapa de validación, se prosiguió a clasificar 27 de las muestras m-FLIM de mezclas, considerando cada pixel y asignando un valor 1 para HC, 2 para HL y 3 para LCL

### C. Estimación del error de clasificación de mezclas

Ya que una de las principales motivaciones del trabajo es lograr una clasificación de las imágenes m-FLIM de mezclas, fue necesario obtener un punto de comparación. Con este fin se utilizó el algoritmo *Blind End-member and Abundance Extraction* (BEAE) para identificar los perfiles de fluorescencia y sus concentraciones [7]. El BEAE es un algoritmo basado en estimación cuadrática con restricciones y en un modelo de descomposición lineal para cada pixel [7], [8]. El modelo de observación está dado por:

$$\mathbf{y}_k = \sum_{n=1}^N \alpha_{k,n} \mathbf{p}_n + \mathbf{v}_k \quad \forall k = 1, \dots, K,$$

donde  $\mathbf{y}_k$  representa el vector de mediciones fluorescentes en el  $k$ -ésimo pixel,  $\mathbf{p}_n$  el perfil fluorescente del  $n$ -ésimo componente elemental,  $\alpha_{k,n}$  la abundancia en el  $k$ -ésimo pixel del  $n$ -ésimo perfil fluorescente ( $0 \leq \alpha_{k,n} \leq 1$ ),  $\mathbf{v}_k$  el vector de ruido o incertidumbre en el  $k$ -ésimo pixel,  $K$  el número total de píxeles en la muestra, y  $N$  el número máximo de perfiles fluorescentes elementales por cada pixel.

El algoritmo retorna  $N$  mapas de abundancia conforme al número de perfiles fluorescentes. Con el objetivo de guiar a BEAE en la identificación de los perfiles fluorescentes elementales, la muestra de la mezcla a procesar fue modificada añadiendo una imagen de cada clase de muestras puras (HC, HL, LCL). De manera, que BEAE procesa una muestra extendida con dimensión  $240 \times 60 \times 700$ ,

extrayendo enseguida solamente los mapas de abundancia asociados con los pixeles de la mezcla.

Debido a que el algoritmo BEAE estima las abundancias de forma porcentual, con el fin de realizar una clasificación dura de la muestra, se empleó un umbral mayor al 50% en la abundancia para asignar al pixel a cada componente (HC,HL,LCL). Estas imágenes clasificadas poseen enteros del 1 al 4, donde los primeros 3 valores corresponden a la presencia de HC, HL y LCL respectivamente, mientras que un valor de 4 denota una incertidumbre en dicho pixel, debido a que el valor de abundancia no supero el umbral seleccionado en ninguno de los tres perfiles fluorescentes elementales.

Por otro lado, para medir el desempeño de la CNN en la clasificación, se prosiguió a calcular el error para la  $j$ -ésima muestra de mezcla, el cual fue calculado como la diferencia entre la imagen generadas mediante CNN  $A_j^{CNN}$  y la imagen discretizadas mediante el BEAE  $A_j^{BEAE}$ :

$$e_j = \frac{\sum_{i \in I} \mathbf{1}(A_j^{CNN}[i] - A_j^{BEAE}[i])}{\text{card}(I)}$$

donde las imágenes  $(A_j^{CNN}, A_j^{BEAE})$  se consideran vectorizadas,  $I$  denota el conjunto de pixeles sin incertidumbre después de la clasificación por BEAE,  $\mathbf{1}(\cdot)$  denota una función indicador ( $\mathbf{1}(x)=0$  si  $x=0$ , y  $\mathbf{1}(x)=1$  si  $x \neq 0$ ), y  $\text{card}(\cdot)$  la cardinalidad del conjunto. Este cálculo del error se realizó para las 27 muestras m-FLIM de mezclas, y enseguida se estima un valor de error promedio.

### III. RESULTADOS

Teniendo en cuenta el objetivo de clasificar las muestras m-FLIM, es necesario que la arquitectura de la CNN a través de su proceso de entrenamiento presente el mejor compromiso entre precisión y generalidad (ver Figura 1). Es por esto que se realizó una validación cruzada de 5 iteraciones, obteniendo los siguientes resultados para las 3 arquitecturas de red (CNN 1, CNN 2, CNN 3) planteadas según se muestra en la Tabla 1.

La CNN que presentó una mayor precisión promedio fue CNN 1, en 3 ocasiones, independientemente del valor que poseían los parámetros de LR y número de épocas de entrenamiento. En esta red, la primera capa de convolución se compone de 20 filtros con tamaño de ventana de 200 muestras, cada uno de estos filtros se desplaza una muestra a la vez y se encarga de captar los rasgos de las señales de m-FLIM. Le sigue una capa de *muestreo-max* (*max-pooling*) que posee una ventana de 5 elementos y se desplaza cada 5

muestras, esta capa disminuye la dimensión de los datos, en este caso de la capa de convolución que le precede.

	Precisión Promedio			
	10 Épocas y y LR de 0.001	10 Épocas y y LR de 0.0001	15 Épocas y y LR de 0.001	15 Épocas y y LR de 0.0001
1 capa de convolucion y 1 capa de muestreo	99.36% ± 0.23%	93.91% ± 1.23%	98.58% ± 0.45%	98.59% ± 0.46%
2 capas de convolucion y 1 capa de muestreo	99.01% ± 0.25%	97.29% ± 0.36%	97.83% ± 0.36%	97.95% ± 0.32%
3 capas de convolucion y 1 capa de muestreo	98.47% ± 0.5328%	94.02% ± 1.47%	96.39% ± 0.86%	96.28% ± 0.73%

Tabla 1. Precisión promedio y desviación estándar de las tres arquitecturas CNN 1, CNN 2 y CNN 3 (primero, segundo y tercer renglón, respectivamente) para distintos valores de LR y épocas.

El desempeño de la CNN 1 se puede observar en la matriz de confusión que se presenta en la Tabla 2, obtenida a partir de la validación cruzada de 5 iteraciones. Con esta CNN, en la validación cruzada se obtuvo una correcta clasificación de 92,674 muestras de las 93,600 muestras puras que se tenían como base de datos de entrenamiento, resultando en una precisión promedio de 99.36%. En la Figura 2 se puede apreciar una imagen resultante de la clasificación por el algoritmo BEAE y a su derecha la imagen obtenida mediante la red convolucional.

	Clases Reales			Total Clasificación
	LCL	HC	HL	
LCL	39185	385	7	39577
HC	178	24689	119	24986
HL	61	176	28800	29037
Total Verdadero	39424	25250	28926	93600

Tabla 2. Matriz de confusión para los resultados de clasificación con la CNN 1.

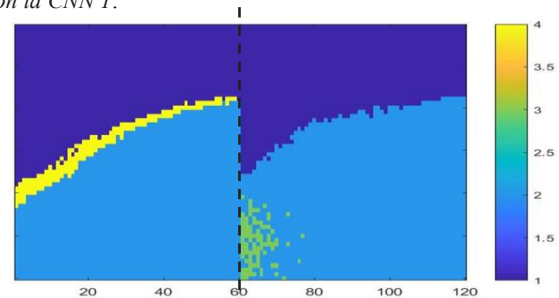


Figura 2. Clasificación de imágenes de fluorescencia algoritmo BEAE (izquierda) & CNN 1 (derecha): HC → 1, HL → 2, LCL → 3 e Incertidumbre → 4.

Con la finalidad de tener una representación gráfica del error a través de las muestras m-FLIM de mezclas, se generó un histograma del error. En la Figura 3, se puede apreciar

como el error máximo obtenido por la clasificación de CNN fue de alrededor de 20%, obtenido en la muestra 15, mientras que la gran mayoría de las restantes poseen un valor <10%. En general, el error promedio para las 27 muestras fue de 5.33%.

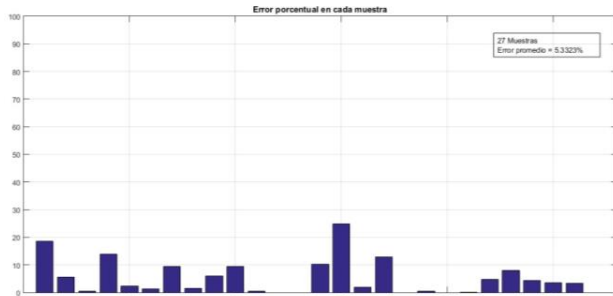


Figura 3. Histograma del error en los datos m-FLIM de mezclas.

#### IV. DISCUSIÓN

La ventaja principal que una CNN nos ofrece ante el algoritmo BEAE es la extracción automática de rasgos, además, de que al ya poseer una red óptima entrenada, el tiempo de procesamiento y la carga computacional se ven reducidas de manera drástica, ya que en este caso la clasificación se haría de manera automática. Las desventajas que posee el empleo de una CNN es el gran número de parámetros que deben ser modificados y planteados, lo que hace que las etapas iniciales de diseño se conviertan en un proceso holístico, donde los valores de los hiper-parámetros son fijados a prueba y error. Sin embargo, técnicas como la validación cruzada han reducido esta problemática a cambio de una carga computacional alta.

La CNN que fue seleccionada por su rendimiento y sintonización de parámetros a través de la validación cruzada (ver Tabla 1), resultó ser la red convolucional con la estructura menos compleja, donde esto se apega a los trabajos anteriormente citados [3],[4]. Esto se puede deber a que, si se poseen demasiadas capas para entrenar, el modelo tiende a sobre ajustarse muy fácilmente debido a la gran cantidad de pesos que se tienen que sintonizar; esto se trató de mitigar en este trabajo haciendo uso de una capa de *dropout* para las CNN 2 y 3. Sin embargo, en los resultados obtenidos, la precisión fueron menor a la esperada.

Los resultados obtenidos de la validación cruzada tienen el comportamiento esperado de las redes (ver Tabla 1). Por ejemplo, la CNN 3 tuvo el rendimiento más bajo cuando solo se entrenó por 10 épocas y con una tasa de aprendizaje de 0.001, posiblemente debido a que es una red que posee un mayor número de parámetros por ajustar, por lo cual es necesario que esta sea iterada más veces o en su defecto, que la tasa con la que los pesos son modificados sea mayor. Así

en la Tabla 1, se puede ver que su rendimiento mejora, cuando cualquiera de estas dos situaciones sucede.

Los resultados del error de la imagen clasificada por la CNN con respecto a BEAE, son bastantes satisfactorios, donde se puede apreciar que ambas metodologías de clasificación generan morfologías muy similares, y además, existe una correspondencia de las agrupaciones de componentes orgánicos; solo ocurriendo ligeras disimilitudes en las zonas de tránsito entre clases de la imagen, que es donde se encuentra concentrado el error.

#### V. CONCLUSIÓN

El desempeño que presentó la CNN fue satisfactorio para la clasificación de las muestras m-FLIM de segmentos de arteria coronaria de tejido postmortem con arteriosclerosis, donde se obtuvieron resultados similares con el algoritmo de separación lineal BEAE. Creemos que los resultados obtenidos en este trabajo tienen concordancia con los que se han tenido con BEAE [7], donde se ha concluido que una CNN con una estructura simple puede llegar a dar mejor desempeño para ciertos casos de estudio y con una menor complejidad. Como trabajo futuro, se contempla aplicar la clasificación con CNN a muestras m-FLIM de otros tipos de tejido, así como optimizar su implementación para una aplicación en tiempo real.

#### AGRADECIMIENTOS

Este trabajo de investigación fue financiado por CONACYT a través de una proyecto de Ciencia Básica (#254637).

#### BIBLIOGRAFÍA

- [1] Park, J., et al. (2012). Biochemical characterization of atherosclerotic plaques by endogenous multispectral fluorescence lifetime imaging microscopy. *Atherosclerosis*, 220(2), 394-401.
- [2] Jo, J. A., et al. (2006). Laguerre-based method for analysis of time-resolved fluorescence data: application to in-vivo characterization and diagnosis of atherosclerotic lesions. *Journal of biomedical optics*, 11(2), 021004.
- [3] Zeng, H., et al. (2016). Convolutional neural network architectures for predicting DNA-protein binding. *Bioinformatics*, 32(12), i121-i127.
- [4] Tsinalis, O., Matthews, P. M., Guo, Y., & Zafeiriou, S. (2016). Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. *arXiv preprint arXiv:1610.01683*.
- [5] Kim, P. (2017). *Deep Learning*. In *MATLAB Deep Learning* (pp. 121-147). Apress, Berkeley, CA.
- [6] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, No. 2, pp. 1137-1145).
- [7] Gutiérrez-Navarro, O., et al. (2014). Blind end-member and abundance extraction for multispectral fluorescence lifetime imaging microscopy data. *IEEE journal of biomedical and health informatics*, 18(2), 606-617.
- [8] Lakowicz, J.R. (2008). *Principles of fluorescence spectroscopy*, Springer Verlag.