

Reconocimiento de Glaucoma usando Imágenes de fondo de la Retina Glaucoma Recognition Using Background Images of the Retina

David Ruíz García¹, Héctor Javier Vásquez Enriquez¹, Saiveth Hernández Hernández¹, Arturo Téllez Velázquez¹, Raúl Cruz Barbosa¹
{ic2013020085, ic2012140086, ps2016200001}@ndikandi.utm.mx
{atellezv, rcruz}@mixteco.utm.mx

¹ Instituto de Computación - Universidad Tecnológica de la Mixteca, Carretera a Acatlima Km. 2.5 69000, Oaxaca, México.

Fecha de recepción: 15 de julio de 2019

Fecha de aceptación: 30 de agosto de 2019

Resumen. En este trabajo se presenta un enfoque de reconocimiento de glaucoma donde la aportación principal es un criterio de selección de características basado en la sensibilidad. El sistema utiliza el conjunto de datos público RIMONE-r3, del cual utiliza las imágenes del fondo de retina y sus segmentaciones de disco y copa óptica para crear un conjunto de 165 características. Se utiliza un método empaquetador para realizar una selección de características hacia adelante, usando sensibilidad como criterio de selección, para reducir el conjunto a un subconjunto de 120 características. Los resultados de la clasificación realizados por una máquina de soporte vectorial son de 92% de exactitud, 100% de sensibilidad y 88.23% de especificidad, logrando un modelo altamente sensible al reconocimiento de glaucoma, el cual mejora a varios trabajos relacionados.

Palabras Clave: Detección de Glaucoma, Extracción de Características, Selección de Características.

Summary. This work presents a glaucoma recognition approach where the main contribution is a sensitivity selection criterion for characteristics. The system uses the public RIMONE-r3 data set, from which it uses images of the retina background and its disc and optical cup segmentations to create a set of 165 features. A packaging method is used to perform a forward feature selection, using sensitivity as the selection criteria, to reduce the set to a subset of 120 features. The results of the classification performed by a vector support machine are 92% accurate, 100% sensitive, and 88.23% specific, achieving a highly sensitive model for glaucoma recognition, which improves several related works.

Keywords: Glaucoma Detection, Feature Extraction, Feature Selection.

1 Introducción

El glaucoma es una enfermedad ocular causada por el aumento de la presión intraocular (PI) debido a la falta de drenaje del humor acuoso, provocando una pérdida progresiva de las fibras nerviosas del nervio óptico [1]. El glaucoma es la segunda causa de ceguera no reversible a nivel mundial, afectando a más de 70 millones de personas, de las cuales la mitad no sabe que lo tiene [2].

Las pruebas para determinar si una persona tiene glaucoma suelen tomar bastante tiempo, además de necesitar personas especializadas y equipos especiales. Por esta razón, es necesario desarrollar técnicas para diagnosticar el glaucoma de manera precisa y rápida.

Este trabajo presenta un enfoque de asistencia al diagnóstico de glaucoma a partir de características de forma y de textura extraídas de imágenes del fondo de la retina y segmentaciones de disco y copa óptica de estas.

2 Estado del arte

Existen varios estudios realizados con el objetivo de desarrollar sistemas computacionales que asistan al diagnóstico de glaucoma. Por ejemplo, en [1] utilizaron la extracción de tres características de forma: relación de copa y disco (CDR, por sus siglas en inglés), índice inferior-superior-nasal-temporal y la distancia de la cabeza del nervio óptico al centro del disco óptico. Para determinar si un paciente tiene glaucoma o no, únicamente estas tres características son calculadas a partir de imágenes del fondo de la retina y usadas para el entrenamiento de una red neuronal artificial de tres capas con función de activación sigmoideal.

El clasificador resultante mostró un porcentaje de precisión de 90.9%, con una especificidad de 80% y sensibilidad de 100%.

En otro trabajo [3], extraen tres tipos de características, como son: a) características color, que incluyen patrones binarios locales (LBP, por sus siglas en inglés), autocorrelogramas y momentos de color; b) características de onda de 7 escalas y 15 orientaciones; y c) CDR. De esta manera, un vector híbrido de

características es obtenido y procesado utilizando Análisis de Componentes Principales para reducir la dimensionalidad del conjunto. El subconjunto de características obtenido es utilizado para entrenar una máquina de soporte vectorial (SVM, por sus siglas en inglés) con kernel lineal.

También en [4] se utiliza una SVM con un kernel lineal entrenada con un total de 35 imágenes (conjunto de entrenamiento) y evaluada sobre otras 15 (conjunto de prueba). En esta investigación, las características ocupadas principalmente son el CDR combinado con características de textura, como LBP y de Haralick. Los resultados obtenidos por estos autores son: a) exactitud 92%, especificidad 93% y sensibilidad 91% utilizando únicamente el CDR; b) exactitud 90%, especificidad 91% y sensibilidad 86% utilizando únicamente características de textura; y c) exactitud 92%, especificidad 88% y sensibilidad de 100% utilizando ambos conjuntos.

Por otro lado, [5] utiliza características en el dominio de la frecuencia. En este trabajo se probaron con distintos clasificadores, de los cuales el mayor porcentaje de precisión fue alcanzado con una red neuronal artificial, obteniendo un 94%. La selección de características fue realizada a través de un algoritmo genético, reduciendo el conjunto a 18 características.

3 Metodología

En esta sección se describe el conjunto de datos y los métodos usados para el preprocesamiento, extracción y selección de características. De igual manera se presentan los clasificadores utilizados.

3.1 Conjunto de datos RIM-ONE Release 3

La tercera versión de RIM-ONE (RIMONE-r3) es un conjunto de datos enfocado al diagnóstico del glaucoma con imágenes estereoscópicas RGB de alta resolución (2144 x 1424 píxeles) del fondo de la retina correspondientes a un ojo (izquierdo o derecho) de cada uno de los 159 diferentes pacientes. Cada imagen tiene segmentaciones manuales (máscaras) de copa y disco realizadas por dos expertos en oftalmología del Hospital Universitario Canarias. Además, cada patrón (paciente) tiene asociado uno de tres posibles diagnósticos: sano, sospecha de glaucoma y glaucoma.

El objetivo de esta investigación es ayudar en el diagnóstico de glaucoma, a partir de características extraídas de imágenes del fondo de la retina. En este sentido, los patrones etiquetados como sospecha de glaucoma son descartados. Como resultado, el conjunto de datos sobre el que se realiza esta investigación es un conjunto desbalanceado de 123 patrones, donde 39 pertenecen a pacientes con glaucoma y el resto a pacientes sanos.

Finalmente, el conjunto se divide estratificadamente en un conjunto de entrenamiento con 67 patrones clasificados como sanos y 31 con glaucoma, y un conjunto de prueba con 17 patrones clasificados como sanos y 8 con glaucoma.

3.2 Medidas de rendimiento

Las medidas de rendimiento presentadas en esta sección son seleccionadas con el objetivo tener un punto de comparación con resultados obtenidos en investigaciones similares [14, 15].

Exactitud y exactitud balanceada

La exactitud es una medida de rendimiento utilizada en problemas de clasificación. Puede ser definida como el porcentaje de aciertos en la predicción de un clasificador respecto a las observaciones de un conjunto de patrones. Sin embargo, esta medida no representa la calidad de las predicciones de un clasificador en conjuntos desbalanceados. En este sentido, [8] recomienda utilizar la exactitud balanceada (BAS, por sus siglas en inglés) como medida de rendimiento en conjuntos desbalanceados.

$$Accuracy = \left(\frac{TP + TN}{P + N} \right) \quad (1)$$

$$BAS = \frac{1}{2} \left(\frac{FP}{P} + \frac{TN}{N} \right) \quad (2)$$

Las ecuaciones (1) y (2) definen exactitud y BAS para problemas de clasificación binaria, respectivamente, donde TP es el número de positivos verdaderos, TN el número de negativos verdaderos, FP el número de positivos

falsos, FN el número de negativos falsos, P el número de observaciones positivas y N el número de observaciones negativas.

Coefficiente de correlación de Matthews

El coeficiente de correlación de Matthews (MCC, por sus siglas en inglés) es una medida para determinar el rendimiento de un clasificador binario cuyas clases en sus observaciones son de tamaños muy distintos, es decir, conjuntos desbalanceados.

De acuerdo con [9], el MCC cuantifica la correlación entre las observaciones y las predicciones en una clasificación binaria. El dominio de esta correlación está acotado por el intervalo $[-1, 1]$, donde 1 representa una predicción perfecta, 0 una predicción aleatoria, y -1 indica total discordancia entre las predicciones y las observaciones.

Tasa de error balanceado

La tasa de error balanceado (BER, por sus siglas en inglés), es el promedio de la proporción de clasificaciones incorrectas para cada clase. Para el caso de clasificación binaria, el BER está definido por la ecuación (3).

$$BER = \frac{1}{2} \left(\frac{FP}{P} + \frac{FN}{N} \right) \quad (3)$$

3.3 Extracción de características y preprocesamiento

A partir de las 123 imágenes originales seleccionadas y apoyándose en las segmentaciones manuales del conjunto de datos, 165 características son extraídas de cada imagen, preprocesadas y categorizadas en cuatro subconjuntos.

El primer subconjunto, de las consideradas por expertos, está conformado únicamente por el CDR. De acuerdo con oftalmólogos [1], este es regularmente utilizado en el diagnóstico del glaucoma. Utilizando las segmentaciones de expertos del RIMONE-r3, el CDR es el resultado de la división del área cubierta por la máscara de copa entre el área cubierta por la máscara de disco. Debido a que la copa está contenida dentro del disco óptico, el rango del CDR está acotado en el intervalo $(0, 1)$.

El segundo subconjunto, de forma, contiene 10 descriptores de forma que caracterizan las segmentaciones de copa y disco. Cinco características son extraídas de cada segmentación: perímetro, área, compacidad, centroide en X y centroide en Y.

El tercer subconjunto, de Haralick, se obtiene calculando las 13 primeras características de Haralick [6] para cada segmentación, utilizando una distancia $1px$ en la segmentación de copa y $3px$ en la segmentación de disco.

El cuarto y último subconjunto, LBP, contiene 20 características extraídas de la segmentación de disco utilizando el operador LBP [11] con un radio $2px$ y puntos 7, y 108 características extraídas de la segmentación de copa utilizando el mismo operador con un radio $2px$ y puntos 10.

El resultado de la extracción aporta 165 características (CDR, 10 de forma, 26 de Haralick y 128 LBP), que son consideradas previo a la selección de características.

De acuerdo con [7], la transformación de los datos, como la normalización, puede mejorar la precisión y eficiencia de algoritmos de aprendizaje supervisado. Estos algoritmos proveen mejores resultados si los datos a ser analizados han sido escalados en intervalos específicos de $[0, 1]$ o de $[-1, 1]$. Con base en las afirmaciones anteriores, previo a los procedimientos de clasificación, un proceso de normalización es empleado para escalar las 165 características extraídas en un rango de $[-1, 1]$.

3.4 Clasificadores

En este trabajo, el mejor clasificador es obtenido al realizar pruebas utilizando *Leave One Out* (LOO) sobre un conjunto de entrenamiento. SVM [12], *Random Forest* (RF), *Ada Boost* (AB) [13], *K-Nearest Neighbors* (KNN) y Bayesiano ingenuo (NB) son considerados en esta investigación.

A través de búsquedas en malla, se determinan las mejores combinaciones de parámetros para los clasificadores SVM, RF, AB y KNN. Considerando las mejores combinaciones de los clasificadores mencionados, se determina el mejor clasificador de acuerdo a su BAS.

3.5 Selección de características

Dos técnicas de selección de características son empleadas y evaluadas sobre los tres mejores clasificadores para determinar la mejor combinación de características: método de *ranking* utilizando Radio Discriminante de Fisher (FDR, por sus siglas en inglés) y selección de características hacia adelante (FFS, por sus siglas en inglés).

FDR Ranking

El FDR describe la separabilidad entre las distribuciones de las características de acuerdo a sus etiquetas asociadas [10]. Para obtener un *ranking* individual de características se calcula el FDR y se evalúa el rendimiento de SVM con las n mejores características de acuerdo al *ranking*, para n desde 1 hasta $t - 1$, siendo t el número total de características.

FFS

En este trabajo, FFS es realizada utilizando un enfoque *wrapper* siguiendo el siguiente procedimiento:

1. Inicia con un conjunto vacío de características SF y un conjunto F que contiene las n características extraídas de RIMONE-r3.
2. Durante $n - 1$ iteraciones:
 - a. Se selecciona la característica f en el conjunto $\{F - SF\}$ que en conjunción con las características en SF brinde la mayor sensibilidad dado un clasificador C .
 - b. f es añadido al conjunto SF .

4 Resultados

Buscando el mejor desempeño en un sistema de ayuda al diagnóstico de glaucoma, se realizó una búsqueda en malla sobre cada clasificador para determinar los parámetros sobre los cuales se obtiene el mejor rendimiento con el conjunto completo de características. La Tabla 1 muestra los mejores parámetros obtenidos, de acuerdo a su exactitud, para los cinco clasificadores considerados. En esta misma tabla, se muestra que SVM con kernel lineal y un C de 0.1, obtiene los mejores resultados de acuerdo a las medidas de exactitud, BER y MCC.

Clasif.	Parámetros	BER	MCC	Exactitud
NB		0.1708	0.6915	0.8699
RF	N.E. ¹ = 50	0.2143	0.5946	0.8293
AB	F.A. ² = 0.05 N.E. = 50	0.217	0.6544	0.8537
SVM	C ³ = 0.1	0.1332	0.7704	0.9024
KNN	k = 7	0.2111	0.6785	0.8618

Tabla 1. Evaluación de clasificadores.

La selección de un subconjunto de características final fue realizada utilizando FFS con un enfoque empaquetador y la exactitud como criterio de selección. Con base a los resultados de la selección del clasificador, se utilizó SVM en este método. El subconjunto de características, A , resultado de este enfoque fueron, en su mayoría, LBP de copa óptica, alcanzando una exactitud máxima de 92.85% sobre el conjunto de entrenamiento utilizando 34 características.

Sin embargo, el subconjunto A presenta dos problemas: a) sensibilidad menor de 80%; y b) no considera a CDR, el cual, de acuerdo al *ranking* FDR, es la mejor característica. Atendiendo esto, heurísticamente se realizó un FFS utilizando CDR como la primera característica seleccionada y la sensibilidad como criterio de selección. El subconjunto de características, B , resultado de este enfoque, alcanzó una exactitud máxima de 94.30% sobre el conjunto de entrenamiento utilizando 119 características. La Tabla 2 muestra el rendimiento obtenido del subconjunto A y B , respectivamente.

¹ Número de estimadores en el ensamble

² Factor de aprendizaje

³ Penalización de la máquina de soporte vectorial

Subconjunto	No. de Características	Exactitud	BAS	BER	MCC
<i>A</i>	86	0.9285	0.8957	0.1042	0.8337
<i>B</i>	119	0.9430	0.9308	0.0691	0.8678

Tabla 2. Resultados del clasificador usando LOO en el conjunto de entrenamiento con el subconjunto de características *A* y *B*.

Con base a las evaluaciones mostradas en la Tabla 1 y 2, un clasificador SVM con $C=0.1$ se entrenó usando el mejor subconjunto de características y fue evaluado sobre el conjunto de prueba. De acuerdo con los resultados obtenidos, mostrados en la Tabla 3, el clasificador obtuvo una exactitud de 92%. Además, aun trabajando sobre un conjunto desbalanceado, el BAS muestra resultados satisfactorios.

Para evaluar los resultados de esta investigación, el rendimiento del modelo utilizado se comparó con otros dos trabajos que utilizaron el mismo conjunto de datos (RIMONE-r3). Como se muestra en la Tabla 3, [14] y [15] obtuvieron un BAS de 79.85% y 83%, respectivamente, mientras que esta investigación obtiene un BAS significativamente superior de 94.11%. Otro resultado relevante de esta tabla es la sensibilidad alcanzada, la cual indica que el modelo obtenido logra predecir correctamente a todos los pacientes con glaucoma. Además, la capacidad para detectar pacientes sanos es mejor que los trabajos relacionados [14] y [15].

Autor	Exactitud	Sensibilidad	Especificidad	BAS
[14]	0.799	0.797	0.800	0.7985
[15]	-	0.83	0.83	0.83
Este trabajo	0.92	1.0	0.8823	0.9411

Tabla 3. Comparación de resultados con otros trabajos.

5 Conclusiones y trabajo futuro

A partir de los resultados obtenidos en la selección de características es posible apreciar que las características de textura empleadas, principalmente sobre la segmentación de copa, complementan satisfactoriamente la información que el CDR provee individualmente. Con esto, el sistema alcanza una exactitud de 92%; además, los resultados medidos por el BAS, BER y MCC muestran un alto rendimiento incluso trabajando sobre un conjunto desbalanceado, en comparación con otros trabajos relacionados.

Lo más relevante de este trabajo es el uso de la sensibilidad como criterio en la selección de características. Este criterio ayuda principalmente en conjuntos desbalanceados donde la cantidad de pacientes sanos es mayor que la de los pacientes con glaucoma.

El presente trabajo puede ser utilizado en conjunción con un sistema de segmentación automática de copa y disco óptico para identificar y reconocer pacientes que podrían padecer de glaucoma.

Referencias

- [1] J. Nayak, P. S. Bhat, N. Shetty and T.-C. Lim. Automated Diagnosis of Glaucoma Using Digital Fundus Images. *Journal of Medical Systems*, 33, 5, pp. 337–346, 2008.
- [2] H. A. Quigley. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, 90, 3, pp. 262–267, 2006.
- [3] A. A. Salam, M. U. Akram, K. Wazir, S. M. Anwar and M. Majid, *Autonomous Glaucoma detection from fundus image using cup to disc ratio and hybrid features*. In proceedings of the 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), (2015), pp. 370-374.
- [4] M. R. K. Mookiah, U. R. Acharya, C. M. Lim, A. Petznick and J. S. Suri. Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowledge-Based Systems*, 33, pp. 73–82, 2012.
- [5] A. Singh, M. K. Dutta, M. Parthasarathi, V. Uher and R. Burget. Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. *Computer Methods and Programs in Biomedicine*, 124, pp. 108–120, 2016.

- [6] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural Feature for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 6, pp. 610-621, 1973.
- [7] L. A. Shalabi, Z. Shaaban and B. Kasasbeh, Data Mining: A Preprocessing Engine. *Journal of Computer Science*, 2, 9, pp. 735–739, 2006.
- [8] V. García, R.A. Mollineda and J. S. Sánchez. Index of balanced accuracy: A performance measure for skewed class distributions. Iberian Conference on Pattern Recognition and Image Analysis. Springer, Berlin, Heidelberg, 2009.
- [9] B. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405, 2, pp. 442–451, 1975.
- [10] Fisher, R. A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7, 2, pp. 179–188. 1936.
- [11] T. Ojala, M. Pietikainen and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 7, 971–987, 2002.
- [12] B. E. Boser, I. M. Guyon and V. N. Vapnik. A training algorithm for optimal margin classifiers. In proceedings of the 5th annual workshop on Computational learning theory - COLT '92, 1992.
- [13] M. Kearns, M. Li, L. Pitt and L. G. Valiant. Recent Results on Boolean Concept Learning. Proceedings of the 4th International Workshop on Machine Learning, pp. 337–352, 1987.
- [14] A. C. D. M. Lima, L. B. Maia, R. M. P. Pereira, G. B. Junior, J. D. S. D. Almeida, and A. C. D. Paiva. Glaucoma Diagnosis over Eye Fundus Image through Deep Features. In proceedings of the 25th International Conference on Systems, Signals and Image Processing (IWSSIP), 2018.
- [15] A. Ramaswamy, K. Ram and M. Sivaprakasam. A Depth Based Approach to Glaucoma Detection Using Retinal Fundus Images. In proceedings of the Ophthalmic Medical Image Analysis Third International Workshop, 2016.