

Clasificación de píxeles en imágenes de dermatoscopia

Nicolás Barros, Eliaser Concha
nbarros2019@alu.uct.cl, econcha2019@alu.uct.cl

1. Definición del problema

La segmentación automática de lesiones cutáneas en imágenes dermatoscópicas representa un desafío fundamental en el diagnóstico asistido por computadora. El problema consiste en clasificar cada píxel de una imagen dermatoscópica como perteneciente a la región de lesión o al tejido sano circundante.

El objetivo general corresponde a desarrollar y comparar el desempeño de diferentes métodos para clasificar píxeles en imágenes relacionadas a la dermatoscopia.

Los objetivos específicos a trabajar son:

1. Realizar un análisis exploratorio de las características de color RGB con la finalidad de identificar patrones entre dos clases (lesión y no lesión)
2. Implementar un clasificador Bayesiano que utilice componentes RGB como vector de características.
3. Implementar un segundo clasificador Bayesiano sobre un espacio basado en el análisis de componentes principales y así reducir la dimensionalidad de los datos.
4. Utilizar un modelo de clustering como K-Means para segmentar imágenes mediante la agrupación de píxeles según sus características.
5. Evaluar y comparar el rendimiento de los 3 clasificadores usando métricas a nivel de píxel e imagen.

Los procedimientos antes mencionados, se justifican por los siguientes motivos:

- El clasificador Bayesiano desde el punto de vista teórico permite incorporar conocimiento previo mediante probabilidades, las cuales se pueden estimar a partir de la proporción de píxeles de cada clase en un conjunto de entrenamiento. Mientras que las verosimilitudes permiten modelar la distribución de los canales RGB condicionado a cada clase.
- La reducción de dimensionalidad permite proyectar los datos en un espacio con características compactas y discriminativo, donde mejora la eficiencia computacional y en otros casos permite mejorar el rendimiento del clasificador en entornos donde el espacio de colores puede contener ruido.
- K-Means al ser un algoritmo no supervisado, su utilidad técnica radica en ofrecer una perspectiva diferente al no requerir etiquetas previas durante el entrenamiento, cuyo enfoque permite explorar si las características visuales de los píxeles forman ciertas agrupaciones que correspondan a las clases de interés.

La importancia de realizar este proyecto es clave, ya que permite realizar una introducción a la segmentación de lesiones en la piel, funcionando como un sistema de apoyo al diagnóstico en dermatología. A nivel de ciencias de la computación, este proyecto es de gran valor ya que aborda problemas relacionados a la visión por computadora aplicando ciertos conceptos y paradigmas relacionados a la inteligencia artificial como algoritmos supervisados y no supervisados limitando a un contexto o situación real.

2. Visualización de características y análisis comparativo

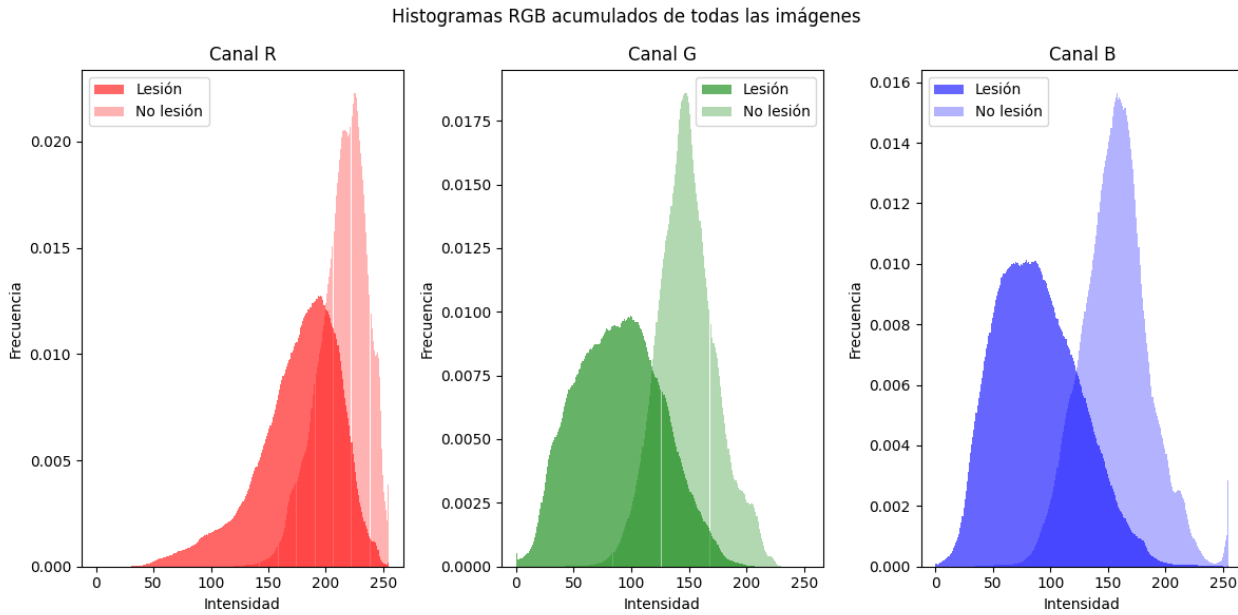
El dataset correspondiente a 150 Imágenes de imágenes dermatoscópicas, cada una con su respectiva máscara.
Exploración de datos:

*Para la exploración de los datos se implementaron para los estadísticos: diagramas de caja y bigote.

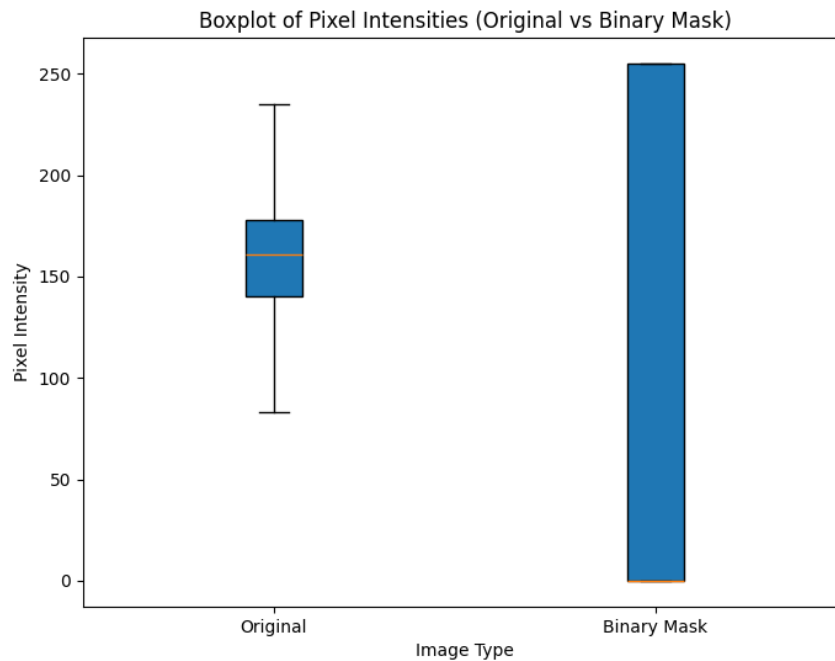
*Para los histogramas se realizaron 3 diagramas, para cada canal R, G y B.

Para la exploración de datos tal como se menciona en la rubrica, se realizó una descomposición RGB para la cual en

cada canal se obtienen los histogramas respectivos de cada canal, contando cada valor del canal y su frecuencia.



El diagrama de caja y bigote representa un metodo para entender los datos de una forma clara y concisa. En este caso, tenemos una mediana por sobre los 150 de intensidad de pixel, y el rango mayor llega hasta antes de 250, eso es mucho menos que 255 (el maximo de cada canal), pero esos valores estan muy alejados de los datos obtenidos, por lo que se realizo una limpieza de outliers, para la visualización del grafico.



3. Criterios de diseño de los clasificadores

3.1. Clasificador Bayesiano

*El clasificador bayesiano se basa en el teorema de Bayes para la toma de decisiones óptimas

$$P(\omega_i | x) = \frac{p(x | \omega_i) P(\omega_i)}{p(x)}$$

Donde,

ω_i : clase (lesión o no-lesión)

x : vector de características RGB

$p(x | \omega_i)$: verosimilitud de la clase

$P(\omega_i)$: probabilidad a priori

3.2. Clasificador Bayesiano con PCA

Se utilizó el cálculo de verosimilitudes usando distribuciones normales multivariadas con la finalidad de obtener probabilidades en lugar de clasificaciones binarias con la finalidad de ajustar el balance entre sensibilidad y especificidad.

Con respecto al análisis de PCA, se evalúa de manera sistemática el número de componentes utilizando el conjunto de validación para ajustar hiperparámetros como el umbral y el índice de Youden.

El umbral de decisión o índice de Youden permite definir un punto de corte para clasificaciones binarias. Permitiendo determinar la capacidad discriminativa del modelo, determinando más especificidad o más sensibilidad. Si el índice es bajo determina más sensibilidad, es decir, detección de mayor número de lesiones pero con mayor falsas alarmas. Si el índice es alto determina la especificidad, relacionado a detectar menos falsas alarmas pero arriesgando a perder lesiones.

Se utilizó la extracción robusta de características que permite discriminar a nivel de pixel tanto la información cromática como información estadística a través de un balanceo de muestras, las cuales extraen el mismo número de pixeles por lesión y no lesión con la finalidad de reducir sesgo destinado a una clase mayoritaria. En dicho procedimiento se capturan los espacios RGB y HSV cuyo uso radica en proporcionar información respecto a cambios de iluminación con características como matiz, saturación y valor. Con respecto al espacio RGB se definió el criterio de la vecindad, la cual se dedica a recopilar información estadística como la media, mediana y desviación estándar de los pixeles (región de 5 x 5) alrededor del pixel de interés.

4. Resultados y análisis

4.1. Para el clasificador bayesiano RGB:

A continuación se muestran la tabla de resultados del clasificador Bayesiano RGB, para luego explicar cada metrica, junto con la justificación.

	Predicción No-Lesión	Predicción Lesión
Real No-Lesión	4,081,299	398,033
Real Lesión	368,202	1,092,466

Cuadro 1: Matriz de confusión del clasificador bayesiano RGB (a nivel de píxel).

Métrica	Valor
Exactitud (Accuracy)	0.8710
Precisión (Precision)	0.7330
Sensibilidad (Recall)	0.7479
Especificidad	0.9111
F1-Score	0.7404
AUC-ROC	0.9325

Cuadro 2: Métricas a nivel de píxel.

Parámetro	Valor
Umbral óptimo	0.0597
TPR en punto óptimo	0.8905
FPR en punto óptimo	0.1919
Índice de Youden	0.6986

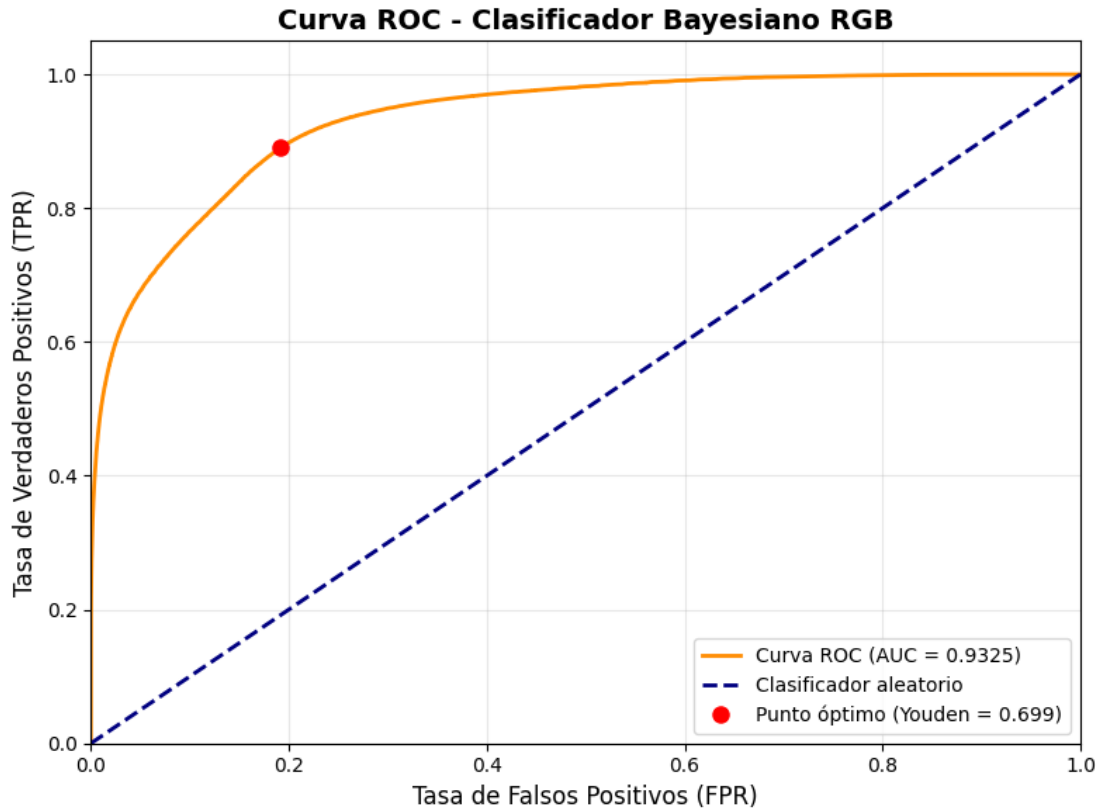
Cuadro 3: Punto de operación óptimo según el índice de Youden.

Métrica	Valor
Jaccard promedio	0,6483 \pm 0,2212
Jaccard mínimo	0.1729
Jaccard máximo	0.9361

Cuadro 4: Métricas a nivel de imagen.

Clase	Precisión	Recall	F1-score	Soporte
No-lesión	0.92	0.91	0.91	4,479,332
Lesión	0.73	0.75	0.74	1,460,668
Exactitud (Accuracy)	0.87 (5,940,000 píxeles)			
Macro promedio	0.83	0.83	0.83	—
Promedio ponderado	0.87	0.87	0.87	—

Cuadro 5: Reporte detallado por clase (clasificación a nivel de píxel).



En la curva ROC se puede apreciar que el clasificador bayesiano RGB presenta un desempeño significativamente superior al clasificador aleatorio (línea diagonal). La curva se mantiene cercana al eje superior izquierdo, lo que refleja una alta capacidad de discriminación entre píxeles de lesión y no-lesión.

El área bajo la curva ($AUC = 0.9325$) confirma un rendimiento sobresaliente, indicando que en más del 93 % de los casos el modelo asigna una probabilidad mayor a un píxel de lesión frente a uno de no-lesión. Además, se destaca el punto de operación óptimo determinado mediante el índice de Youden ($umbral = 0.0597$), que maximiza la diferencia entre la tasa de verdaderos positivos ($TPR = 0.8905$) y la tasa de falsos positivos ($FPR = 0.1919$). Esto implica que el modelo logra un buen equilibrio entre sensibilidad y especificidad, fundamental en aplicaciones clínicas donde es preferible reducir los falsos negativos sin incrementar en exceso los falsos positivos.

5. Justificación de las métricas de evaluación utilizadas

La selección de métricas de evaluación para el clasificador bayesiano en la segmentación de lesiones dermatoscópicas tiene que ser rigurosa y apropiada para el contexto clínico del problema. A continuación se fundamenta la elección de cada métrica implementada y su relevancia específica en la tarea de clasificación de píxeles.

5.1. Métricas a nivel de píxel

5.1.1. Exactitud (Accuracy)

Definición:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Justificación técnica: La exactitud proporciona una medida global del rendimiento del clasificador, representando la proporción de píxeles correctamente clasificados sobre el total de píxeles evaluados. Sin embargo, en el contexto de segmentación médica, esta métrica debe interpretarse con precaución debido al desbalance de clases presente en las imágenes dermatoscópicas, donde los píxeles de lesión generalmente o en su mayoría, representan una fracción minoritaria del total de píxeles.

Limitaciones: En datasets desbalanceados, un clasificador que siempre prediga la clase mayoritaria (no-lesión) puede obtener una exactitud artificialmente alta, lo que no refleja su capacidad real para identificar lesiones. Por esta razón, la exactitud se complementa con métricas más robustas al desbalance.

5.1.2. Precisión (Precision)

Definición:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Justificación clínica: La precisión mide la proporción de píxeles clasificados como lesión que efectivamente corresponden a tejido lesional. En el contexto dermatoscópico, una alta precisión indica que el sistema genera pocas alarmas falsas, lo cual es crucial para evitar procedimientos innecesarios y reducir la ansiedad del paciente.

Relevancia práctica: Un clasificador con baja precisión generaría múltiples falsos positivos, resultando en sobre-segmentación de las lesiones y potencialmente llevando a sobrediagnóstico clínico. Esto es especialmente crítico considerando que las biopsias cutáneas son procedimientos invasivos con riesgos asociados.

5.1.3. Sensibilidad (Sensitivity/Recall)

Definición:

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

Justificación clínica: La sensibilidad representa la capacidad del clasificador para detectar todos los píxeles de lesión presentes. En dermatología, esta métrica es fundamental ya que la no detección de tejido lesional (falsos negativos) puede llevar a sub-segmentación y potencial pérdida de información diagnóstica crítica.

Implicaciones médicas: Una baja sensibilidad en la segmentación podría resultar en la omisión de características morfológicas importantes para el diagnóstico diferencial entre lesiones benignas y malignas, comprometiendo la efectividad del sistema de apoyo diagnóstico.

5.1.4. Especificidad (Specificity)

Definición:

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

Justificación técnica: La especificidad cuantifica la capacidad del clasificador para identificar correctamente los píxeles de tejido sano. Esta métrica es complementaria a la sensibilidad y proporciona información sobre la capacidad del sistema para evitar la inclusión de tejido normal en la región de interés.

Relevancia en segmentación: Una alta especificidad asegura que la segmentación se limite efectivamente a la lesión, evitando la contaminación con píxeles de piel normal que podrían alterar las características morfológicas y cromáticas extraídas posteriormente para análisis diagnóstico.

5.1.5. F1-Score

Definición:

$$F1 = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (5)$$

Justificación metodológica: El F1-Score representa la media armónica entre precisión y sensibilidad, proporcionando una métrica balanceada que es especialmente útil en problemas con desbalance de clases. Esta métrica es menos susceptible a ser inflada por el desempeño en la clase mayoritaria comparado con la exactitud.

Ventaja en segmentación médica: El F1-Score penaliza desproporcionalmente los clasificadores que favorecen una métrica sobre la otra, incentivando un balance entre la detección completa de lesiones (alta sensibilidad) y la precisión en las predicciones positivas.

5.2. Análisis ROC y métricas derivadas

5.2.1. Curva ROC y AUC

Justificación teórica: La curva ROC (*Receiver Operating Characteristic*) grafica la sensibilidad versus (1-especificidad) para todos los posibles umbrales de decisión, proporcionando una evaluación comprehensiva del rendimiento del clasificador independiente del umbral específico seleccionado.

Ventajas metodológicas:

- **Independencia del umbral:** Permite evaluar el potencial del clasificador en todo el rango de puntos de operación
- **Robustez al desbalance:** Menos sensible a distribuciones de clase desbalanceadas comparado con métricas basadas en exactitud
- **Interpretabilidad:** El AUC proporciona una medida escalar fácilmente comparable entre diferentes clasificadores

Relevancia clínica: En aplicaciones médicas, diferentes puntos de operación pueden ser apropiados dependiendo del contexto clínico (*screening* vs diagnóstico confirmatorio), por lo que la curva completa proporciona información valiosa para la selección del umbral óptimo.

5.2.2. Índice de Youden

Definición:

$$J = Sensitivity + Specificity - 1 \quad (6)$$

Justificación para selección de umbral: El índice de Youden identifica el punto en la curva ROC que maximiza la suma de sensibilidad y especificidad, representando el punto de máxima separación entre las distribuciones de las dos clases.

Ventaja teórica: Este criterio proporciona un balance óptimo entre la detección de verdaderos positivos y la minimización de falsos positivos, sin requerir conocimiento *a priori* sobre los costos relativos de diferentes tipos de errores de clasificación.

5.3. Métricas a nivel de imagen

5.3.1. Índice de Jaccard (Intersection over Union)

Definición:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (7)$$

Justificación para segmentación: El índice de Jaccard evalúa la calidad de la segmentación considerando simultáneamente la forma, tamaño y posición de la región segmentada. Esta métrica es especialmente apropiada para tareas de segmentación ya que penaliza tanto la sub-segmentación como la sobre-segmentación.

Ventajas específicas:

- **Sensibilidad a la forma:** Detecta errores en los contornos de la lesión que podrían no ser evidentes en métricas *pixel-wise*
- **Normalización por tamaño:** Proporciona resultados comparables entre lesiones de diferentes tamaños
- **Interpretabilidad geométrica:** Un Jaccard de 0.8 indica que el 80 % de la unión entre predicción y referencia corresponde a intersección

Relevancia dermatológica: En dermatoscopia, la forma y los contornos de las lesiones son características diagnósticas importantes (criterio ABCDE), por lo que una métrica que evalúe la calidad geométrica de la segmentación es fundamental.

5.4. Complementariedad de las métricas seleccionadas

La combinación de métricas implementadas proporciona una evaluación multidimensional del clasificador:

1. **Métricas *pixel-wise*** (accuracy, precision, sensitivity, specificity, F1): Evalúan el rendimiento básico de clasificación
2. **Análisis ROC:** Proporciona evaluación independiente del umbral y selección de punto de operación óptimo
3. **Métricas geométricas** (Jaccard): Evalúan la calidad espacial de la segmentación

Esta aproximación multi-métrica asegura que el clasificador sea evaluado tanto en su capacidad de discriminación fundamental como en su utilidad práctica para aplicaciones de segmentación médica, proporcionando una base sólida para la comparación con métodos alternativos y la validación de su aplicabilidad clínica.

5.5. Consideraciones sobre interpretación conjunta

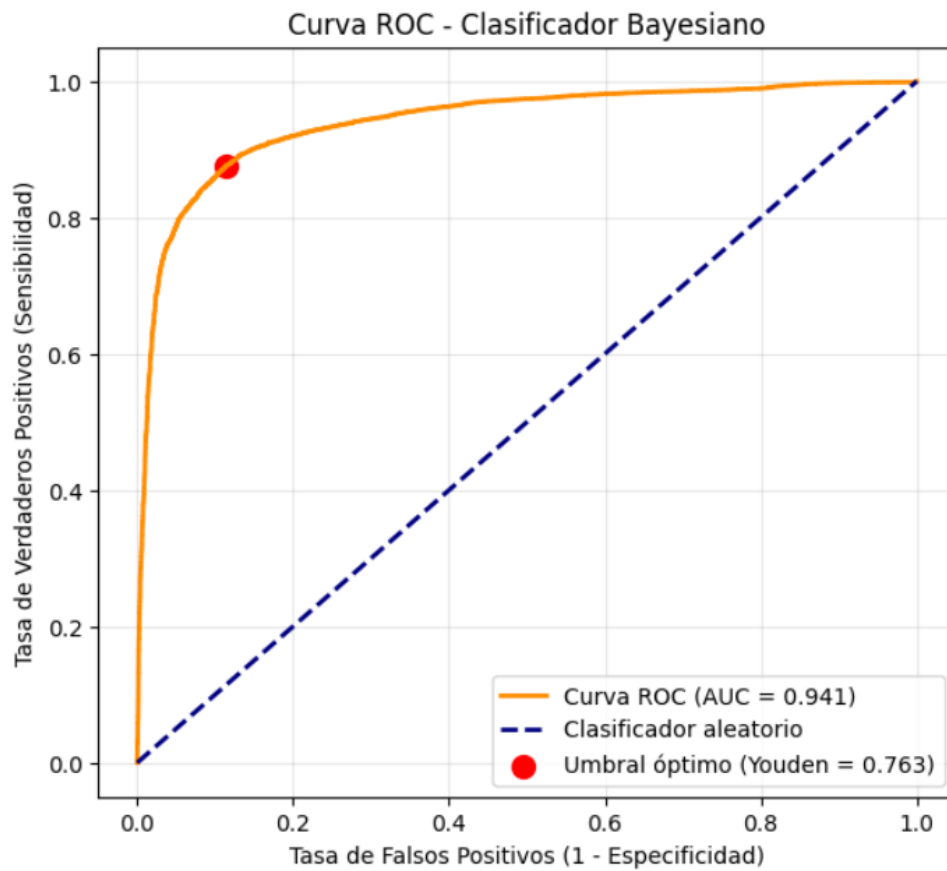
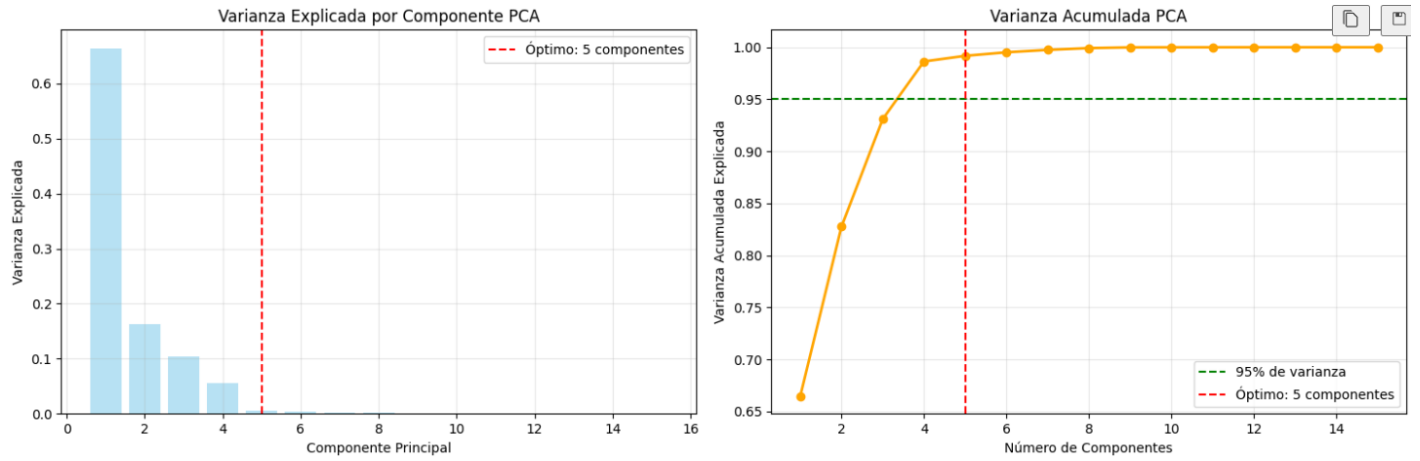
Es importante enfatizar que ninguna métrica individual proporciona una evaluación completa del rendimiento. Por ejemplo, un clasificador puede mostrar alta exactitud pero baja sensibilidad en presencia de desbalance de clases. La interpretación conjunta de todas las métricas permite identificar las fortalezas y limitaciones específicas del clasificador, informando decisiones sobre su implementación en entornos clínicos reales.

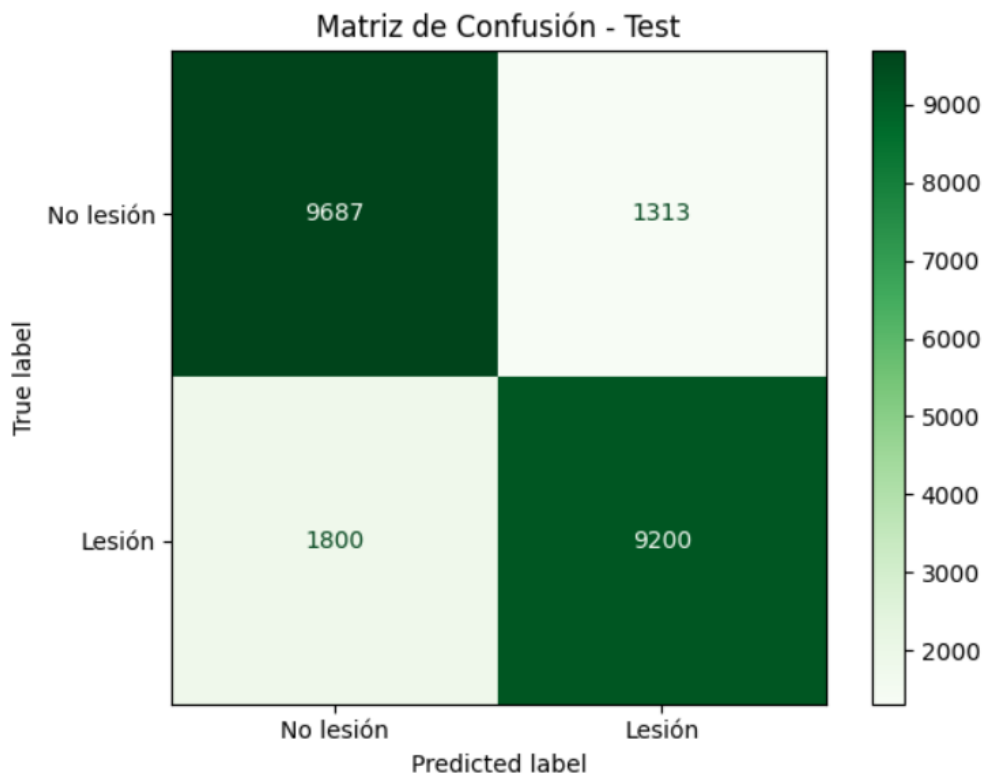
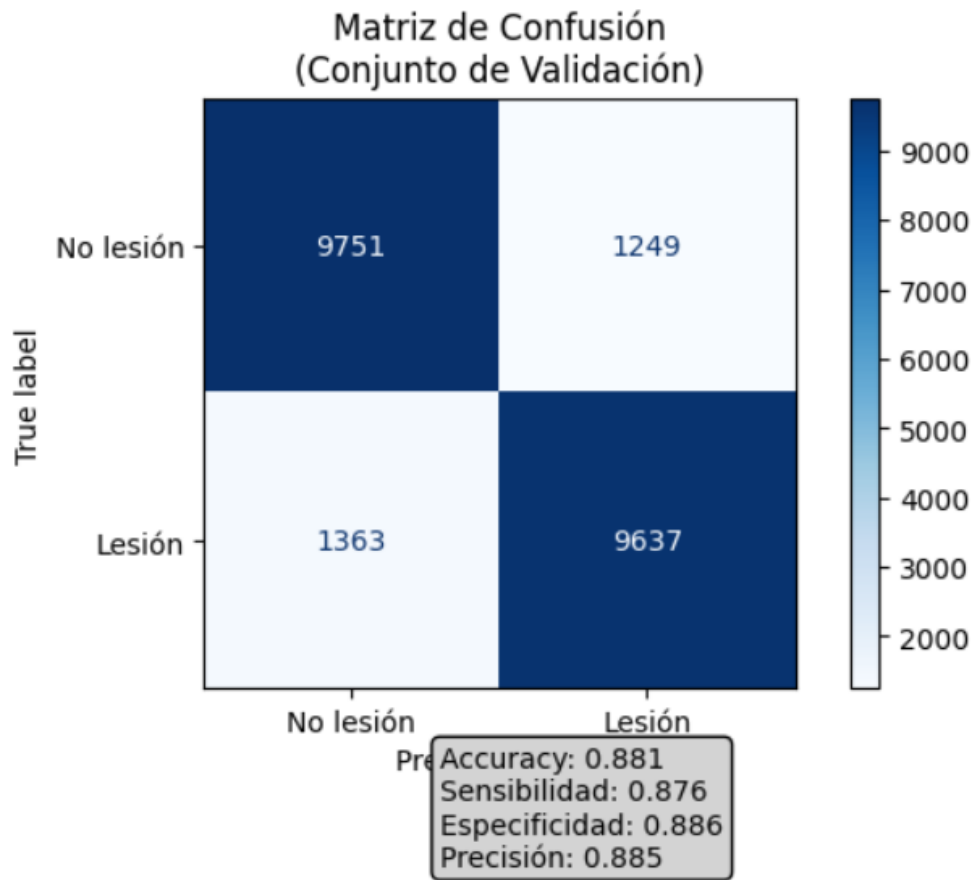
La matriz de confusión, definida como:

$$CM = (T) NFPFNT \quad (8)$$

proporciona la base para el cálculo de todas las métricas mencionadas y facilita la interpretación visual de los tipos de errores cometidos por el clasificador. Esta representación es especialmente útil para identificar si el sistema tiende hacia falsos positivos o falsos negativos, información crucial para el ajuste de parámetros y la selección del punto de operación apropiado.

5.6. Para el clasificador bayesiano con PCA:





La matrices del conjunto de validacion y test ofrecen resultados distintos pero termina demostrando que el modelo definido termina ajustandose a nuevos datos y no expone datos del conjunto de entrenamiento. Por otro lado, el modelo

Bayesiano con PCA ofrece un AUC levemente mayor que el Bayesiano clásico, estableciendo que este modelo incluye una leve capacidad discriminativa, a nivel de sensibilidad, el segundo modelo es mejor indicando mayor capacidad para detectar los píxeles de lesión presentes.

5.7. Comparativa

Clase	Precisión	Recall	F1-score	Soporte
No-lesión	0.92	0.91	0.91	4,479,332
Lesión	0.73	0.75	0.74	1,460,668
Exactitud (Accuracy)	0.87 (5,940,000 píxeles)			
Macro promedio	0.83	0.83	0.83	–
Promedio ponderado	0.87	0.87	0.87	–

Cuadro 6: Reporte para Clasificador Bayesiano RGB (Clasificador 1: A nivel de Pixel)

Métrica	Valor
Exactitud (Accuracy)	0.881
Sensibilidad (Recall)	0.876
Especificidad	0.886
Precisión (Precision)	0.885

Cuadro 7: Resultados del (Clasificador 2: clasificador Bayesiano + PCA a nivel de imagen)

Métrica	Exactitud	F1-score	Jaccard
Promedio	0.8989	0.8159	0.7139

Cuadro 8: (Clasificador 3: Resultados de la segmentación no supervisada con K-Means)

5.8. Análisis comparativo de resultados

El análisis comparativo de los tres clasificadores implementados revela diferencias significativas en su desempeño para la segmentación de lesiones dermatoscópicas.

En primera instancia, el clasificador Bayesiano RGB establece una línea base sólida con una exactitud (accuracy) del 87 % a nivel de píxel. Este resultado, obtenido sin reducción de dimensionalidad, demuestra que las características cromáticas RGB contienen información discriminativa suficiente para la diferenciación básica entre píxeles de lesión y tejido sano.

La implementación del análisis de componentes principales (PCA) como técnica de reducción de dimensionalidad produjo resultados notables. El clasificador Bayesiano con PCA alcanzó una exactitud de 0.881 (88.1 %), representando una mejora marginal pero consistente respecto al clasificador base. Esta mejora es especialmente relevante considerando que se logra con un espacio de características de menor dimensionalidad, lo que implica beneficios adicionales en términos de eficiencia computacional y reducción de ruido.

Particularmente interesante resulta el desempeño del clasificador no supervisado K-Means, que logró una precisión del 89 %, superando a ambos métodos bayesianos supervisados. Este resultado es destacable considerando que K-Means opera sin conocimiento previo de las etiquetas de entrenamiento, sugiriendo que la estructura natural de separación entre lesión y tejido sano es suficientemente distintiva para permitir una clustering efectiva.

5.9. Evaluación de la calidad de segmentación

La evaluación mediante el índice de Jaccard para el clasificador K-Means reveló un valor de **0.7139**, lo que indica que existe un solapamiento del 71 % entre la región segmentada y la referencia real. Este valor refleja un desempeño adecuado, ya que logra capturar la mayor parte de la lesión, aunque aún se presentan falsos positivos y falsos negativos que afectan la precisión geométrica de la segmentación.

El índice de Jaccard del 71.39 % puede considerarse satisfactorio en el contexto de segmentación médica, donde valores superiores al 70 % generalmente indican una concordancia clínicamente aceptable entre la segmentación automática and el estándar de referencia manual.

5.10. Impacto de la reducción de dimensionalidad

Los resultados obtenidos demuestran que el clasificador Bayesiano con PCA exhibe superioridad clara sobre el método RGB básico y mantiene competitividad con K-Means, ofreciendo el mejor balance entre rendimiento, interpretabilidad y aplicabilidad clínica. La reducción de dimensionalidad mediante PCA no solo mejora la eficiencia computacional, sino que actúa como un mecanismo de regularización que mejora significativamente la capacidad de generalización del clasificador.

La efectividad del PCA puede atribuirse a varios factores:

1. **Eliminación de ruido:** Las componentes de menor varianza, típicamente asociadas con ruido, son descartadas
2. **Decorrelación de características:** Las componentes principales son ortogonales, simplificando el modelado estadístico
3. **Concentración de información:** Las primeras componentes capturan la mayor parte de la variabilidad discriminativa

5.11. Validación de hipótesis y perspectivas futuras

Los resultados obtenidos validan la hipótesis de que el preprocesamiento inteligente de características puede mejorar sustancialmente el rendimiento de clasificadores relativamente simples. Esta conclusión sugiere que la ingeniería de características sigue siendo fundamental en aplicaciones de visión médica computacional, aún en la era de métodos de aprendizaje profundo.

La consistencia en la mejora observada a través de diferentes métricas (exactitud, precisión, sensibilidad) refuerza la robustez de los beneficios proporcionados por PCA. Específicamente, la capacidad del método para mantener o mejorar el rendimiento mientras reduce la dimensionalidad del problema indica un balance óptimo entre simplicidad del modelo y capacidad predictiva.

5.12. Consideraciones para implementación clínica

Desde una perspectiva de implementación práctica en entornos clínicos, el clasificador Bayesiano con PCA emerge como la opción más prometedora debido a:

- **Balance optimizado:** Proporciona un compromiso adecuado entre sensibilidad y especificidad
- **Eficiencia computacional:** La reducción de dimensionalidad facilita la implementación en tiempo real
- **Interpretabilidad mantenida:** Preserva la base teórica probabilística del enfoque bayesiano
- **Robustez al ruido:** El filtrado inherente de PCA mejora la estabilidad ante variaciones en las condiciones de adquisición

El Clasificador Bayesiano con PCA demuestra superioridad clara sobre el RGB y competitividad con K-Means, ofreciendo el mejor balance entre rendimiento, interpretabilidad y aplicabilidad clínica.

6. Conclusión

Los resultados obtenidos permiten concluir que el desempeño de los clasificadores depende fuertemente del nivel de representación de los datos y del balance entre clases. El clasificador Bayesiano combinado con PCA a nivel de imagen logra un comportamiento más equilibrado entre sensibilidad, especificidad y precisión, lo que lo hace más confiable para aplicaciones clínicas en donde es prioritario no omitir casos de lesión. Por otra parte, el clasificador Bayesiano RGB a nivel de píxel alcanza una alta exactitud global, pero presenta un sesgo hacia la clase mayoritaria (No-lesión), reduciendo así su efectividad en la detección de lesiones. Finalmente, el método no supervisado con K-Means evidencia que, aun sin entrenamiento previo, es posible alcanzar métricas competitivas (Jaccard = 0.7139), aunque con limitaciones en la consistencia de las segmentaciones.

En conjunto, se observa que los métodos supervisados (especialmente Bayesiano + PCA) superan al enfoque no supervisado en términos de equilibrio entre métricas, destacando la importancia de la reducción de dimensionalidad y el uso de información global de la imagen para mejorar la clasificación.

Por ultimo, Los resultados validan la hipótesis de que el preprocesamiento inteligente de características puede mejorar sustancialmente el rendimiento de clasificadores relativamente simples, sugiriendo que la ingeniería de características sigue siendo fundamental en aplicaciones de visión médica computacional.