

Miniproyecto 2 - entrega 3

Nicolás Garnica (201713127)
 Universidad de los Andes
 pn.garnica@uniandes.edu.co

César Zapata (201730655)
 Universidad de los Andes
 ca.zapatar@uniandes.edu.co

1. Parte teórica

1.1. Reporte la matriz de confusión, precisión y cobertura promedio del método de prueba, obtenidos a través de sus implementaciones

82	20	21	16	22
20	81	18	26	19
17	16	264	17	19
24	22	19	86	24
188	30	16	14	89

$$\text{Precisión} = \frac{TP}{TP + FP} = 0.56407494 \quad (1)$$

$$\text{Cobertura} = \frac{TP}{TP + FN} = 0.56500063 \quad (2)$$

1.2. Responda: ¿Cuál es la relación entre la precisión y la cobertura?

La relación entre precisión y cobertura tiene que ver con los conceptos estadísticos de error tipo 1 y error tipo 2. El error tipo 1 hace referencia al error que se comete al rechazar la hipótesis nula cuando no debió haber sido rechazada, a lo que también se le llama un *falso positivo*. El error tipo 2, por otro lado, es cometido cuando no se rechaza la hipótesis nula aunque esta debió ser rechazada, también conocido como *falso negativo*. Eso nos lleva a la relación entre la precisión y la cobertura, donde, si por ejemplo se desea aumentar la cobertura, esto hará que aumente la probabilidad de error tipo 2 (o de falso negativo), disminuyendo así la precisión. En pocas palabras, la precisión alude al porcentaje de los resultados que son relevantes, mientras que la cobertura se refiere al porcentaje de los resultados relevantes que son clasificados correctamente, y si alguno de los dos sube, el otro bajará.

1.3. En un problema multiclase: ¿Qué otro nombre recibe la cobertura promedio?

A la cobertura también se le llama sensibilidad, pues indica la capacidad que tiene nuestro algoritmo de discriminar casos positivos de casos negativos.

1.4. Clasificación de tres clases A,B,C

		Anotaciones		
		A	B	C
Predicciones	A	TP	FN	FN
	B	FP	TN	TN
	C	FP	TN	TN

Figure 1. Matriz de confusión centrada en A

		Anotaciones		
		A	B	C
Predicciones	A	TN	FP	TN
	B	FN	TP	FN
	C	TN	FP	TN

Figure 2. Matriz de confusión centrada en B

		Anotaciones		
		A	B	C
Predicciones	A	TN	TN	FP
	B	TN	TN	FP
	C	FN	FN	TP

Figure 3. Matriz de confusión centrada en C

2. Parte biomédica

2.1. Analice los resultados obtenidos. ¿Cuál es el método que da mejores resultados? ¿Por qué cree que este método es mejor?

Dado que el algoritmo desarrollado para el punto teórico, no nos funciona para evaluar nuestros resultados, fue necesario crear un nuevo algoritmo el cual si puedo evaluar la precisión y cobertura de nuestras predicciones. los resultados encontrados se muestran en la figura 4.

METODO UTILIZADO	PRECISIÓN	COBERTURA
RGB - Chi cuadrado:	40.0	40.0
HSV - Chi cuadrado:	60.0	60.0
Lab - Chi cuadrado:	50.0	20.0
RGB - Inter:	37.5	60.0
HSV - Inter:	57.14	80.0
Lab - Inter:	60.0	60.0

Figure 4. Valores de precisión y cobertura para los distintos métodos

de los resultados obtenidos, consideramos que los métodos del espacio de color Lab con una comparación de kernel de intersección y del espacio de color HSV con comparación de Chi - cuadrado, fueron los mejores métodos, pues tuvieron su índice de precisión y cobertura mas alto. en cuanto a cobertura encontramos que el método de espacio de color HSV comparado con kernel de intersección, tiene la cobertura mas alta, de un 0,8. Mientras que la precisión mas alta la encontramos en los dos mejores métodos que seleccionamos.

En el desarrollo de este análisis no tuvimos en cuenta los resultados del método de template matching, debido al corto tiempo que teníamos para el desarrollo. sin embargo, teorizamos que este sería el mejor método de clasificación debido a los resultados de la entrega anterior. esto puede deberse a que el cambio en la coloración producto de una infección impacta mas en la intensidad del tono de gris cuando la imagen esta a blanco y negro, y es mas facil identificar estos cambios de tonalidad. Sin embargo, encontramos que los métodos de análisis de color seleccionados como los mejores, resaltan de una mejor forma el parásito dentro de las células que los demás métodos. la forma de comparación también puede ser un factor determinante.

En la parte de anexos, las figuras 7 y 8, muestran la clasificación de las imágenes, según se planteo con nuestro algoritmo. las predicciones mostradas en la figura 8, se realizaron teniendo en cuenta cual de los valores de comparación con los histogramas sano e infectado era menor o mayor. para el caso de la distancia chi-cuadrado se opto por escoger el número menor y para la distancia de kernel de intersección se escogió el número mayor como se indica en la documentación de la función CompareHist de openCV.

2.2. Defina y muestre la ecuación de la distancia de kernel de intersección y de la distancia Chi-cuadrado.

La ecuación del kernel de intersección suma los minimos comunes que se encuentran entre los dos histogramas, tal como se muestra en la ecuación representada en la figura 1. La ecuación de Chi-cuadrado suma las diferencias al

$$d(H_1, H_2) = \sum_I \min(H_1(I), H_2(I))$$

Figure 5. Distancia kernel de intersección

cuadrado de las intensidades de los histogramas divididas por la intensidad del histograma 1. la suma se realiza las veces que se repita el valor I, tal como se muestra en la ecuación representada en la figura 2.

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}$$

Figure 6. Distancia Chi-cuadrado

2.3. ¿Por qué se usan estas distancias para histogramas y no distancias como la euclidiana?

Se utilizan estas distancias ya que la distancia euclidiana es para distribuciones no discretas como el caso de funciones continuas de distribución de probabilidad, mientras que en el caso no discreto se deben utilizar otro tipo de aproximaciones.

References

Realizado en \LaTeX

Imagen	Tipo de distancia	Espacio de color	Distancia a sano	Distancia a infectado	Prediccion
1	X^2	RGB	32,06	79	sano
1	X^2	HSV	18,3	13,94	infectado
1	X^2	Lab	8831	5418	infectado
2	Interseccion	RGB	7,35	1,12	sano
2	Interseccion	HSV	2,08	1,43	sano
2	Interseccion	Lab	2,24	1,24	sano
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
10	Interseccion	RGB	5,81	1,29	sano
10	Interseccion	HSV	2,24	1,51	sano
10	Interseccion	Lab	1,84	1,35	sano

Figure 7. Tabla de predicciones

```

RGB - Chi cuadrado: ['Sano', 'Sano', 'Infectado', 'Infectado', 'Infectado', 'Sano', 'Infectado', 'Sano', 'Infectado', 'Sano']
HSV - Chi cuadrado: ['Infectado', 'Sano', 'Infectado', 'Sano', 'Infectado', 'Sano', 'Sano', 'Infectado', 'Sano', 'Infectado']
Lab - Chi cuadrado: ['Infectado', 'Sano', 'Infectado', 'Infectado', 'Sano', 'Infectado', 'Infectado', 'Infectado', 'Infectado', 'Sano']
RGB - Inter: ['Sano', 'Sano', 'Sano', 'Sano', 'Sano', 'Sano', 'Infectado', 'Sano', 'Infectado', 'Sano']
HSV - Inter: ['Infectado', 'Sano', 'Sano', 'Sano', 'Infectado', 'Sano', 'Sano', 'Sano', 'Infectado', 'Sano']
Lab - Inter: ['Infectado', 'Sano', 'Sano', 'Sano', 'Infectado', 'Infectado', 'Infectado', 'Sano', 'Infectado', 'Sano']

```

Figure 8. Predicciones realizadas por el algoritmo