

Clasificación de imágenes de redes mitocondriales asociadas al estudio del cáncer de mama

F.J. Vázquez-Vázquez, E. Castillo-Montiel, J.C. Chimal-Eguía,
P. Maycotte-Gonzáles

Instituto Politécnico Nacional, Centro de Investigación en Computación,
México

fvazquezv89@gmail.com, {ecastillom, jchmale}@ipn.mx, bisbenzimida@gmail.com

Resumen. Las mitocondrias son organelos dinámicos involucrados en diversos procesos celulares, sus funciones están relacionadas a su morfología. Las alteraciones en la dinámica mitocondrial se han asociado como biomarcadores a diversas enfermedades incluyendo el cáncer de mama. En este trabajo presentamos un método de clasificación para imágenes de redes mitocondriales extraídas de distintas líneas celulares (MCF10A, BT549, MDAMB23 y CMF) pertenecientes a distintos subtipos del cáncer de mama. El método se basa en tres etapas: en la primera se aplica un algoritmo de procesamiento morfológico y de segmentación, en la segunda se extraen características usando medidas empleadas comúnmente en redes complejas, finalmente se emplea el algoritmo no supervisado K-means para el proceso de clasificación. Los resultados indican una diferencia de las medidas obtenidas en cada subconjunto o clase, obteniendo una exactitud de 63.33% con 30 imágenes muestra.

Palabras clave: Morfología mitocondrial, grafos.

Mitochondrial Morphology Images Classification for Assessing Breast Cancer

Abstract. Mitochondria are involved in a variety of cellular functions, which are related to their morphology. The disturbance in mitochondria dynamics has been associated to biomarkers for assessing breast cancer. In this work, we present a method for classification of mitochondrial morphology images, the set of images belongs to cell lines: (MCF10A, BT549, MDAMB23 y CMF) which are subtypes of breast cancer. The proposed method is based on the following image processing algorithms: segmentation (Otsu's method), features extraction (based on graph measures), and clustering algorithm (K-means). The study reveals a remarkable difference between the graph measures applied in features extraction. We obtain an accuracy of 63.33

Keywords: Mitochondrial Morphology, Graph Measures.

1. Introducción

Dentro de diversos procesos celulares, un incremento en la fisión (división de un solo organelo en dos o mas independientes estructuras) causa fragmentación mitocondrial y un aumento en la fusión (reacción opuesta la fisión) produce una red de mitocondrias que contraresta procesos metabólicos[1]. Un balance entre fisión y fusión define una morfología micondrial cuyo propósito es cumplir con las demandas metabólicas y asegurar la remoción de organelos dañados. Estos eventos se han asociado a la proliferación y redistribución de mitocondrias, permitiendo el estudio de distintos subtipos de cáncer de mama[1-3]. El objetivo de este trabajo es clasificar imágenes de redes mitocondriales(mas de una mitocondria), con base en los cuatro tipos de morfologías mostradas en la figura 1 (1.-completamente tubular, 2.-tubular con fragmentos, 3.-fragmentada con túbulos, 4.-completamente fragmentada) que son fueron definidas para el estudio del cancer de mama[1], este clasificador permite a los expertos en el campo agilizar el proceso de clasificación normalmente aplicado de forma visual, además se pretende disminuir el error que se pudiera derivar de factores relacionados a parámetros de percepción o condición humana. El conjunto de imágenes fueron extraídas de las siguientes líneas celulares (MCF10A, BT549, MDAMB23 y CMF), resultado de los estudios aplicados [1].

2. Método

La figura 2 representa el diagrama a bloques de las etapas de procesamiento. En la primera etapa se aplican dos algoritmos(segmentación y adelgazamiento) con objetivo de mantener la morfología de la red de mitocondrias en la imagen y tratar a esta morfología como un grafo. En la segunda etapa se obtienen medidas de redes complejas de la morfología obtenida en la etapa previa. Finalmente se define el algoritmo no supervisado que emplea las medidas obtenidas para utilizar el método K-means.

2.1. Etapa 1

Las imágenes empleadas se encuentran representadas por los canales RGB (red, green, blue). El primer paso es obtener una imagen en escala de grises, para esto se calcula el promedio del valor de cada canal para una unidad de pixel, la ecuación 1 determina la función $f(x, y)$ que se obtiene al calcular el valor promedio de la intensidad de los canales $r(x, y)$, $g(x, y)$ y $b(x, y)$:

$$f(x, y) = \frac{r(x, y) + g(x, y) + b(x, y)}{3}. \quad (1)$$

2.1.1. Segmentación En esta sección el objetivo es aplicar un algoritmo que obtenga la región de interés de la imagen, es decir la red de mitcondrias. Este procedimiento es aplicado debido a que las imágenes contienen elementos(pixeles)

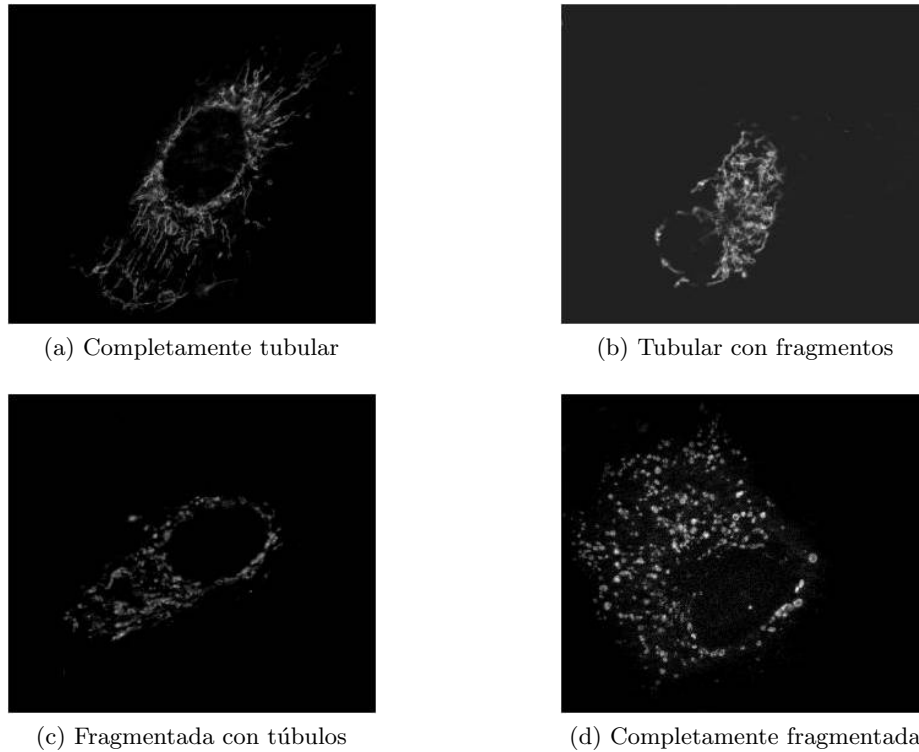


Fig. 1: Clases definidas para el proceso de clasificación.

que corresponden al núcleo de la célula, además representan parte de la reacción química con la cual se tiñe a la red de mitocondrias durante la captura de la imagen[1]. Esta segmentación se procesa sobre la función $f(x, y)$, para esto se emplea el método de segmentación Otsu [5].

El algoritmo Otsu consiste en obtener umbrales de segmentación dentro de un rango de intensidades 0-255, siendo el número de umbrales la cantidad de secciones entre las cuales se desea dividir el rango de escala de grises, para este trabajo se definen dos umbrales t_1 y t_2 que permiten extraer la geometría de la red de mitocondrias. Para encontrar los umbrales t_1 y t_2 se obtiene un histograma con las probabilidades de intensidad de cada nivel en la imagen, estos son separados obteniendo la probabilidad de tres clases con fronteras establecidas por t_1 y t_2 , con el objetivo de encontrar los valores t_1 y t_2 para los cuales la varianza de cada una de las tres clases es la máxima, estos umbrales definen la segmentación de las intensidades en el rango 0-255. En la ecuación 2 se muestra la función obtenida después de aplicar los umbrales encontrados usando el método Otsu, esta nueva imagen $g(x, y)$ contiene solo la geometría de la red en una escala binaria.

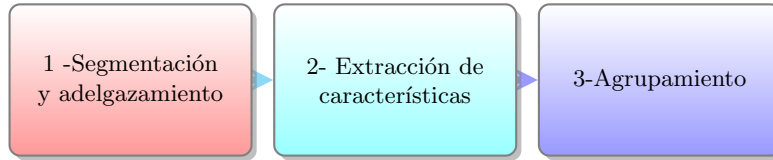


Fig. 2: Etapas del método de clasificación.

La figura 3 muestra la imagen obtenida después de aplicar el proceso de segmentación Otsu, la región de interés se define en la imagen con el nivel de intensidad blanco(255):

$$g(x, y) = \begin{cases} 0 & f(x, y) < t_1, \\ 0 & t_1 \leq f(x, y) \leq t_2, \\ 1 & f(x, y) > t_2. \end{cases} \quad (2)$$

En la figura 3 se muestra la imagen $f(x,y)$, y la imagen binaria $g(x,y)$ obtenida en el proceso de segmentación, esta imagen es de nuestro interés ya que contiene la morfología de la red de mitocondrias.

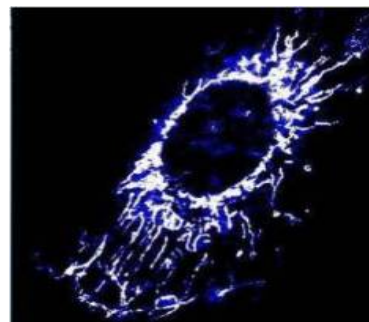
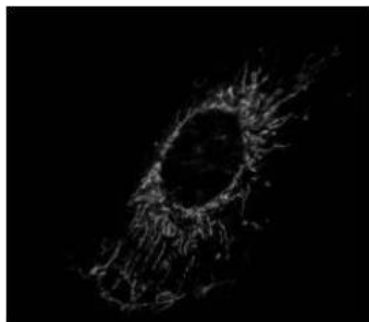


Fig. 3: Proceso de segmentación.

2.1.2. Adelgazamiento Previo a la extracción de características se emplea un algoritmo de adelgazamiento de la morfología de la imagen. El objetivo de este proceso es transformar la morfología de la red mitocondrial en un grafo, manteniendo las partes conexas. El método de adelgazamiento utilizado se fundamenta en la morfología matemática aplicada en imágenes, este tipo de morfología consiste en definir una imagen como un conjunto de elementos(pixeles) y aplicar

determinadas operaciones utilizando conjuntos establecidos [5], estos conjuntos establecidos son llamados elementos estructurales.

El algoritmo consiste en realizar una operación denominada acierto fallo $g(x, y) \otimes b$ utilizando ocho elementos estructurales b , esta operación acierto o fallo se aplica sobre la imagen utilizando cada uno de los ocho elementos estructurales, y aplicando la operación sobre el resultado de operar el elemento $n - 1$ sobre la imagen, tal como se define en la ecuación:

$$h(x, y) = (..((g(x, y) \otimes b_1) \otimes b_2 \dots \otimes b_8), \quad (3)$$

donde $g(x, y)$ es la imagen obtenida en la etapa 1 y $\{b_1 \dots b_8\}$ son los conjuntos o elementos estructurales que operan sobre la imagen, permitiendo obtener un adelgazamiento de la geometría sin afectar sus propiedades conexas. En la figura 4 se muestra la imagen $h(x, y)$ obtenida a partir de $g(x, y)$, geometría que determina el grafo que representa a la red de mitocondrias en la imagen.

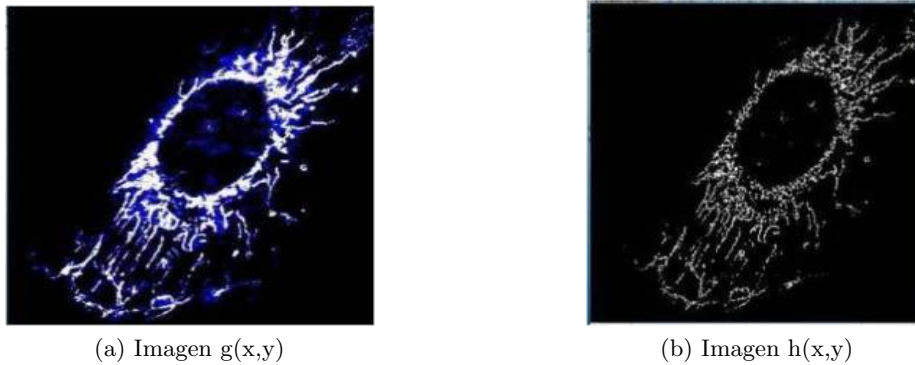


Fig. 4: Proceso de adelgazamiento.

3. Etapa 3

En esta etapa se propone un conjunto características que definen la morfología de la red mitocondrial, estas propiedades son normalmente aplicadas a redes complejas que topologicamente son grafos, con estas características se pretende establecer una diferencia cuantitativa en cada grupo de imágenes. Las medidas propuestas se obtienen estableciendo un nodo v como la unidad mínima dentro de la imagen(pixel), por otro lado las adyacencias de v con sus pixeles vecinos se definidas como las aristas e , un nodo v dentro de la imagen se toma en cuenta si el valor del pixel es igual al nivel de intensidad blanco en escala binaria. A continuación se describen las características:

3.1. Longitud promedio de la trayectoria

Se define a w como el promedio de pasos a través de la distancia (camino mas corto) de todos los posibles pares de nodos. Donde N es el número total de nodos v y $d_{v_i v_j}$ es la distancia del nodo v_i al nodo v_j :

$$w = \frac{1}{N(N-1)} \sum_{i \neq j} d_{v_i v_j}. \quad (4)$$

3.2. Centralidad

Sea d_{ij} el camino mas corto posible entre los nodos v_i y v_j (también conocida como distancia geodésica). La centralidad x mide la distancia de un nodo, con respecto al nodo mas cercano:

$$x = \frac{N-1}{\sum_{i \neq j} d_{ij}}. \quad (5)$$

3.3. Excentricidad

La excentricidad y se define como la máxima distancia geodésica entre los nodos v_i y v_j donde $i \neq j$. En otras palabras es la máxima distancia de todos los pares de nodos:

$$y = \max_{v_i v_j} \{d_{v_i v_j}\}. \quad (6)$$

3.4. Densidad

Se puede definir a z como el radio de aristas en el grafo G con relación al número máximo de posibles aristas. Donde N es el número total de nodos en el grafo y E el número total de aristas:

$$z = \frac{2E}{N(N-1)}. \quad (7)$$

4. Clasificación

En este proceso se asigna el valor escalar obtenido en cada características w , x , y , z a un vector de cuatro dimensiones (w, x, y, z) por cada imagen, posteriormente se definen cuatro grupos correspondientes a los tipos de morfologías definidas[1], después se aplica el algoritmo K-means para establecer el agrupamiento de las cuatro clases[6]. La ecuación 8 muestra la terminología establecida para el algoritmo.

$$\begin{aligned}
 &I(w, x, y, z) \text{ Vector imagen} \\
 &m_k \text{ centroide del cluster } k \\
 &r_{nk} \text{ Elemento } r \text{ asignado al cluster } k \\
 &X_n \text{ Vector } X \text{ de la imagen } n \\
 &J \text{ Distancia de los vectores del grupo } k \text{ respecto a su centroide}
 \end{aligned} \tag{8}$$

Una vez establecido el vector (w, x, y, z) , el algoritmo busca minimizar el error J determinado por la suma de las distancias euclidianas de los vectores r con respecto al centroide m_k , pertenecientes a la clase k , obteniendo cada uno de los centroides por cada iteración, hasta que el error total sea menor a una condición de paro establecida, de esta forma se obtienen los cuatro grupos de imágenes que corresponden a los cuatro tipos de morfología (completamente tubular, tubular con fragmentos, fragmentada con túbulos, completamente fragmentada).

algoritmo 1 Proceso de clasificación

```

1: while  $J > \text{error}$ 
2: for 1 to N
3:

$$r_{nk} = \begin{cases} 1 & k = \min ||x_n - m_k||^2 \\ 0 & \text{otro} \end{cases}$$

4:  $m_k = \frac{\sum_N r_{nk} x_n}{\sum_N r_{nk}}$ 
5:  $J = \sum_{n=1}^N \sum_{k=1}^K ||x_n - m_k||^2$ 
6: end for
7: end while

```

5. Análisis de resultados

Se utilizó un conjunto de 30 imágenes prueba correspondientes a los cuatro tipos de morfología de las redes mitocondriales, este conjunto fue clasificado por el expertos en el área despues de presentar los resultados obtenidos por el algoritmo propuesto. Para probar la eficiencia del clasificador y obtener algunas medidas de calidad (exactitud,error, sensibilidad, especificidad) se implementó el algoritmo y se compararon los resultados contra los resultados asignados por el experto: 5 completamente tubulares, 8 tubulares con fragmentos, 10 fragmentadas con túbulos y 7 fragmentadas, con estos datos se calculó la matriz de confusión mostrada en la tabla 1. Las filas de la tabla representan la clase a la cual pertenece la imagen, y las columnas representan la clase a la cual fue asignada la imagen por el clasificador. De esta forma se puede obtener el porcentaje de aciertos y fallas del conjunto de prueba.

Tabla 1: Matriz de confusión.

.	Tubular	Tubular con fragmentos	Fragmentada con túbulos	Fragmentada
Tubular	5	2	1	1
Tubular con fragmentos	2	2	1	0
Fragmentada con túbulos	1	4	6	0
Fragmentada	0	2	0	3

La figura 5 muestra la gráfica de la matriz de confusión, donde se puede ver mas a detalle en donde se encuentra acumulado la mayor cantidad de aciertos y fallos. Por ejemplo en la clase 1(Completamente tubular) y la clase 4(Fragmentadas) se obtienen los mejores resultados de clasificación obteniendo mayor número de aciertos. Por otro lado se observa que el mayor porcentaje de fallo en la clasificación se deriva de las clases 2(Tubular con fragmentos) y 3(Fragmentada con túbulos).

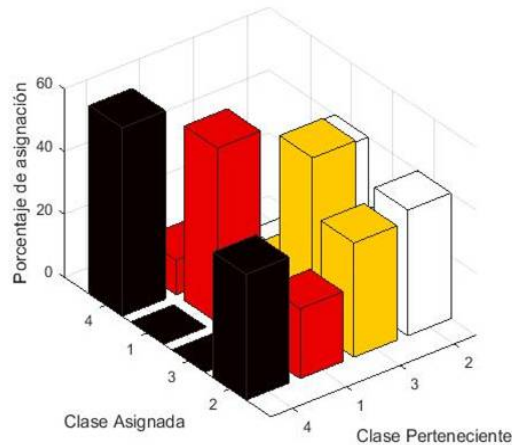


Fig. 5: Gráfica de la matriz de confusión.

$$Exactitud : 63,33\% . \tag{9}$$

Como se puede visualizar en la gráfica de la figura 5, el mayor porcentaje de aciertos corresponde a las clases 1 y 4, esto es debido a que algunas de las medidas

propuestas aplicadas en el grafo de la morfología presentan una diferencia notable las medidas propuestas, sin embargo las medidas como excentricidad, centralidad y longitud promedio, son vulnerables a escalabilidad en la imagen, lo que provoca una disminución o aumento en la magnitud obtenida, factor que influye en el porcentaje de aciertos.

$$\text{Error} : 36,66\%. \quad (10)$$

El mayor porcentaje de error es obtenido en la clasificación de grupos intermedios (2 y 3), debido a que presentan similitudes en las medidas propuestas, en estas circunstancias, el algoritmo de agrupamiento juega un factor importante, pues al aplicar el algoritmo K-means no se obtienen los clusters y el error no disminuye.

$$\text{Sensibilidad} : 58,12\%. \quad (11)$$

El número de aciertos por cada clase fue determinado a partir de la comparación de los resultados y la clasificación realizada por el experto para este conjunto de prueba, la sensibilidad obtenida se encuentra en su mayor parte definida por la clase 1 y 4, que es donde se obtuvo la mayor cantidad de positivos reales (elementos clasificados correctamente).

$$\text{Especificidad} : 84,09\%. \quad (12)$$

Este porcentaje se obtuvo debido a la incorrecta clasificación en las clases intermedias (2 y 3), las magnitudes en los escalares obtenidos para las características en estas clases, no presentan diferencia notable, por lo tanto un elemento puede ser mal clasificado con una mayor probabilidad.

6. Conclusiones

Las medidas propuestas para la caracterización de las redes mitocondriales son de gran apoyo para definir a las morfologías establecidas por los investigadores expertos en el área. Sin embargo el error obtenido (36.66%), deben disminuir con el objetivo de arrojar mejores resultados en los estudios realizados[1]. Las propiedades empleadas para la obtención de características son vulnerables a factores como escalabilidad en la imagen, tiempo de procesamiento y rotación, un ejemplo de esta afectación es la distancia entre pares de nodos, cuya magnitud puede ser afectada por escalabilidad o rotación de la imagen. Debido a estos factores, una propuesta del procedimiento es implementar un aprendizaje supervisado, en el cual los factores de afectación descritos, no jueguen un papel importante en el proceso de clasificación, además se pueda entrenar al sistema con un conjunto determinado para aumentar el porcentaje de exactitud y disminuir el error.

Agradecimientos. Agradecemos el apoyo al Consejo Nacional de Ciencia y Tecnología CONACYT y al Centro de Investigación en Computación del Instituto Politécnico Nacional.

Referencias

1. Maycotte G. P. , Goyri A. M., Medina B. D., Reyes L. J, Anaya R. M, Cortes H. P: Estudio de parámetros asociados a la dinámica mitocondrial en cáncer de mama. Ctedras CONACYT (2016)
2. Rehman J., Zhang H., Toth P., Zhang Y., Marsboom G., Hong Z., Salgia R., Husain A., Wietholt C, Archer L. S: Inhibition of mitochondrial fission prevents cell cycle progression in lung cancer. *The FASEB Journal*, pp. 26–36 (2012)
3. Giedt R. J., Feruglio P. F., Pathania D., Yang K. S., Kilcoyne A., Vinegoni C., Mitchison T. J. , Weissleder R: Computational imaging reveals mitochondrial morphology as a biomarker of cancer phenotype and drug response. *Scientific Reports*, pp. 85–95 (2016)
4. Castillo M., E.: Estudio de neoplasias malignas utilizando dinámica no lineal. Master Tesis, pp. 32–40 (2009)
5. Gonzalez C. R., Woods R: *Digital Image Processing*. Prentice Hall, pp. 78–135 (2007)
6. Bishop C. M: *Pattern Recognition and Machine Learning*. Springer, pp. 423–428 (2006)