

Segmentación de núcleos celulares mediante métodos de aprendizaje de máquina

Eduard Leonardo Sierra Ballén *

*Maestría en Ingeniería - Ingeniería de Sistemas y Computación

Departamento de Ingeniería de Sistemas e Industrial

Universidad Nacional de Colombia

Email: elsierrab@unal.edu.co

Abstract—El principal objetivo de este trabajo es desarrollar métodos de medida para dar soporte al patólogo en el diagnóstico de lesiones escamosas cervicales. Para esto, una de las primeras labores se concentra en la segmentación de las células obteniendo una representación más simple de la imagen para poder establecer medidas que sean útiles en el diagnóstico de dichas lesiones. El presente escrito muestra algunos avances iniciales en la segmentación de núcleos celulares.

Index Terms—Cancer cervical, lesiones escamosas, segmentación de imágenes en color, segmentación de células, Maquinas de vectores de soporte (Support Vector Machine SVM), perceptrones multicapa.

I. INTRODUCCIÓN

El cáncer de cuello uterino (cáncer cervical) es hoy el segundo más común en mujeres en todo el mundo [1]. El diagnóstico temprano de esta enfermedad en una etapa precancerosa es muy importante para la efectividad de los tratamientos. Existen diversos exámenes para el diagnóstico del cáncer de cuello uterino algunos de estos son: la citología vaginal, la colposcopia, la inspección visual directa y las espectroscopías fluorescente y reflectiva. La mayoría de países del tercer mundo usan estudios de citología vaginal o inspección visual directa ya que son de muy bajo costo. En los sitios donde se ha implementado la citología vaginal se ha logrado bajar la incidencia de cáncer en un 70% pero esta prueba tiene inconvenientes ya que la sensibilidad de la misma es sólo del 51%, y la interpretación para dar un diagnóstico es altamente subjetiva [1].

En el mundo hay varios proyectos que tratan de estudiar nuevas formas de bajar la incidencia y/o de encontrar nuevos tratamientos para el cáncer de cuello uterino entre estos se encuentran: la construcción de redes de investigación para estudiar la incidencia del virus del papiloma humano en el desarrollo de este cáncer [1]; la implementación de herramientas computacionales que den soporte objetivo en la determinación del diagnóstico asociado a los exámenes de patología como por ejemplo el proyecto realizado en el Reino Unido por Price et. al. [2].

Este artículo trata de la segmentación de núcleos celulares, esta tarea es un paso inicial en el desarrollo de un proyecto en el cual se pretende determinar medidas morfométricas sobre imágenes de cervix. Dichas medidas se toman con el fin de ayudar a los expertos patólogos en el diagnóstico de lesiones escamosas intraepiteliales, las cuales son un indicador de la

presencia de tejido canceroso en el cuello uterino. Para esto, el proyecto se apoya en el conocimiento a priori acerca de los criterios empleados por los patólogos para dicho diagnóstico, algunos de estos son:

- Pérdida de la polaridad de las células.
- Mitosis anormales.
- Alteración de la relación núcleo-citoplasma.
- Hiperchromasia nuclear.
- Alteración de la maduración.

El artículo se divide de la siguiente manera: en la sección II se revisan algunos antecedentes acerca del diagnóstico asistido por computador y la segmentación de células; la sección III presenta el conjunto de imágenes y algunos métodos empleados; en la sección IV se muestran los experimentos realizados y los resultados obtenidos; finalmente en la sección VI se dan algunas conclusiones y perspectivas de trabajo a desarrollar.

II. ANTECEDENTES

A. Diagnóstico de cáncer asistido por computador

Recientemente se han reportado diferentes proyectos que buscan introducir el diagnóstico automático de cáncer basado en imágenes histopatológicas. En el reporte técnico de Demir y Yener [3] se encuentran documentados algunos de los pasos del diagnóstico de cáncer, ilustrados en la figura 1, estos pasos incluyen: preprocesamiento para eliminar ruido y segmentar regiones de interés en imágenes microscópicas; selección y extracción de características. Para el caso del cáncer de origen cervical se encuentran características basadas en textura, en fractales y en topología. Finalmente para determinar el diagnóstico se emplean algoritmos de aprendizaje con clasificadores, métodos de evaluación y pruebas estadísticas. Otro ejemplo es proyecto realizado por Price et. al. ellos desarrollaron un sistema de decisión basado en redes de decisión bayesianas (bayesian belief networks). Este sistema busca ayudar a los patólogos a dar diagnósticos más objetivos, hacerlos fácilmente reproducibles y ayudar a patólogos jóvenes en el aprendizaje [2].

Por otro lado, Loukas y Linney [4], hicieron un análisis de imágenes histológicas basado en tres factores biológicos que influyen en la respuesta de los tumores a la radiación: la proliferación, la hipoxia y la vasculatura. Al igual que

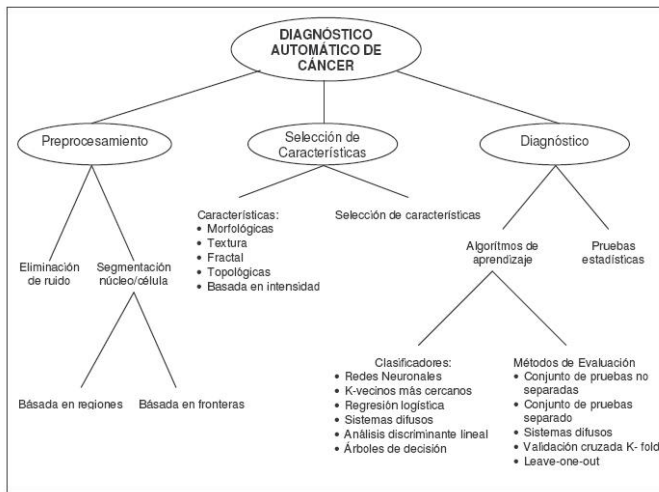


Fig. 1. Revisión de pasos en diagnóstico automático de cáncer. Adaptado de Demir y Yener [3]

la revisión de Demir y Yener [3] presentan diferentes pasos donde se destaca el problema de segmentar las células.

B. Segmentación de células

La segmentación de células se ha tratado desde diversos enfoques que incluyen técnicas tradicionales de imágenes usando umbralización, detección de bordes con máscaras de gradiente, hasta técnicas donde se aprovecha la información de color incluyendo métodos usando modelos donde se separa la componente de iluminación como el HSI (Hue Saturation Intensity) y el HSL (Hue Saturation Luminance). También uso de algoritmos de clasificación de píxeles en el espacio de color RGB, usando análisis de componentes principales (ACP), fusión de regiones, emparejamiento de plantillas (contornos activos, transformada de Hough), análisis de texturas y clasificación (Redes neuronales, lógica difusa) [4].

Lezoray [5] en su tesis doctoral hace segmentación de células en imágenes en color con morfología matemática aplicada a imágenes de redes neuronales. Un trabajo anterior hecho por Walker et al. en 1994 [6], aplicado a células cervicales usa también morfología y extracción de características de textura.

Otros enfoques tratan de solucionar problemas que se presentan con segmentación de células usando representaciones paramétricas para hacer emparejamiento de plantillas como contornos activos o transformada Hough [4]. Entre estas también hay técnicas que plantean el problema de segmentar las células como un problema de optimización de una representación paramétrica de la forma celular. Segmentación satisfactoria de células aisladas, múltiples células y células traslapadas se logra con la técnica de Wu, Barba y Gil [7]. Este método hace minimización sobre una representación paramétrica pero el costo computacional es prohibitivo para imágenes grandes donde hay muchas células. Otro esquema basado también en optimización usa un método evolutivo basado en búsqueda tabú de Jiang y Yang [8]. Este método usa memoria de corto plazo y tiene una lista para diversificar

el espacio de búsqueda pero el problema está precisamente en la determinación del tamaño de esta lista. El número de familias es importante, si es pequeño se acerca a la búsqueda tabú, de otra manera la velocidad de convergencia empeora. Además el determinar al número de hijos también afecta la velocidad de convergencia para ese caso usaron valores entre 2 y 8. Además los autores encontraron que este método tenía casos donde no se obtenían segmentaciones óptimas.

Entre las técnicas de segmentación de células con aplicaciones en cancerología hay una comparación de métodos supervisados, basada en píxeles hecho por Meurie et al. [9], con imágenes de cáncer bronquial, en donde usa un esquema de segmentación en cinco etapas: simplificación de imágenes usando pre-tratamiento con suavizado, eliminación de ruido; seguido de clasificación de píxeles usando Bayes, kNN, SVM; extracción de marcadores basados en morfología matemática; luego watersheds para hacer crecimiento usando la información de color de la imagen obteniendo las regiones correspondientes al núcleo y al citoplasma. En este trabajo se concluyó que es de importancia la escogencia del binomio compuesto por el modelo de color y del clasificador para obtener mejores resultados, hay diferentes combinaciones que dan mejores resultados ellos encontraron que para Bayes los mejores espacios de color son YCh1Ch2, RGB y XYZ, para SVM son YCh1Ch2, HSL y RGB, y para kNN los mejores son HSL, XYZ y RGB [9].

III. MATERIALES Y MÉTODOS

A. Material Histológico

El proyecto se desarrolla con imágenes digitales en color tomadas de muestras de cérvix con método de tinción (Hematoxilina-Eosina).

B. Análisis de imágenes

Inicialmente, el estudio se ha centrado en la segmentación de los núcleos celulares. En las imágenes disponibles se identificaron varios casos de presencia de núcleos estas son núcleos aislados, en mitosis, traslapados con otras células y traslapados con otros núcleos celulares. Además, se debe tener en cuenta que hay casos donde en este tipo de tejidos se presentan células que tienen más de un núcleo, la figura 2 presenta algunos de los casos nombrados anteriormente.

Hasta el momento, se ha trabajado específicamente en la determinación de núcleos aislados. Inicialmente el problema de segmentación se trabajó sobre las imágenes en escalas de grises, tomando la componente de iluminación Y del modelo de color YIQ ($0.299R + 0.587G + 0.114B$), la figura 3 presenta un ejemplo donde se nota como al emplear la imagen en escala de grises se pueden confundir algunos núcleos esto debido al fenómeno de metamerismo [10].

La complejidad de estas imágenes se hace manifiesta debido a que el método de captura empleado y la tinción, hacen que se afecte bastante la calidad de las imágenes. Ya que la coloración de estas imágenes es variable en una misma imagen y entre imágenes. Lo cual obliga a buscar alternativas de segmentación empleando la información de color.

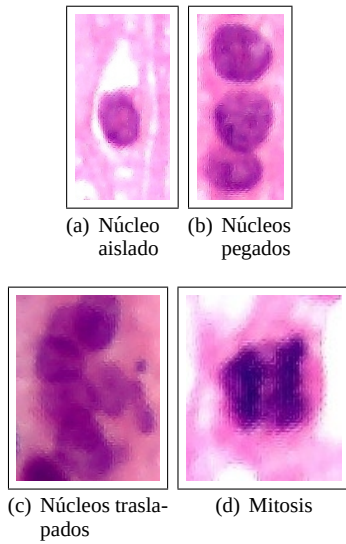


Fig. 2. Algunos casos que se presentan para la segmentación de núcleos celulares.

C. Segmentación de núcleos celulares

Debido a que la coloración de los objetos en cada imagen puede variar bastante, pero que también se encuentra que en general la coloración se encuentra separada en cuanto a su tonalidad, los núcleos son de tonalidad violeta casi azul, y el citoplasma de coloración rosa. Se plantea un esquema de segmentación de imágenes que haga una separación en el espacio de color RGB, esta segmentación se hace mediante una metodología experimental usando diversos métodos de aprendizaje de máquina, se empleó regresión lineal y polinomial y máquinas de vectores de soporte.

1) *Regresión*: Para hacer clasificación via regresión se plantea el problema de determinar la función (en este caso línea o polinomio) que mejor represente los datos y se plantea de manera que se puede determinar un criterio sencillo para clasificación. De manera formal se tiene un conjunto de entrenamiento $\{y_i, \mathbf{x}_i\}$ y se construirá una función $f(\mathbf{x}_i)$ que permita estimar de la mejor manera la solución y_i , en el caso de regresión se prefiere la función que tenga un menor error cuadrático medio. Ahora para hacer clasificación por ejemplo con dos posibles clases se define que $y_i \in \{0, 1\}$. De tal manera que se determinaría la pertenencia a una u otra clase siguiendo el criterio si $f(\mathbf{x}_i) < 0.5$, $\mathbf{x}_i \in 0$ o de lo contrario $\mathbf{x}_i \in 1$.

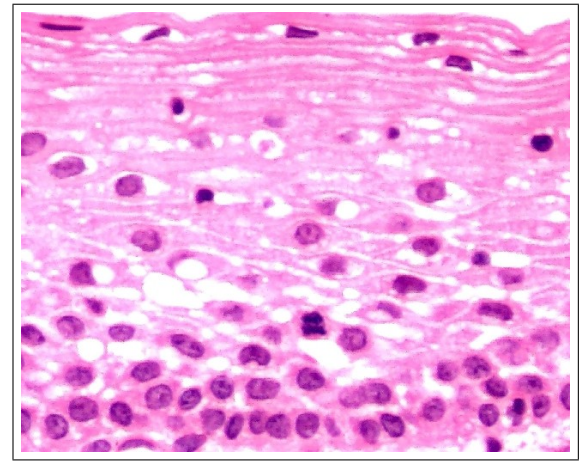
En este artículo se trabajó con regresión lineal

$$y = f(\mathbf{x}; \hat{\mathbf{w}}) = \hat{w}_0 + \sum_{i=1}^N \hat{w}_i x_i \quad (1)$$

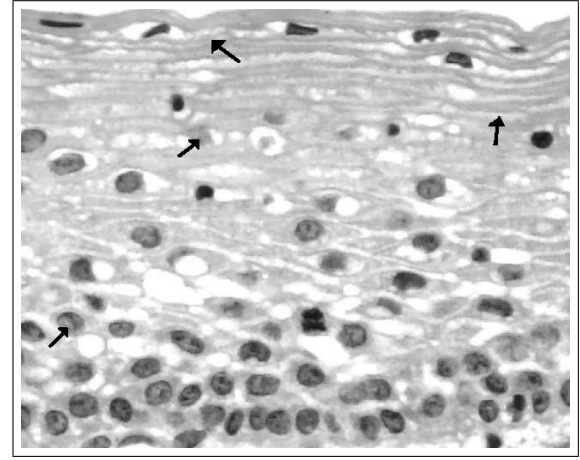
y regresión polinomial

$$y = f(\mathbf{x}; \hat{\mathbf{w}}) = \hat{w}_0 + \sum_{i=1}^N \sum_{d=1}^m \hat{w}_{i,d} x_i^d \quad (2)$$

2) *Máquinas de Vectores de Soporte (Support Vector Machine -SVM-)*: Este es un método de aprendizaje de máquina que hace una clasificación para patrones de entrenamiento de la forma $\{y_i, \mathbf{x}_i\}$, donde $y_i \in \{1, -1\}$, y $\mathbf{x}_i \in \mathbf{R}^n$. El método



(a) Imagen Original



(b) Imagen en escala de grises

Fig. 3. Imagen original en color y escala de grises usando la componente Y del modelo de color YIQ, las flechas indican algunos sitios donde se pueden confundir núcleo y citoplasma celular

da por resultado un hiperplano de separación determinado por la ecuación:

$$\mathbf{w}^T \mathbf{x} + w_0 = 0 \quad (3)$$

Donde la clasificación estaría dada por:

$$\mathbf{w}^T \mathbf{x}_i + w_0 \geq 1, \text{ para } y_i = 1 \quad (4)$$

$$\mathbf{w}^T \mathbf{x}_i + w_0 \leq -1, \text{ para } y_i = -1 \quad (5)$$

Este método plantea un esquema donde se busca maximizar la separación entre las clases usando unos puntos de entrenamiento denominados vectores de soporte los cuales van a soportar el hiperplano. Esto es equivalente a la minimización de $\frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} \mathbf{w}^T \mathbf{w}$. El objetivo final del método es obtener el vector de pesos $\mathbf{w}$ y el sesgo w_0 que forman el hiperplano de separación. Este problema se plantea como:

Minimizar :

$$J(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad (6)$$

Sujeto a:

$$y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1 \geq 0 \quad i = 1, \dots, N \quad (7)$$

Usando el método de multiplicadores de Lagrange esto es:

Minimizar :

$$J(\mathbf{w}, w_0, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i [y_i (\mathbf{w}^T \mathbf{x}_i + w_0) - 1] \quad (8)$$

Esta formulación es compelja ya que requiere determinar tanto los pesos α_i de Lagrange como los pesos w del hiperplano. Por lo anterior esta formulación se emplea para contruir el dual de este problema, el cual se puede expresar únicamente en términos de los datos de entrenamiento y permite determinar los pesos α_i . Así el problema a solucionar ahora es:

Maximizar :

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (9)$$

sujeto a $0 \leq \alpha_i \leq C$ y $\sum_i \alpha_i y_i = 0$ y C es un parámetro a seleccionar que se utiliza cuando las clases no son linealmente separables.

De la solución del anterior problema se puede obtener los valores de $\mathbf{w}$ y w_0 usando las expresiones:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \quad (10)$$

$$w_0 = 1 - \mathbf{w}^T \mathbf{x}^{(s)} \quad (11)$$

donde $\mathbf{x}^{(s)}$ es un vector de soporte, es decir un vector donde $y_i = \pm 1$ y $\alpha_i \neq 0$ Detalles más a fondo acerca de este método se pueden ver en [11], [12].

IV. EXPERIMENTACIÓN Y RESULTADOS

A. Conjunto de datos

Se cuenta con un conjunto de 29781 datos obtenidos por marcación directa hecha por un patólogo experto sobre imágenes de biopsias de cérvix tomadas con método de tinción Hematoxilina-Eosina. Los datos corresponden a las categorías núcleo (0 -cero-) o citoplasma (1 -uno-) y los valores de color RGB, coordenadas x, y donde ocurren. La figura 4 presenta la imagen original y la imagen con las fronteras de los núcleos y citoplasmas de 35 células marcadas en color verde por el patólogo.

A continuación se presentan los resultados experimentales de aplicar diversos métodos de aprendizaje de máquina para determinar núcleos (regiones de color oscuro) y citoplasmas (regiones de color más claro alrededor de los núcleos), sobre un conjunto de 1000 de los puntos del conjunto. Se tomaron los primeros 1000 datos del dataset después de eliminar la información de posición y eliminar datos redundantes ya que solo se usará la información de color RGB y en el dataset completo hay bastantes de estos datos repetidos la figura 5 presenta los datos de experimentación con respecto a cada una de sus variables.

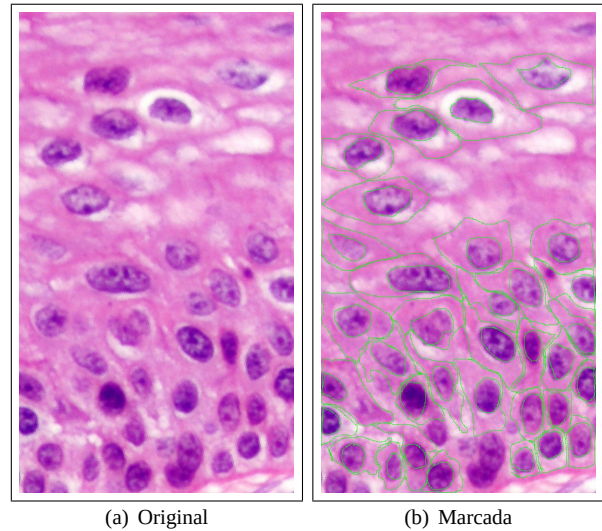


Fig. 4. Imágenes original y marcada por el patólogo

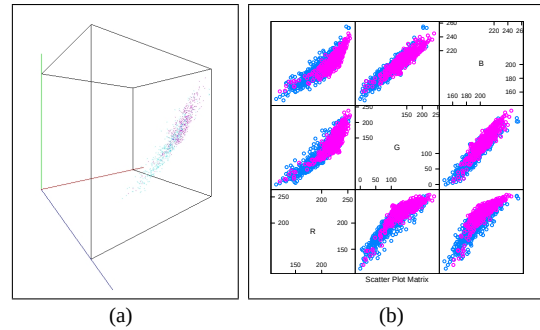


Fig. 5. Visualización del conjunto de datos de entrenamiento correspondiente a fronteras de núcleos (cyan) y de citoplasmas (magenta)

B. Regresión Lineal

En la tabla I y la figura 6 se presenta el error de entrenamiento y de prueba variando los tamaños de entrenamiento tomando fracciones del conjunto de datos del 10%, 20%, 30%, 40%, 50%, ..., 90%; al aplicar regresión lineal de la forma

$$y = f(\mathbf{x}; \hat{\mathbf{w}}) = \hat{w}_0 + \sum_{i=1}^N \hat{w}_i x_i \quad (12)$$

Donde N es la dimensión del vector $\mathbf{x}$, para este caso 3.

Tamaño	Error de Entrenamiento	Error de Prueba
100	0.1756210	0.1793968
200	0.1744159	0.1772189
300	0.1774611	0.1749550
400	0.1746430	0.1767924
500	0.1731699	0.1781946
600	0.1795937	0.1701685
700	0.1812579	0.1630131
800	0.1754624	0.1753424
900	0.1747898	0.1794707

TABLE I

ERROR DE PRUEBA Y ENTRENAMIENTO OBTENIDO POR REGRESIÓN LINEAL

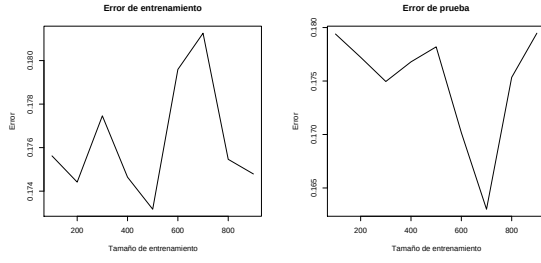


Fig. 6. Error de entrenamiento y de prueba obtenidos por regresión lineal variando el tamaño del conjunto de entrenamiento

La figura VI presenta el resultado de segmentar la imagen de entrenamiento con el resultado de la regresión lineal usando un conjunto de entrenamiento de 700 datos ya que tiene el error de entrenamiento más bajo. Los pesos para la línea de regresión son:

$$w_0 = 0.793873304\mathbf{w} = \begin{pmatrix} 0.013235941 \\ 0.003960904 \\ -0.017546224 \end{pmatrix} \quad (13)$$

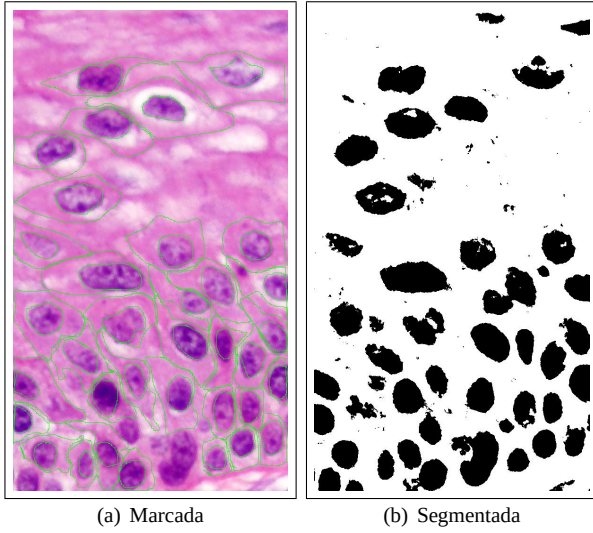


Fig. 7. Segmentación usando regresión lineal

C. Regresión Polinomial

En las tablas II y III se presenta el error de entrenamiento y de prueba variando los tamaños de entrenamiento tomando fracciones del conjunto de datos del 10%, 20%, 30%, 40%, 50%, ..., 90%; al aplicar regresión polinomial de la forma

$$y = f(\mathbf{x}; \mathbf{w}) = w_0 + \sum_{i=1}^N \sum_{d=1}^m w_{i,d} x_i^d \quad (14)$$

Donde N es la dimensión del vector $\mathbf{x}$, para este caso 3 y d es el grado del polinomio.

La figura VI presenta el resultado de segmentar la imagen de entrenamiento con el resultado de la regresión polinomial

Grado	Tamaño del conjunto de entrenamiento								
	100	200	300	400	500	600	700	800	900
2	0.1573712	0.1609569	0.1557858	0.1547649	0.1549284	0.1577421	0.1604243	0.1528224	0.1509655
3	0.1555507	0.1571203	0.1531613	0.1536293	0.1542554	0.1569261	0.1596344	0.1518967	0.1500146
4	0.1538338	0.1559723	0.1526449	0.1534464	0.1540707	0.1566005	0.1590849	0.1511732	0.1493176
5	0.1507522	0.1555481	0.1512287	0.1529379	0.1538998	0.1563946	0.1590112	0.1510302	0.1490585
6	0.1487909	0.1554517	0.1516900	0.1523875	0.1533846	0.1560218	0.1594638	0.1509329	0.1490069
7	0.1481255	0.1535712	0.1516029	0.1697833	0.1540871	0.1553934	0.1581109	0.1504515	0.1485526
8	0.1480149	0.1518731	0.1493092	0.1506106	0.1525274	0.1548649	0.1579342	0.1503533	0.1485116
9	0.1471867	0.1515881	0.1491158	0.1495943	0.1520245	0.1545724	0.1578125	0.1502213	0.1483778
10	0.2330234	0.1502133	0.1486296	0.1485556	0.1516485	0.1541405	0.1576477	0.1499483	0.1480670

TABLE II

ERROR DE ENTRENAMIENTO OBTENIDO POR REGRESIÓN POLINOMIAL

Grado	Tamaño del conjunto de entrenamiento								
	100	200	300	400	500	600	700	800	900
2	0.1597720	0.1507025	0.1507793	0.1496193	0.1495119	0.1493318	0.1491169	0.1487623	0.1486717
3	0.1624829	0.1561014	0.1519235	0.1492895	0.1491402	0.1487576	0.1485202	0.1483759	0.1481400
4	0.1717006	0.1604446	0.1518972	0.1490497	0.1488342	0.1479170	0.1475283	0.1473724	0.1471020
5	0.9972649	0.1593995	0.1545033	0.1488180	0.1484802	0.1479379	0.1473991	0.1470259	0.1467378
6	0.9972649	0.1593995	0.1545033	0.1488180	0.1484802	0.1479379	0.1473991	0.1470259	0.1467378
7	4.9684890	0.1854874	0.1646232	0.2022845	0.1528802	0.1478162	0.1469682	0.1464120	0.1461813
8	4.0337141	0.2357348	0.1584329	0.1496690	0.1486619	0.1481731	0.1472213	0.1464684	0.1461683
9	7.7335909	0.1692824	0.1545301	0.1493319	0.1487062	0.1482790	0.1468550	0.1460272	0.1457488
10	1409.3164406	0.2939282	0.1550507	0.1532065	0.1496406	0.1489853	0.1465847	0.1457245	0.1455346

TABLE III

ERROR DE PRUEBA OBTENIDO POR REGRESIÓN POLINOMIAL

grado 2 usando un conjunto de entrenamiento de 900 datos. Los pesos para el polinomio de regresión son:

$$w_0 = -3.721500e + 00, \mathbf{w} = \begin{pmatrix} -8.752422e - 02 & 2.620946e - 04 \\ 9.534193e - 04 & 7.040096e - 06 \\ 1.259927e - 01 & -3.602497e - 04 \end{pmatrix} \quad (15)$$

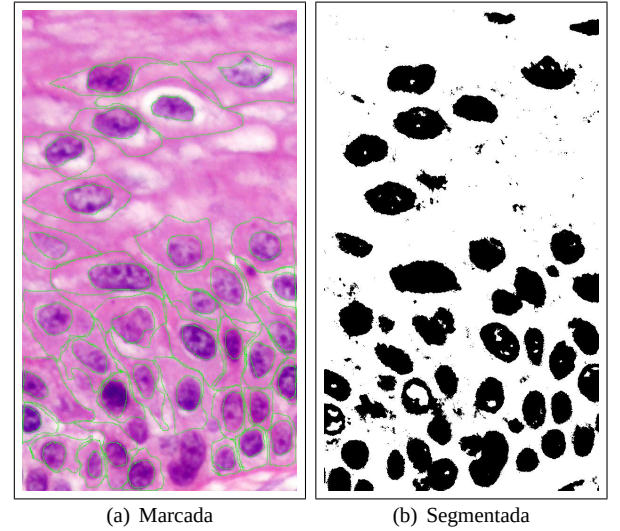


Fig. 8. Segmentación usando regresión polinomial grado 2

V. SUPPORT VECTOR MACHINE

Se experimento con los posibles kernel de SVM con sus valores por defecto, $C = 1$, polinomial grado =3, gamma= 1, coef0=0,

Por otro lado, la tabla V presenta el error de generalización usando k-fold cross validation con 10 folds para los métodos de SVM. Los resultados son:

Adicionalmente se experimentó variando los parámetros gamma de 1, 2, 3 y C 0.5, 1, 1.5, 2 del kernel de funciones de

Tamaño	SVM-Linear	SVM-Polynomial	SVM-Radial	SVM Sigmoid
100	24.1	28.3	23.6	34.2
200	22.8	26	23.3	35.8
300	21.6	31.1	22.3	36.7
400	22	32.3	20.3	36.7
500	23	33.5	21	37.4
600	23.5	33.8	20.4	37.6
700	23.5	33.6	20.3	38
800	22.7	34	20.3	36.7
900	22.8	34.3	20	37
1000	23	33.9	20.6	32

TABLE IV
ERROR DE CLASIFICACIÓN PARA SVM

Método	Error Cross Validation
SVM radial	0.213
SVM polynomial	0.338
SVM linear	0.233
SVM sigmoid	0.355

TABLE V
ERROR DE GENERALIZACIÓN CON CROSS VALIDATION 10 FOLDS PARA
LOS METODOS DE CLASIFICACIÓN

base radial, ya que es uno de los métodos que permite hacer proyecciones de los datos en espacios de mayor dimensión, los errores de generalización son los que se ven en la tabla VI.

Gamma	C				
		0.5	1	1.5	2
	1	0.214	0.21	0.21	0.212
	2	0.218	0.21	0.215	0.212
	3	0.216	0.213	0.214	0.221

TABLE VI
ERROR DE CLASIFICACIÓN PARA SVM

Finalmente se entrenó un SVM con una función de base radial ($\gamma=1, C=1$) aplicando el algoritmo SMO de la librería Weka [13]. La figura 9 muestra la segmentación de núcleos obtenida. Se puede observar como existen casos donde los núcleos no quedan totalmente separados especialmente en las regiones donde hay alta proliferación de células, además de huecos dentro de algunos núcleos producto de tonalidades similares a las del citoplasma. También hay regiones del citoplasma cuya tonalidad al parecer es similar a la del núcleo y se marcan como núcleo (ver figuras 9(a) y 9(b) izquierda abajo).

VI. CONCLUSIONES Y TRABAJO FUTURO

Se presentó el primer paso para la construcción de una morfometría en imágenes de cérvix empleando la información de color en el modelo de color RGB para hacer una segmentación de núcleos celulares.

Las imágenes de cérvix debido al método de tinción usado y a fenómenos como el metamerismo no son buenas candidatas para ser procesadas en escala de grises.

El conjunto de datos, no es separable linealmente de una forma sencilla y presenta una amplia posibilidad de tener datos que pueden llegar a estar etiquetados con las dos posibles clases debido a que el color no es el único indicador de la pertenencia de un punto a una clase.

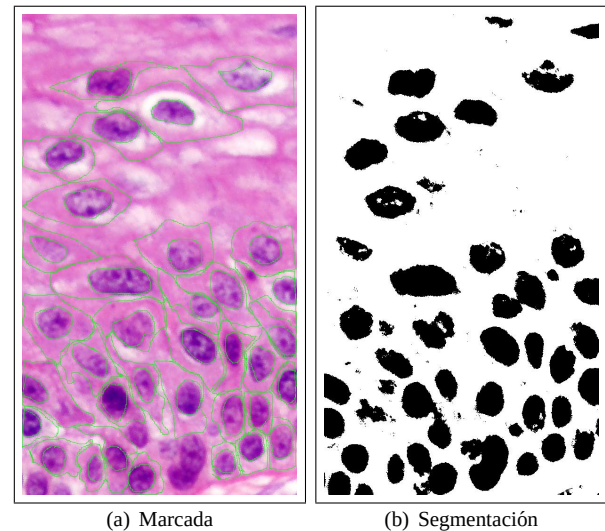


Fig. 9. Imagen original, contornos delineados por un experto y resultados de segmentación usando SVM algoritmo SMO con función Radial

La aproximación presentada es un paso inicial, se plantea la necesidad de buscar un modelo de segmentación donde los núcleos aislados queden bien segmentados ya que como se ve en las figuras 8 y 9 estos métodos requieren afinarse porque se obtienen regiones que no corresponden.

Se debe revisar la segmentación afinando parámetros del SVM, además de explorar modelos de color diferentes la RGB como es el caso de modelos que trabajan el color desde el punto de vista de percepción humana como el CIELab o modelos donde se separe el color en componentes separadas para tonalidad, saturación e iluminación como HSV ó HSL.

A pesar de lo anterior, con los experimentos realizados hasta el momento, la mejor clasificación desde el punto de vista de error de entrenamiento y prueba se obtienen al hacer regresión lineal simple, lo cual lo hace ser el preferido ya que es un método más sencillo de implementar.

REFERENCES

- [1] D. G. Ferris, "Analysis of digitized cervical images detect cervical neoplasia," in *Proceedings of SPIE*, ser. Medical Imaging 2004: Image Processing, J. M. Fitzpatrick and M. Sonka, Eds., vol. 5370, 2004, pp. 181–194.
- [2] G. J. Price, W. G. McCluggage, M. L. Morrison, G. McClean, L. Venkatraman, J. Diamond, H. Bharucha, R. Montironi, P. H. Bartels, D. Thompson, and P. W. Hamilton, "Computerized diagnostic decision support system for the classification of preinvasive cervical squamous lesions," *Human Pathology*, vol. 34, no. 11, pp. 1193–1203, 2003.
- [3] C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Rensselaer Polytechnic Institute, Department of Computer Science, Troy, NY 12180, US, Tech. Rep. TR-05-09, March 2005. [Online]. Available: <http://www.cs.rpi.edu/research/pdf/05-09.pdf>
- [4] C. G. Loukas and A. Linney, "A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature," *Computer Methods and Programs in Biomedicine*, vol. 74, no. 03, pp. 183–199, June 2004.
- [5] O. M. Lezoray, "Segmentation d'images couleur par morphologie mathématique et classification de données par réseaux de neurones : application à la classification de cellules en cytologie des séreuses," Ph. D. These, Ecole Doctorale Structure, Information, Matière et Matériaux, U.F.R de Sciences, Université de Caen/Basse-Normandie, 2000. [Online]. Available: <http://users.info.unicaen.fr/lezoray/Publis/These.pdf>

- [6] R. Walker, P. Jackwa, B. Lovell, and D. Longstaff, "Classification of cervical cell nuclei using morphological segmentation and textural feature extraction," in *2nd ANZIS Australian and New Zealand Conference on Intelligent Information Systems*, 1994.
- [7] H.-S. Wu, J. Barba, and J. Gil, "A parametric fitting algorithm for segmentation of cell images," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 03, pp. 400–407, March 1998.
- [8] T. Jiang and F. Yang, "An evolutionary tabu search for cell image segmentation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 32, no. 05, pp. 675–678, October 2002. [Online]. Available: http://nlprweb.ia.ac.cn/english/staff/jiangtz/Jiang_Yang.IEEE-TSMCB02.pdf
- [9] C. Meurie, G. Lebrun, O. Lezoray, and A. Elmoataz, "A comparison of supervised pixel-based color image segmentation methods. application in cancerology," *WEAS Transactions on Computers, special issue on ICOSIP'03*, vol. 02, no. 03, pp. 739–744, July 2003. [Online]. Available: http://users.info.unicaen.fr/lezoray/Publis/meurie_icosip03.pdf
- [10] T. Gevers and F. C. A. Groen, "Segmentation of color images," in *Proceedings of 7th Scandinavian Conference on Image Analysis*, 1991. [Online]. Available: <http://staff.science.uva.nl/gevers/>
- [11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Prentice Hall, 1999, ch. Support Vector Machines.
- [12] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, no. 2, pp. 121–167, 1998.
- [13] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann, 2005.

PLACE
PHOTO
HERE

Eduard Sierra Ingeniero de Sistemas de la Universidad Nacional de Colombia en 2002. Docente del Programa de Ingeniería en Multimedia Universidad Militar "Nueva Granada desde 2003. En 2005 inició sus estudios de Maestría en Ingeniería - Ingeniería de Sistemas y Computación en la Universidad Nacional de Colombia.