

Tipo de artículo: Artículo original
Temática: Procesamiento de imágenes
Recibido: 01/02/2015 | Aceptado: 14/01/2016

Evaluación de alternativas para la clasificación de células cervicales utilizando solo rasgos del núcleo

Evaluation of alternatives for cervical cell classification using only nucleus' features

Solangel Rodríguez Vázquez ^{1*}, Andy Vidal Martínez Borges ²

¹ Universidad de las Ciencias Informáticas. Km 2½ carretera San Antonio de los Baños, Rpto. Torrens, La Lisa, Ciudad de la Habana. svazquez@uci.cu

² Empresa de tecnología para la defensa, XETID. Km 2½ carretera San Antonio de los Baños, Rpto. Torrens, La Lisa, Ciudad de la Habana. avmartinez@xetid.cu

* Autor para correspondencia: svazquez@uci.cu

Resumen

En este artículo se presenta un estudio comparativo del desempeño de tres de las técnicas de clasificación más utilizadas para el diagnóstico en la prueba de Papanicolaou basándose solamente en las características extraídas de la región del núcleo. Dentro de las técnicas seleccionadas se encuentran los clasificadores vecinos más cercanos (*kNN*), redes neuronales artificiales (RNA-RBF Network) y máquinas de soporte vectorial (SVM). El estudio comparativo se realiza con el objetivo de determinar la técnica con mayor capacidad para clasificar de forma correcta los patrones que identifican los cambios en las células cervicales a partir de matrices de rasgos de imágenes de la prueba de Papanicolaou. Para el estudio se utilizaron los resultados obtenidos en las iteraciones realizadas a cada clasificador con iguales conjuntos de datos, con el objetivo de determinar cuál de los métodos propuestos ofrece la mejor solución al problema de la clasificación de células cervicales. Se realizó un estudio experimental en el que se consideraron dominios con igualdad en los números de clases, atributos y ejemplos de entrenamiento, así como igual proporción de casos pertenecientes a cada clase. La comparación estadística se realizó en función de los resultados obtenidos por los índices de efectividad conocidos como medida F, media H, predictividad negativa y área bajo la curva ROC (AUC).

Palabras clave: células cervicales, prueba de Papanicolaou, clasificación, *kNN*, redes neuronales, máquinas de soporte vectorial.

Abstract

This paper presents a comparative study of the performance of three of the most widely used classification techniques for diagnosis in Pap tests based solely on features extracted from the nucleus' region. Among the selected techniques are the nearest neighbor classifier (kNN), artificial neural networks (ANN-RBF Network) and support vector machines (SVM). The comparative study is performed in order to determine the technique with greater ability to correctly classify the patterns that identify changes in cervical cells from feature matrices from images of Papanicolaou smears. In this study the results obtained from the iterations in each classifier using the same data sets were used in order to determine, which method offers the best solution to the problem of cervical cells classification. An experimental study was conducted considering domains with equal numbers of classes, attributes, and training examples, as well as an equal proportion of cases belonging to each class. The statistical comparison was made based in the results obtained in terms of the indexes of effectiveness known as F-measure, H-mean, negative predictivity and area under the ROC curve (AUC).

Keywords: *cervical cells, Papanicolaou test, classification, kNN, neural networks, support vector machines.*

Introducción

El cáncer de cérvix uterino es una de las enfermedades más frecuentes entre las mujeres, aunque más del 80% se produce en países en vías de desarrollo. El cribado de cáncer de cérvix ha demostrado su efectividad para reducir la incidencia y la mortalidad por esta enfermedad, especialmente cuando se realiza a través de programas poblacionales organizados.

El cuello uterino es la abertura del útero que se encuentra en la parte superior de la vagina. Está cubierto por una capa fina de tejido compuesto por células. A medida que estas células se desarrollan, las que se encuentran en la capa inferior se trasladan lentamente a la superficie del cuello uterino, Durante este proceso, algunas células se pueden volver anómalas y ellas pueden devenir en un proceso canceroso. Estas alteraciones precancerosas se denominan displasia o lesiones escamosas intraepiteliales. La prueba de cribado o citología de Papanicolaou se utiliza para detectar células cervicales anómalas en personas que no presentan señales o síntomas. De esta manera es posible administrar tratamiento en las primeras etapas de la enfermedad.

Debido al carácter masivo de su aplicación en la población femenina, la prueba de Papanicolaou genera una considerable carga de trabajo para los laboratorios que analizan en el microscopio los frotis resultantes de esta técnica citológica. En un frotis típico se pueden encontrar hasta 300000 células, lo que limita la productividad a no más de

unas 60-80 citologías por día de trabajo y observador (Lorenzo-Ginori and Rodríguez-Santos, 2012). Además de esto, a partir de la experiencia de la aplicación de esta prueba durante muchos años, se ha podido determinar que existen diferentes factores que afectan la calidad de los resultados. Estos son, principalmente, los errores en la toma de las muestras, en su procesamiento y en su lectura e interpretación. Sobre este último caso, la necesidad de analizar una gran cantidad de muestras con muy baja tasa de casos positivos tiende a sesgar el resultado de la evaluación, y además provoca errores debidos a la rutina y a la fatiga de los analistas.

"Si tenemos en cuenta que la esperanza de vida promedio de la mujer en nuestro país es de 80 años, con cada fallecimiento se pierden 30 años de vida potenciales, por lo que es la principal causa de muerte prematura en la población femenina cubana. La carga para la sociedad es entonces muy alta, además de la discapacidad que genera en aquellas que no fallecen". (Acosta, 2012)

Como consecuencia de los diferentes errores mencionados, aparece una cierta tasa de falsos negativos en el resultado final. Estos implican que una lesión precursora pueda progresar, o que un carcinoma in situ (curable) pueda devenir en un cáncer invasivo, sin que la paciente sea sometida a tratamiento alguno. La aparición de falsos negativos motiva la necesidad de asegurar la calidad en la toma y procesamiento de las muestras, y además de disponer de métodos efectivos de revisión de las pruebas. Tanto en lo referido a la necesidad de incrementar la productividad de los laboratorios, como a la de reducir la tasa de falsos negativos, los métodos que utilizan el procesamiento digital de imágenes han comenzado a desempeñar, en los últimos años, un importante papel. (Lorenzo-Ginori and Rodríguez-Santos, 2012)

Según (Acosta, 2012), en 2011 se reportaron en Cuba 1334 casos nuevos de cáncer cérvicouterino, y unas 455 mujeres fallecieron por esta causa. Atendiendo a tal situación, se plantea que los programas de pesquisa y diagnóstico precoz del cáncer cérvicouterino y sus lesiones precursoras cobran cada vez mayor relevancia. Para mejorar la efectividad del cribado se han introducido nuevas técnicas de obtención de muestras citológicas, como la citología líquida, y se han desarrollado sistemas automatizados de lectura de las muestras citológicas que utilizan técnicas de clasificación.

Los métodos que se refieren a la clasificación de imágenes de Papanicolaou se basan en el cálculo de las características extraídas de las áreas del núcleo y el citoplasma (Marinakis, Dounias et al. 2009; Marinakis, Marinaki et al. 2008). Se basan generalmente en características de forma y la intensidad de los objetos de interés. Sin embargo, las características calculadas no presentan la misma capacidad de discriminación (Plissiti and Nikou, 2012 a). Para la determinación del conjunto de características más eficaz que se utiliza como entrada en un clasificador, se han

propuesto algunos esquemas de selección de características, y que se refieren a algoritmos genéticos (Marinakis, Dounias et al. 2009) y optimización de nubes de partículas (Marinakis, Marinaki et al. 2008). Estos métodos utilizan imágenes pre-segmentadas que contienen solo una célula, por lo que la segmentación correcta del núcleo y el citoplasma es factible. En las imágenes que contienen grupos de células, la detección de la frontera del citoplasma es un problema difícil y, hasta ahora, no se ha encontrado ninguna técnica en la literatura estudiada que mencione algoritmos con buenos resultados en la delimitación de las fronteras. Sin embargo, la detección y segmentación de los núcleos de las imágenes que contienen células superpuestas y agrupaciones de células ha sido abordado con éxito en varios estudios.

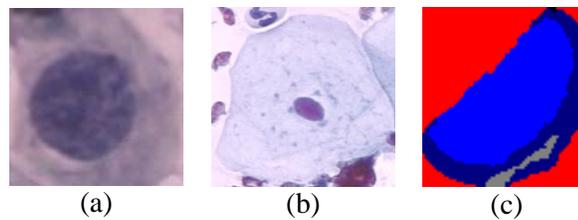


Figura 1. Ejemplos de células de las pruebas de Papanicolaou incluidas en la base de datos Herlev. (a) Célula Anómala, (b) Célula Normal y (c) Célula Segmentada.

En dicha investigación se evalúa la clasificación de las células del cuello uterino, sobre la base de las características extraídas del núcleo. Los experimentos a realizar se basarán en la base de datos Herlev de referencia presentada en (Jantzen, Norup et al. 2005). La base de datos consta de 917 imágenes que contienen cada una una única célula (figura 1), y las muestras se distribuyen irregularmente en siete clases. Tres de ellas son consideradas como normales, y las cuatro restantes como anormales. La descripción detallada de la base de datos se representa en la Tabla 1.

Tabla 1. Distribución de las células en la base de datos Herlev, de imágenes celulares de la prueba de Papanicolaou

NORMAL	# DE CÉLULAS
Epitelio Escamoso Superficial	74
Epitelio Escamoso Intermedio	70
Epitelio Columnar	98
TOTAL	242
ANORMAL	# DE CÉLULAS
Carcinoma in situ	182
Displasia Ligera	146
Displasia Moderada	197
Displasia Severa	150
TOTAL	675

Las imágenes de la base de datos han sido manualmente segmentadas por los expertos, y las áreas del núcleo y el citoplasma se definen con precisión. A partir de estas áreas, fueron determinadas veinte características (Jantzen, Norup et al. 2005) relativas a la intensidad y las características de forma de la región específica tanto del núcleo como de la zona del citoplasma (Tabla 2). Nueve de cada veinte características se refieren a la región del núcleo, los que pueden ser calculados de forma independiente.

Tabla 2. Características extraídas de cada imagen en la base de datos (Jantzen, Norup et al. 2005)

RASGOS DEL CITOPLASMA	RASGOS DEL NÚCLEO
Área	Área
Brillo	Brillo
Diámetro Corto	Diámetro Corto
Diámetro más Largo	Diámetro más Largo
Elongación	Elongación
Redondez	Redondez
Perímetro	Perímetro
Máxima ¹	Máxima ¹
Mínima ¹	Mínima ¹
Posición del Núcleo	
Tamaño Núcleo/Citoplasma	

¹El número de píxeles con el valor de intensidad máximo / mínimo en una zona de 3x3 de la zona específica.

En algunas investigaciones como es el caso de (Plissiti and Nikou, 2012 a) se hace uso de los 9 rasgos del núcleo, debido a la mayor facilidad que presenta esta para su segmentación en imágenes que suelen exhibir un fuerte solapamiento entre los citoplasmas de las células. No obstante en la presente investigación solo se utilizan 5 rasgos de los 9 como *área*, *perímetro*, *diámetro corto*, *diámetro más largo* y *la redondez* debido a que el objetivo principal de la investigación va encaminado a buscar cuál es el clasificador que, a partir de un conjunto simplificado de datos geométricos básicos, mejor se adapta a la clasificación binaria de células cervicales en la prueba de Papanicolaou (Velezmoro and Villafuerte, 2001). La extracción de rasgos es uno de los pasos fundamentales en el procesamiento de imágenes debido a que mientras mejor sea la selección de los atributos más acertada será la clasificación final de las células. Esto hace que la adecuada selección de los rasgos sea una de las limitantes en las investigaciones para la clasificación de imágenes, debido a que si no se cuenta con los rasgos apropiados los resultados obtenidos por el clasificador no tendrán la calidad que se necesita.

Materiales y métodos

Para el caso de la presente investigación, las matrices de rasgos utilizadas fueron extraídas de las imágenes ya previamente segmentadas (Figura 1(c)) pertenecientes a la base de datos Herlev. Para ello se hizo uso de la herramienta Matlab y de las funciones propias de la misma. De esta forma se desarrolló un algoritmo que toma aleatoriamente de la base de datos Herlev un 80% de las imágenes y extrae de ellas los rasgos antes mencionados. A continuación, el algoritmo crea una matriz con los rasgos extraídos de las imágenes, la que será utilizada en el entrenamiento del clasificador. Posteriormente el algoritmo extrae los rasgos del 20% de las imágenes restantes y conforma la matriz de rasgos que se utilizará para realizar las pruebas. Estas matrices tienen como característica fundamental que poseen vectores de rasgos diferentes lo que posibilita una evaluación correcta del funcionamiento del clasificador.

Para esto se utilizó un esquema experimental, validación cruzada de k particiones (*k-fold cross validation*). Se realizó un particionamiento de los datos de la forma antes descrita y se realizaron siete ejecuciones. Se comprobaron los mismos conjuntos de datos para cada clasificador, tanto en las comparaciones internas de estos algoritmos como en la posterior comparación entre los mismos. Para las comparaciones de los resultados en cada algoritmo, de los conjuntos de datos obtenidos en el particionamiento, se utilizaron las tres particiones que mejores resultados mostraron en cuanto a las medidas P_n , área bajo la curva ROC (AUC) y las medias armónicas H y F .

En la presente investigación, se definió de forma heurística utilizar para la comparación y posterior selección, tres algoritmos de clasificación: las máquinas de soporte vectorial (SVM) (Vapnik, 2000), la búsqueda por el vecino más cercano (*kNN*) (Duda, Hart et al. 2012) y las redes neuronales (RNA) (Olabe, 1998).

A través de una comparación realizada por los autores se obtuvieron las características internas con las cuales los algoritmos muestran los mejores resultados en la clasificación de este tipo de células. El algoritmo SVM con el kernel RBF mostró resultados en la clasificación que se comportan entre un 78% y 86% de predictividad negativa (P_n) además de su bajo costo computacional en comparación del kernel lineal que mostró una P_n entre 79% y 85% con un elevado costo computacional. En el caso de las redes neuronales comparadas se evidencia que la RBF Network es más eficiente en el momento de la clasificación comportándose entre un 81 y 91% de P_n en comparación con la red neuronal MLP que mantuvo un comportamiento entre un 81% y 86%, así como en el caso del algoritmo *kNN*, la distancia Manhattan evidenció obtener resultados en la clasificación que se comportan entre un 84 y 93% de P_n con respecto a la distancia mahalanobis que mostró un 84 y 90% de P_n . Para la selección de dichas características se tuvieron en cuenta además los valores obtenidos en cuanto a los índices de efectividad que a continuación también se

utilizan en esta investigación. Dichos algoritmos arrojaron resultados que demuestran la eficiencia de cada uno en la clasificación de las células cervicales.

En la presente investigación se utilizaron, para evaluar la efectividad de los diferentes clasificadores utilizados, los índices de efectividad, conocidos como sensibilidad, especificidad, predictividad positiva y negativa, así como la tasa de clasificación correcta. Otros índices de efectividad empleados en esta investigación para evaluar las distancias en el clasificador fueron el área bajo la curva ROC, la predictividad negativa y las medidas F y H , dadas por:

$$F - measure = \frac{2 \times (Precision \times S_e)}{Precision + S_e} \quad (1)$$

$$H - mean = \frac{2 \times (S_p \times S_e)}{S_p + S_e} \quad (2)$$

Resultados y discusión

A fin de determinar la calidad relativa de los resultados se realizó una comparación entre los clasificadores anteriormente mencionados. Se utilizó una programación propia escrita en lenguaje Java y se usó además las implementaciones disponibles en el paquete de Minería de Datos WEKA.

Análisis de los resultados de acuerdo a la medida F

El objetivo de este estudio es realizar un análisis referente al comportamiento del índice de efectividad conocido como medida F o media armónica (F -measure, en la literatura en inglés). La media armónica determina hasta qué punto los resultados obtenidos se asemejan a los que se hubieran logrado con una categorización manual real por cada iteración de prueba. De acuerdo a los datos ofrecidos en la tabla 3 se muestra que el valor de la medida F se mantuvo por encima del 86% entre los tres clasificadores donde la media prevaleció por encima del 90%. En la corrida de las iteraciones de prueba SVM con el kernel RBF obtuvo un 95.1% de medida F (mayor valor obtenido).

Tabla 3. Comportamiento de la medida F en las iteraciones de pruebas realizadas por clasificador

Clasificadores	Conjuntos de Datos						
	1	2	3	4	5	6	7
<i>SVM (RBF)</i>	0.874	0.883	0.922	0.951	0.934	0.908	0.933
<i>kNN (Manhattan)</i>	0.884	0.877	0.917	0.901	0.939	0.906	0.926
<i>RNA (RBF Network)</i>	0.884	0.882	0.922	0.913	0.906	0.895	0.922

Análisis de los resultados de acuerdo a la Predictividad Negativa (Pn)

El objetivo de este estudio es realizar un análisis referente al comportamiento de la medida de evaluación predictividad negativa (Pn). La Pn es la probabilidad de que el clasificador detecte la célula como normal siendo en realidad el resultado de la prueba diagnóstica negativo. De acuerdo a los datos que se ofrecen en la tabla 4 se muestra que el valor de Pn se mantuvo por encima del 74% entre los tres clasificadores donde la media prevaleció por encima del 80%. En la corrida de las iteraciones de prueba *kNN* con la distancia Manhattan obtuvo un 93% de Pn (mayor valor obtenido).

Tabla 4. Comportamiento de la medida Pn en las iteraciones de pruebas realizadas por clasificador

Clasificadores	Conjuntos de Datos						
	1	2	3	4	5	6	7
<i>SVM (RBF)</i>	0.74	0.80	0.87	0.85	0.86	0.81	0.91
<i>kNN (Manhattan)</i>	0.77	0.79	0.85	0.80	0.88	0.83	0.93
<i>RNA (RBF Network)</i>	0.77	0.81	0.87	0.80	0.78	0.80	0.88

Área bajo la curva ROC (A-ROC o AUC)

El objetivo de este estudio es realizar un análisis referente al comportamiento de la medida de evaluación área bajo la curva ROC (A-ROC o AUC). El AUC según (Bamber, 1975) y (Hanley and McNeil, 1982), se define como la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar de la población, mediante los resultados obtenidos al aplicar la prueba diagnóstica. Los datos ofrecidos en la tabla 5 muestran que el valor de la AUC se mantuvo por encima del 83% entre los tres clasificadores donde la media prevaleció por encima del 91.8%. En la corrida de las iteraciones de prueba *kNN* con la distancia Manhattan obtuvo un 97.6% de AUC (mayor valor obtenido). En la tabla se muestra la estabilidad del clasificador *kNN* por encima de SVM y RBF Network, donde obtuvo los mayores valores del área debajo de la curva en todas las iteraciones de pruebas realizadas.

Tabla 5. Comportamiento de la medida de evaluación AUC en las iteraciones de pruebas realizadas por clasificador

Clasificadores	Conjuntos de Datos						
	1	2	3	4	5	6	7
<i>SVM (RBF)</i>	0.910	0.924	0.938	0.928	0.975	0.902	0.940
<i>kNN (Manhattan)</i>	0.917	0.948	0.951	0.946	0.976	0.925	0.950
<i>RNA (RBF Network)</i>	0.853	0.832	0.893	0.899	0.925	0.860	0.886

Análisis comparativo de los clasificadores

Luego del análisis realizado de acuerdo a las medidas de evaluación aplicadas a los clasificadores es necesario evaluar si existe alguna diferencia significativa que demuestre la superioridad de un clasificador sobre los demás. Para ello se

siguieron las recomendaciones realizadas por Demšar (Demšar, 2006) y las extensiones presentadas por García y Herrera (García and Herrera, 2008). Se procedió a realizar un análisis estadístico a través de pruebas no paramétricas para k muestras no relacionadas mediante el Test de Friedman (Demšar, 2006) con el objetivo de probar la hipótesis nula de que todos los clasificadores alcanzan los mismos resultados en promedio. Como pruebas post-hoc se aplicó primeramente el test Bonferroni-Dunn (Demšar, 2006) para definir que existen diferencias significativas entre el clasificador mejor rankeado y el siguiente. Para contrastar los resultados se aplicó el test de Holm (Demšar, 2006).

Tabla 6. Resultados obtenidos para AUC por los clasificadores

<i>Particiones</i>	<i>SVM</i>	<i>kNN</i>	<i>RNA</i>
1	0.91	0.917	0.853
2	0.924	0.948	0.832
3	0.938	0.951	0.893
4	0.928	0.946	0.899
5	0.975	0.976	0.925
6	0.902	0.925	0.86
7	0.94	0.95	0.886
<i>Ranking</i>	2.0	1.0	2.9
<i>Posición</i>	2	1	3

Tabla 7. Resultados obtenidos para Pn por los clasificadores

<i>Particiones</i>	<i>SVM</i>	<i>kNN</i>	<i>RNA</i>
1	0.74	0.77	0.77
2	0.8	0.79	0.81
3	0.87	0.85	0.87
4	0.85	0.8	0.8
5	0.86	0.88	0.78
6	0.81	0.83	0.8
7	0.91	0.93	0.88
<i>Ranking</i>	1.93	1.86	2.21
<i>Posición</i>	2	1	3

Tabla 8. Resultados obtenidos para medida F por los clasificadores

<i>Particiones</i>	<i>SVM</i>	<i>kNN</i>	<i>RNA</i>
1	0.874	0.884	0.884
2	0.883	0.877	0.882
3	0.922	0.917	0.922
4	0.951	0.901	0.913
5	0.934	0.939	0.906
6	0.908	0.906	0.895
7	0.933	0.926	0.922

Ranking	1.5	2.21	2.28
Posición	1	2	3

Las tablas 6, 7 y 8 muestran los resultados experimentales de los tres clasificadores para las particiones del conjunto de dato realizadas. Las dos últimas filas muestran el ranking de cada uno de los clasificadores y su posición. Después de analizar estadísticamente los resultados para detectar si existen diferencias significativas entre los clasificadores para cada una de las medidas utilizadas, se comprobó que el test de Friedman rechazó la hipótesis nula para AUC con un valor de $p = 9.12 \times 10^{-4}$, mientras que para Pn y la medida F no se encontraron diferencias significativas. Por esta razón solo se aplicaron los test post-hoc para la medida AUC. Los test Bonferroni-Dunn y de Holm rechazaron la hipótesis nula para valores de $p \leq 0.025$ y $p \leq 0.05$ respectivamente, con un valor de confianza $\alpha = 0.05$ (Tabla 9). Estos resultados demuestran que el clasificador *kNN* tiene un mejor desempeño que RNA y equivalente que SVM para el particionamiento de los datos realizados.

Tabla 9. Orden de los clasificadores por el valor de p y ajustando el valor de α para el procedimiento Holm con $\alpha = 0.05$

i	Clasificador	$z = (R_0 - R_i)/SE$	p	α/i
2	RNA	3.74	1.83×10^{-4}	0.025
1	SVM	1.87	0.061	0.05

De acuerdo al análisis realizado y luego de evaluadas las medidas en cada clasificador, se evidencia que el clasificador con mejor desempeño en la clasificación de células cervicales es el *kNN* haciendo uso de la distancia Manhattan. Dicho clasificador presenta los mejores resultados para las medidas evaluadas, con valores promedios de medida *F* de 91%, Pn de 83% y AUC de 94%. Además, este clasificador mostró ser el más estable manteniendo los valores de AUC superiores al 90% y llegando hasta un 97.6% (Tabla 5). Por estas razones se recomienda utilizar el clasificador *kNN* con la distancia Manhattan en la herramienta de visión computacional desarrollada.

Comparación con otros enfoques de la literatura estudiada

La tabla 10 muestra la comparación de la evaluación estadística obtenida durante la investigación con los resultados obtenidos en otras investigaciones analizadas. Dichas investigaciones utilizan al igual que la presente investigación la base de datos Herlev, aunque no especifican la forma en que se realizó la partición de los conjuntos de rasgos para las pruebas y el entrenamiento.

Tabla 10. Comparación de la clasificación en términos de la media armónica H (H -mean)

<i>Clasificadores utilizados por las investigaciones</i>	<i># de Rasgos utilizados</i>	<i>% Media Armónica (H-mean)</i>
Fuzzy C-means (Plissiti and Nikou 2012 a)	7	90.58
Spectral Clustering (Plissiti and Nikou 2012 a)	9	88.77
SVM (RBF) (Plissiti and Nikou 2012 b)	9	96.89
kNN (Manhattan)	5	92.33

Se establece una comparación en función de la media armónica H debido a que es la medida de evaluación utilizada por dichas investigaciones. El resultado obtenido por la presente investigación de acuerdo a la media armónica muestra que se logra superar los resultados obtenidos en (Plissiti and Nikou, 2012 a), mientras que son ligeramente inferiores a (Plissiti and Nikou, 2012 b) aunque se debe tener en cuenta que este utiliza un mayor número de rasgos.

Conclusiones

En el presente artículo se realiza una comparación entre los clasificadores de acuerdo a los resultados obtenidos en cuanto a las medidas de evaluación Exactitud, Medida F, Predictividad Negativa y AUC. Se establece además una comparación entre el clasificador resultante de la validación experimental con resultados obtenidos en investigaciones realizadas por otros autores. A raíz de esto se puede concluir que el clasificador kNN con empleo de la distancia Manhattan ha sido el más estable de los clasificadores evaluados en este trabajo, habiendo obtenido durante las iteraciones de prueba los mayores valores en cuanto al área bajo la curva y la mejor probabilidad de detección correcta de los casos negativos. Cuando se comparó el resultado de la presente investigación con los resultados obtenidos en investigaciones reportadas en la literatura en cuanto a los resultados de la clasificación se evidenció una mejora referente al cumplimiento del objetivo que es la reducción de falsos negativos en la prueba de Papanicolaou. Como dirección de trabajo futuro, se propone introducir otros rasgos y realizar una selección de estos sobre la base de su efectividad, con el propósito de reducir la dimensionalidad de las matrices de rasgos sin afectar significativamente el desempeño de los clasificadores. A más largo plazo, se investigará sobre el proceso de clasificación en varias clases para las imágenes de la prueba de Papanicolaou.

Referencias

ACOSTA, L. F. (2012). El diagnóstico temprano es garantía de vida. Periódico Granma. Ciudad de la Habana.

- BAMBER, D. (1975). "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph." Journal of Mathematical Psychology **12**(4): 387-415.
- DEMŠAR, J. (2006). "Statistical Comparisons of Classifiers over Multiple Data Sets." The Journal of Machine Learning Research **7**: 1-30.
- DUDA, R. O., P. E. HART, et al. (2012). Pattern classification, John Wiley & Sons.
- GARCÍA, S. AND F. HERRERA (2008). "An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons." Journal of Machine Learning Research **9**: 2677-2694.
- HANLEY, J. A. AND B. J. MCNEIL (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve." Radiology **143**(1): 29-36.
- JANTZEN, J., J. NORUP, et al. (2005). "Pap-smear Benchmark Data For Pattern Classification." **9**.
- LORENZO-GINORI, J. V. AND I. RODRÍGUEZ-SANTOS (2012). "Aplicación de técnicas de visión computacional en la prueba de Papanicolaou." Medicentro Electrónica **16**(3): 196-198.
- MARINAKIS, Y., G. DOUNIAS, et al. (2009). "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification." Computers in Biology and Medicine **39**(1): 69-78.
- MARINAKIS, Y., M. MARINAKI, et al. (2008). "Particle swarm optimization for pap-smear diagnosis." Expert Systems with Applications **35**(4): 1645-1656.
- OLABE, XABIER B. (1998). Redes Neuronales Artificiales y sus Aplicaciones. Bilbao, España, Publicaciones de la Escuela de Ingenieros.
- PLISSITI, M. AND C. NIKOU (2012) a. Cervical Cell Classification Based Exclusively on Nucleus Features. Image Analysis and Recognition. A. Campilho and M. Kamel, Springer Berlin Heidelberg. **7325**: 483-490.
- PLISSITI, M. E. AND C. NIKOU (2012) b "On the importance of nucleus features in the classification of cervical cells in Pap smear images." University of Ioannina.
- VAPNIK, V. N. (2000). The Nature of Statistical Learning Theory, Springer.
- VELEZMORO, G. A. B. AND D. F. VILLAFUERTE (2001). Factores de riesgo que pronóstican el hallazgo de citologías cervicales anormales en dos poblaciones: mujeres de obreros de construcción civil vs. mujeres control en la posta médica "Construcción Civil" ESSALUD, de junio a septiembre del 2000. Facultad de Medicina Humana. Lima, Perú, Universidad Nacional Mayor de San Marcos: 67.