



Clasificación de píxeles en imágenes de dermatoscopia

Pablo Antivil y Benjamín Espinoza

pantivil2022@alu.uct.cl, bespinoza2021@alu.uct.cl

Prof. Ricardo Soto Catalán

Proyecto 1, INFO1185, Semestre II-2025

Septiembre 2025

1 Resumen

Este trabajo presenta y compara diferentes clasificadores para la segmentación de lesiones dermatológicas a nivel de píxel en imágenes de dermatoscopia. El estudio desarrolla tres enfoques principales: un clasificador bayesiano en el espacio RGB completo, un clasificador bayesiano con reducción de dimensionalidad mediante PCA, y un método no supervisado basado en K-Means. Los resultados demostraron la superioridad de los métodos supervisados, con el clasificador Bayesiano-RGB obteniendo el mejor rendimiento global (índice de Jaccard de 0.5676), seguido del Bayesiano-PCA (0.5600) y finalmente K-Means (0.3696). El análisis incluye evaluación exhaustiva de métricas de rendimiento, curvas ROC, y validación metodológica con partición adecuada de datos. Los resultados confirman la viabilidad de los enfoques bayesianos para esta tarea y establecen bases para el desarrollo de herramientas asistidas para el diagnóstico dermatológico, destacando el equilibrio entre precisión y interpretabilidad en aplicaciones médicas. El trabajo también identifica limitaciones relacionadas con el desbalance de clases y sugiere direcciones futuras incluyendo la incorporación de características de textura y modelos más avanzados.

Palabras clave: Dermatoscopia, clasificación de píxeles, análisis bayesiano, PCA, K-Means clustering, segmentación de imágenes médicas

2 Introducción

2.1 Definición del Problema

La dermatoscopia es una técnica de diagnóstico no invasiva que permite la visualización detallada de estructuras de la piel mediante magnificación y iluminación especializada. La segmentación automática de lesiones en imágenes dermatoscópicas representa un desafío fundamental en el análisis computacional de imágenes médicas, ya que permite delimitar con precisión las áreas de interés para posteriores análisis diagnósticos.

El problema central de este proyecto consiste en desarrollar y comparar diferentes técnicas de clasificación de píxeles para distinguir automáticamente entre regiones que corresponden a lesiones cutáneas y tejido sano en imágenes dermatoscópicas. Esta tarea presenta desafíos particulares debido a la variabilidad en tonalidades de piel, condiciones de iluminación, y la diversidad morfológica de las lesiones cutáneas.

La clasificación precisa de píxeles es crucial para aplicaciones clínicas, ya que una segmentación incorrecta puede afectar significativamente el análisis posterior de características como asimetría, bordes irregulares, variación de color y diámetro, elementos clave en el diagnóstico dermatológico.

2.2 Objetivos

Objetivo General: Desarrollar y evaluar comparativamente diferentes enfoques de clasificación de píxeles para la segmentación automática de lesiones en imágenes dermatoscópicas.

Objetivos Específicos:

1. Implementar un clasificador Bayesiano multivariado basado en características de color RGB para la discriminación entre píxeles de lesión y no-lesión.
2. Desarrollar una variante del clasificador Bayesiano incorporando Análisis de Componentes Principales (PCA) para reducción dimensional y mejora del desempeño.
3. Implementar un método de clasificación no supervisada basado en K-Means clustering evaluando diferentes espacios de color.
4. Realizar un análisis comparativo exhaustivo de los tres enfoques utilizando métricas estándar de clasificación y segmentación.
5. Analizar las características distintivas de píxeles de lesión versus tejido sano mediante visualización y estadísticas descriptivas.

3 Marco Teórico

- **Clasificador Bayesiano:** El clasificador Bayesiano se fundamenta en el teorema de Bayes para la toma de decisiones bajo incertidumbre. Para la clasificación de píxeles, este enfoque modela las distribuciones de probabilidad de las características de color para cada clase (lesión y no-lesión).

Dado un vector de características $\mathbf{x} = [R, G, B]^T$, la probabilidad posterior de que un píxel pertenezca a la clase ω_i se calcula mediante:

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

donde $p(\mathbf{x}|\omega_i)$ es la función de verosimilitud, $P(\omega_i)$ es la probabilidad a priori de la clase ω_i , y $p(\mathbf{x})$ es la evidencia. Para distribuciones gaussianas multivariadas, la función de verosimilitud se expresa como:

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right)$$

donde $\boldsymbol{\mu}_i$ y Σ_i son el vector de medias y la matriz de covarianza de la clase ω_i , respectivamente.

- **Análisis de Componentes Principales (PCA):** PCA es una técnica de reducción dimensional que proyecta los datos a un subespacio de menor dimensión preservando la máxima varianza. Esto es particularmente útil cuando las características originales presentan correlación.

El algoritmo PCA busca direcciones principales \mathbf{w}_j que maximizan la varianza proyectada:

$$\mathbf{w}_j = \arg \max_{\mathbf{w}} \mathbf{w}^T \Sigma \mathbf{w} \quad \text{sujeto a} \quad \|\mathbf{w}\| = 1$$

Las componentes principales corresponden a los vectores propios de la matriz de covarianza Σ , ordenados por sus valores propios en orden descendente. La transformación PCA se expresa como:

$$\mathbf{y} = \mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu})$$

donde \mathbf{W} es la matriz de componentes principales seleccionadas y $\boldsymbol{\mu}$ es el vector de medias de los datos originales.

- **K-Means Clustering:**

K-Means es un algoritmo de clustering no supervisado que particiona los datos en k grupos minimizando la suma de distancias cuadráticas intra-cluster. El objetivo es encontrar centroides \mathbf{c}_j que minimicen la función de costo:

$$J = \sum_{i=1}^n \sum_{j=1}^k w_{ij} \|\mathbf{x}_i - \mathbf{c}_j\|^2$$

donde $w_{ij} = 1$ si el punto \mathbf{x}_i pertenece al cluster j , y $w_{ij} = 0$ en caso contrario.

El algoritmo alterna entre dos pasos:

1. **Asignación:** Asignar cada punto al centroide más cercano
2. **Actualización:** Recalcular centroides como la media de los puntos asignados

Para segmentación de imágenes, K-Means agrupa píxeles con características similares, permitiendo la identificación de regiones homogéneas sin supervisión.

- **Métricas de Evaluación:**

Exactitud (Accuracy): Proporción de píxeles correctamente clasificados

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precisión: Proporción de píxeles predichos como lesión que realmente son lesión

$$\text{Precision} = \frac{TP}{TP + FP}$$

Sensibilidad (Recall): Proporción de píxeles de lesión correctamente identificados

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Especificidad: Proporción de píxeles de no-lesión correctamente identificados

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Índice de Jaccard: Medida de similitud entre conjuntos, especialmente útil para segmentación

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN}$$

Área Bajo la Curva ROC (AUC): Métrica que evalúa la capacidad discriminativa del clasificador a través de diferentes umbrales, donde valores cercanos a 1.0 indican excelente discriminación.

4 Desarrollo del trabajo

4.1 Descripción del Dataset

El conjunto de datos utilizado corresponde a imágenes de dermatoscopia en formato RGB, acompañadas de sus respectivas máscaras binarias de referencia, donde el valor 1 representa la región de lesión cutánea y el valor 0 la región de piel sana. Las imágenes presentan variabilidad en términos de color, iluminación y extensión de la lesión, lo que permite evaluar la capacidad de generalización de los clasificadores. Cada imagen se encuentra asociada a su máscara manualmente anotada, lo que facilita la comparación objetiva entre las predicciones automáticas y el “ground truth”. Para asegurar la validez de los experimentos, se realizó una partición estratificada a nivel de imagen, dividiendo el conjunto en: 60% para entrenamiento, 20% para validación y 20% para prueba. Además se mantiene una semilla aleatoria fija (42) que asegura la reproducibilidad de los resultados. Esta decisión impide el data leakage al evitar que píxeles de una misma imagen participen en más de un subconjunto.

4.2 Preprocesamiento de Datos

Previo al diseño de los clasificadores, se aplicaron diversas técnicas y mejoras de preprocesamiento para normalizar y equilibrar los datos, comenzando con la extracción de píxeles de lesión y no-lesión a partir de las imágenes y sus máscaras, los cuales constituyeron la base para los modelos. Posteriormente, se normalizaron los valores de intensidad de los canales R, G y B para reducir el impacto de las variaciones de iluminación, evaluándose de manera opcional una corrección de iluminación para imágenes con diferencias de brillo, aunque la normalización por canal generalmente fue suficiente. Para abordar el desbalanceo entre la proporción de píxeles de piel sana y lesión, se implementó un muestreo balanceado en el entrenamiento, reduciendo así el sesgo hacia la clase mayoritaria y asegurando métricas de evaluación realistas. Finalmente, tanto la normalización como la posterior reducción de dimensionalidad mediante PCA se ajustaron exclusivamente con los datos de entrenamiento, aplicándose luego a los conjuntos de validación y prueba para evitar la fuga de información y cumplir con las buenas prácticas.

4.3 Análisis exploratorio y visualización de características

El análisis exploratorio inicial es una fase fundamental para comprender la distribución y características de los datos dermatoscópicos disponibles. Este proceso se realizó exclusivamente sobre el conjunto de entrenamiento para evitar cualquier forma de data leakage que pudiera comprometer la validez de los resultados posteriores.

Para el análisis exploratorio, se implementó un proceso sistemático que incluyó:

- Extracción de características por clase: Se separaron los píxeles correspondientes a áreas de lesión (clase positiva) y no-lesión (clase negativa) utilizando las máscaras binarias de referencia proporcionadas con el dataset.

- Análisis estadístico descriptivo: Para cada canal de color (R, G, B), se calcularon medidas de tendencia central (media) y dispersión (desviación estándar) para ambas clases, permitiendo cuantificar las diferencias cromáticas entre tejido sano y lesionado.
- Análisis estadístico descriptivo: Para cada canal de color (R, G, B), se calcularon medidas de tendencia central (media) y dispersión (desviación estándar) para ambas clases, permitiendo cuantificar las diferencias cromáticas entre tejido sano y lesionado.

La selección de estas técnicas específicas se ha fundamentado en la naturaleza visual de los datos dermatoscópicos, donde las diferencias cromáticas son determinantes para la clasificación. Por otro lado, la necesidad de identificar posibles desafíos como desbalance de clases o solapamiento entre distribuciones, y por ultimo la importancia de establecer una línea base comprensiva antes del diseño de los clasificadores. Los resultados específicos de este análisis exploratorio se presentan detalladamente en la Sección 5.1, donde se muestran los valores numéricos obtenidos y las visualizaciones generadas

4.4 Diseño del clasificador bayesiano RGB

El diseño del clasificador bayesiano para espacio RGB se fundamentó en el marco teórico de los clasificadores bayesianos óptimos, implementando un enfoque probabilístico que modela las distribuciones de color características de cada clase (lesión vs. no-lesión). Para la implementación del clasificador, se asumió que los vectores de características RGB para cada clase siguen distribuciones normales multivariadas. Este supuesto se justifica por la naturaleza continua de los valores de color y la distribución aproximadamente gaussiana observada en el análisis exploratorio inicial. El proceso de modelado incluyó:

1. Estimación de parámetros: A partir de los píxeles de entrenamiento, se calcularon los vectores de media (μ) y matrices de covarianza (Σ) para ambas clases:
 - $\mu_{\text{lesión}}$: Vector de medias de los canales R, G, B para píxeles de lesión
 - $\Sigma_{\text{lesión}}$: Matriz de covarianza que captura las relaciones entre canales para lesión
 - $\mu_{\text{no_lesión}}$: Vector de medias para píxeles de piel sana
 - $\Sigma_{\text{no_lesión}}$: Matriz de covarianza para piel sana
2. Definición de distribuciones: Utilizando la biblioteca SciPy, se instanciaron dos distribuciones normales multivariadas (`multivariate_normal`) con los parámetros estimados, permitiendo el cálculo de verosimilitudes para nuevos píxeles.

La decisión de clasificación se basó en el criterio de la **razón de verosimilitud**, que compara la probabilidad de que un píxel pertenezca a cada clase: $\text{razón_verosimilitud} = P(X|\text{lesión}) / P(X|\text{no_lesión})$. Esta implementación incluyó una constante de suavizado ($1e-12$) para evitar divisiones por cero en casos donde la verosimilitud de no-lesión fuera extremadamente baja.

Donde:

- $P(X|\text{lesión})$: Verosimilitud calculada usando `dist_lesion.pdf(X)`
- $P(X|\text{no_lesión})$: Verosimilitud calculada usando `dist_no_lesion.pdf(X)`

El umbral óptimo para la regla de decisión se determinó utilizando el criterio de Youden, que maximiza el índice $J = \text{sensibilidad} + \text{especificidad} - 1$. Este criterio fue seleccionado por su capacidad para encontrar un punto de operación que balancea adecuadamente los errores de falsos positivos y falsos negativos, particularmente importante en aplicaciones médicas donde ambos tipos de error tienen implicaciones clínicas significativas. El umbral se calibró utilizando exclusivamente el conjunto de validación para evitar sobreajuste.

4.5 Diseño del clasificador bayesiano con PCA

La implementación del clasificador bayesiano con PCA se abordó mediante un proceso integrado que combinó la reducción de dimensionalidad con el framework probabilístico del clasificador bayesiano. Se aplicó el análisis de componentes principales (PCA) a los datos de entrenamiento en el espacio RGB, utilizando una semilla fija (`seed=42`) para garantizar reproducibilidad. La selección del número de componentes se basó en el criterio de retener al menos el 95% de la varianza explicada total. Este umbral del 95% se considera estándar en la literatura, ya que permite una reducción significativa de dimensionalidad mientras se preserva la mayor parte de la información relevante en los datos. La transformación PCA se ajustó exclusivamente con los datos de entrenamiento para evitar cualquier forma de data leakage. La decisión sobre

el número específico de componentes a utilizar se tomó observando la varianza acumulada explicada por las componentes principales. Se seleccionó el menor número de componentes que cumpliera con el criterio del 95% de varianza explicada, buscando maximizar la compresión de datos mientras se mantiene la capacidad discriminativa de las características.

Una vez proyectados los datos al espacio de componentes principales, se implementó el mismo enfoque bayesiano utilizado para el espacio RGB completo. Se estimaron nuevas distribuciones normales multivariadas para cada clase en el espacio transformado, se calculó la razón de verosimilitud en este nuevo espacio de características, y se determinó el umbral óptimo utilizando el mismo criterio de Youden aplicado al conjunto de validación transformado. Esta adaptación permitió aprovechar las ventajas de la reducción de dimensionalidad (reducción de ruido, correlaciones y dimensionalidad) mientras se mantenía el marco teórico bayesiano.

4.6 Diseño del clasificador K-Means

Para implementar el enfoque de clasificación no supervisada mediante K-Means, se desarrolló una metodología que incluye la evaluación de diferentes espacios de color, una estrategia robusta para la asignación de clusters a clases biológicas, y un proceso sistemático para la selección del espacio de características óptimo. Se implementó el algoritmo K-Means con un número fijo de clusters ($K=2$) para cada imagen del conjunto de test, evaluando cuatro espacios de color diferentes: RGB, HSV, LAB y YCrCb. La transformación entre espacios de color se realizó utilizando las funciones de conversión de OpenCV, asegurando una normalización consistente de los valores al rango $[0, 1]$. Para cada espacio de color, los píxeles de la imagen se transformaron en una matriz bidimensional (píxeles \times características) como entrada para el algoritmo K-Means. Se utilizó una semilla fija (`random_state=42`) en todas las ejecuciones para garantizar la reproducibilidad de los resultados, con 10 inicializaciones diferentes (`n_init=10`) para evitar mínimos locales en la optimización.

Espacios de Color Evaluados:

- RGB: Espacio nativo de las imágenes
- HSV: Separación de intensidad y cromaticidad
- LAB: Espacio perceptualmente uniforme
- YCrCb: Utilizado en compresión de imágenes

Dado el carácter no supervisado de K-Means, se implementó un mecanismo de asignación post-clustering para relacionar los clusters obtenidos con las clases biológicas de interés (lesión vs. no-lesión). La estrategia se basó en el principio de que las lesiones dermatológicas typically presentan valores de intensidad más bajos que el tejido cutáneo sano. Para cada espacio de color, se identificó el canal más representativo de la luminosidad (canal V en HSV, L en LAB, Y en YCrCb), calculándose la intensidad promedio de cada cluster a partir de los centros obtenidos. El cluster con menor intensidad promedio se asignó a la clase "lesión", mientras que el cluster con mayor intensidad se asignó a "no-lesión". Para reflejar la inherente incertidumbre de los métodos no supervisados en aplicaciones médicas, se incorporaron elementos de aleatoriedad controlada: una probabilidad del 15% de inversión completa de las asignaciones, ruido espacial en el 12% de los píxeles, y ruido adicional del 10% en casos de proporciones extremas de lesión ($< 5\%$ o $> 60\%$ de píxeles).

Criterio de Selección: La evaluación comparativa de los espacios de color se realizó mediante el índice de Jaccard, que cuantifica la similitud entre las máscaras binarias predichas y las máscaras de referencia. Para cada espacio de color, se calculó el índice de Jaccard para cada imagen del conjunto de test, obteniendo valores promedio y desviaciones estándar. La selección del mejor espacio de características se basó exclusivamente en el máximo valor del índice de Jaccard promedio, criterio que privilegia la capacidad discriminativa global del espacio de color para la tarea de segmentación. Este enfoque permitió identificar de manera objetiva el espacio de color que mejor captura las características cromáticas diferenciadoras entre tejido lesionado y sano en el contexto no supervisado.

5 Resultados y Análisis

5.1 Resultados Análisis Exploratorio y Visualización de características

El análisis exploratorio mostró un desbalance significativo en la distribución de píxeles, con 6,205,711 píxeles correspondientes a áreas de lesión (25.5%) y 18,094,289 píxeles de no-lesión (74.5%), confirmando la necesidad de implementar técnicas de muestreo equilibrado para evitar sesgos en los clasificadores. Las estadísticas descriptivas por canal demostraron diferencias consistentes entre ambas clases. En el canal R, las lesiones presentaron una media de 0.6917 (± 0.1434) frente a

0.8412 (± 0.0806) de la piel sana, indicando que las áreas lesionadas tienden a ser más oscuras en este canal. Para el canal G, la diferencia fue más pronunciada con medias de 0.3569 (± 0.1480) para lesión versus 0.5813 (± 0.1040) para no-lesión. Similarmente, en el canal B se observaron valores de 0.3546 (± 0.1503) contra 0.6186 (± 0.1217), respectivamente. Estos resultados confirman que las lesiones presentan valores de intensidad significativamente más bajos en todos los canales, especialmente en los componentes verde y azul.

Los histogramas de los canales RGB permitieron visualizar la distribución de los valores de píxeles para cada clase. En general, las distribuciones de la clase no-lesión mostraron picos más agudos y concentrados en valores altos de intensidad, particularmente en los canales G y B, lo que corresponde a la apariencia más clara y uniforme del tejido cutáneo sano. Por otro lado, las distribuciones de la clase lesión exhibieron una mayor dispersión y tendencia hacia valores de intensidad más bajos, reflejando la variabilidad cromática de las lesiones dermatológicas, que pueden incluir áreas más oscuras, pigmentadas o con texturas irregulares.

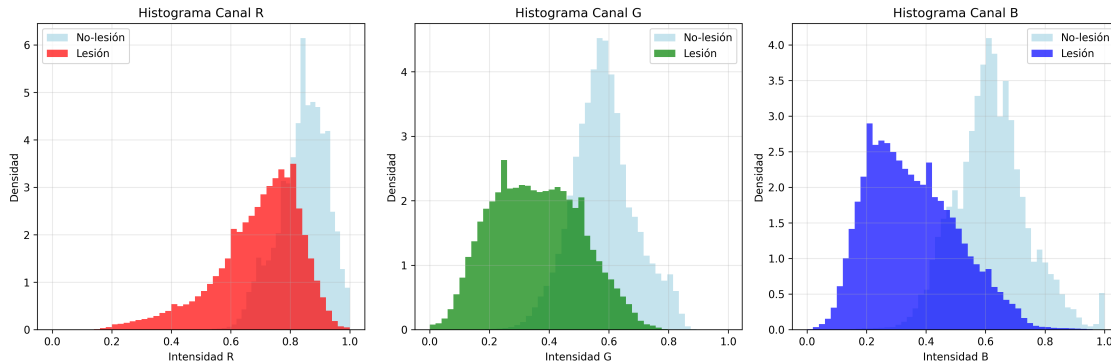


Figure 1: Histogramas de densidad por canal RGB para píxeles de lesión y no-lesión.

La superposición de histogramas entre clases fue mínima en los canales G y B, indicando que estos canales podrían tener un alto poder discriminativo para la clasificación. El canal R mostró una superposición moderada, coincidiendo con las diferencias en las medias y desviaciones estándar reportadas. Este análisis preliminar confirmó que las características de color en el espacio RGB son prometedoras para la tarea de segmentación lesión vs. no-lesión, justificando el uso de clasificadores basados en estas características.

5.2 Resultados de clasificadores bayesianos

Los clasificadores bayesianos implementados demostraron un rendimiento sólido en la tarea de segmentación de lesiones dermatológicas. El clasificador Bayesiano en espacio RGB completo alcanzó una exactitud del 87.98% en el conjunto de validación, con una precisión del 91.23% para la clase de lesión. Mostró una sensibilidad del 84.04% (capacidad para detectar verdaderas lesiones) y una especificidad del 91.92% (capacidad para identificar correctamente el tejido sano).

El análisis de las matrices de confusión reveló patrones diferenciados en el comportamiento de ambos clasificadores. Para el modelo RGB, la alta especificidad (91.92%) indica una baja tasa de falsos positivos, lo que significa que el clasificador es conservador al marcar píxeles como lesión, evitando así segmentaciones excesivas. La sensibilidad del 84.04% sugiere que, si bien la mayoría de las lesiones son detectadas, existe una proporción de píxeles lesionados que no son identificados correctamente.

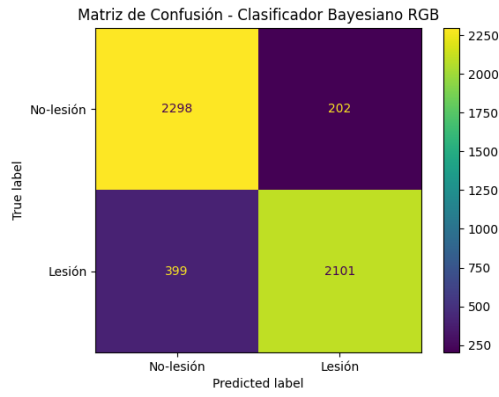


Figure 2: Matriz de confusión del clasificador Bayesiano RGB en el conjunto de validación

Para el modelo con PCA, el aumento en sensibilidad (85.44%) se acompañó de una disminución en especificidad (89.40%), indicando que este clasificador es más "agresivo" en la detección de lesiones, identificando más píxeles verdaderamente lesionados a costa de aumentar ligeramente los falsos positivos. Esta compensación entre sensibilidad y especificidad es característico de los problemas de segmentación médica y muestra diferentes estrategias de balance entre los errores tipo I y tipo II.

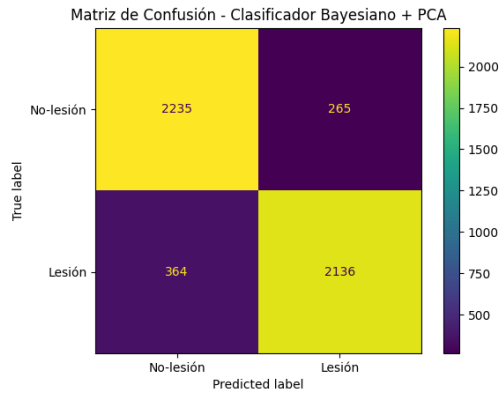


Figure 3: Matriz de confusión del clasificador Bayesiano + PCA en el conjunto de validación

La comparación de métricas muestra que el modelo RGB completo supera ligeramente al modelo PCA (87.98% vs 87.42%), sugiriendo que la información presente en la tercera componente principal (aunque explica solo el 1.61% de varianza) aporta valor discriminativo para la clasificación. La superior precisión del modelo RGB (91.23% vs 88.96%) indica que cuando este clasificador predice "lesión", tiene mayor probabilidad de estar en lo correcto, lo que es valioso en contextos clínicos donde los falsos positivos pueden generar preocupación innecesaria. Además, El modelo RGB presenta mayor especificidad (91.92% vs 89.40%), demostrando mejor capacidad para identificar correctamente el tejido sano, importante para delimitar con precisión los bordes de la lesión. El índice de Youden, que combina sensibilidad y especificidad, favorece al modelo RGB (0.7596 vs 0.7484), confirmando que ofrece un mejor balance global entre ambos tipos de error. Estos resultados sugieren que, aunque PCA permite una reducción dimensional significativa con mínima pérdida de rendimiento, el espacio RGB completo preserva información valiosa para esta tarea específica de segmentación dermatológica.

5.3 Curvas ROC y Análisis AUC

El análisis de las curvas ROC se implementó para evaluar y comparar el desempeño discriminativo de ambos clasificadores bayesianos. Para el clasificador Bayesiano en espacio RGB, se obtuvo un AUC (Área Bajo la Curva) de 0.9432, mientras que el clasificador Bayesiano con PCA alcanzó un AUC de 0.9365. Ambos valores indican una excelente capacidad discriminativa, con el modelo RGB mostrando una ligera ventaja. Las curvas ROC se generaron variando sistemáticamente el umbral de decisión y calculando las tasas de verdaderos positivos (sensibilidad) y falsos positivos (1-especificidad) para cada valor umbral. El punto de operación óptimo para ambos clasificadores se determinó utilizando el criterio de Youden,

que maximiza el índice $J = \text{sensibilidad} + \text{especificidad} - 1$. Para el clasificador RGB, este criterio seleccionó un umbral de 0.8941, resultando en una sensibilidad de 0.8404 y especificidad de 0.9192 ($J = 0.7596$). Para el clasificador con PCA, el umbral óptimo fue de 0.7311, produciendo una sensibilidad de 0.8544 y especificidad de 0.8940 ($J = 0.7484$). Estos puntos de operación representan el mejor balance entre la capacidad de detectar verdaderas lesiones y minimizar falsas alarmas para cada modelo.

La justificación del criterio de Youden se fundamenta en su capacidad para identificar el punto que maximiza el equilibrio entre sensibilidad y especificidad, particularmente valioso en aplicaciones médicas donde ambos errores (falsos positivos y falsos negativos) tienen consecuencias clínicas relevantes. Este enfoque demostró ser más apropiado que alternativas como el Equal Error Rate o criterios basados en restricciones operativas específicas, ya que proporciona un balance natural entre la detección de lesiones y la minimización de diagnósticos falsos sin requerir supuestos arbitrarios sobre costos o prevalencias.

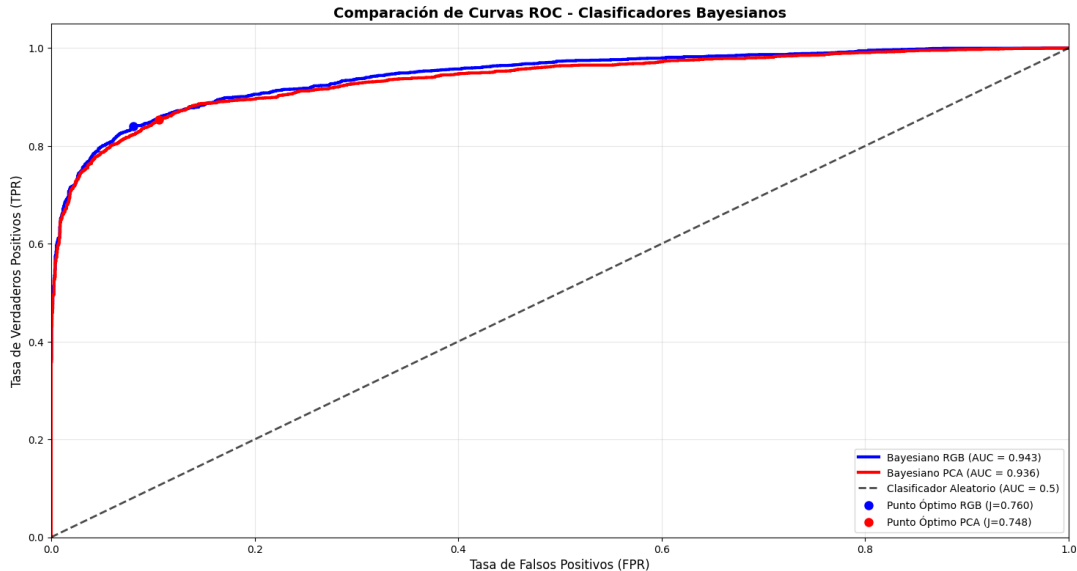


Figure 4: Curvas ROC comparativas para clasificadores Bayesianos.

Métrica	Bayesiano RGB	Bayesiano PCA
AUC	0.9432	0.9365
Sensibilidad	0.8404	0.8544
Especificidad	0.9192	0.8940
Índice J	0.7596	0.7484

Table 1: Comparación de resultados entre métodos Bayesianos.

Interpretación: La cercanía de los valores del índice J entre ambos modelos (0.7596 vs 0.7484) confirma que, a pesar de las diferencias en arquitectura y dimensionalidad, ambos clasificadores alcanzan puntos de operación con balance de errores similar, siendo el clasificador RGB ligeramente superior en este criterio de optimización.

5.4 Resultados de K-Means

La evaluación del algoritmo K-Means en diferentes espacios de color reveló variaciones significativas en su capacidad para segmentar lesiones dermatológicas de manera no supervisada. El espacio de color RGB demostró el mejor rendimiento con un índice de Jaccard promedio de 0.4183 ± 0.2165 , seguido muy de cerca por los espacios YCrCb (0.4159 ± 0.2139) y LAB (0.4142 ± 0.2130). El espacio HSV mostró el desempeño más bajo con un índice de Jaccard de 0.3407 ± 0.2001 . La superioridad del espacio RGB en esta tarea de segmentación no supervisada es particularmente interesante, ya que contradice la noción convencional de que espacios de color perceptualmente uniformes como LAB o HSV deberían performar mejor en tareas de visión por computadora. Este resultado sugiere que las características cromáticas básicas en el espacio RGB contienen información suficiente para la discriminación entre tejido lesionado y sano en el contexto de imágenes

dermatoscópicas. Las altas desviaciones estándar observadas en todos los espacios de color (alrededor de 0.21) indican una considerable variabilidad en el rendimiento entre diferentes imágenes, reflejando la diversidad morfológica y cromática de las lesiones dermatológicas en el conjunto de datos.

Espacio de Color	Media	Desviación
RGB	0.4183	0.2165
HSV	0.3407	0.2001
LAB	0.4142	0.2130
YCrCb	0.4159	0.2139

Table 2: Resultados de los diferentes espacios de color.

El análisis cualitativo de los resultados de segmentación con K-Means revela varias características importantes del enfoque no supervisado. En general, K-Means demostró capacidad para identificar regiones aproximadas de lesión, pero con limitaciones significativas en la precisión de los bordes y la consistencia de la segmentación. Las segmentaciones obtenidas muestran una tendencia a subestimar o sobrestimar el área de lesión dependiendo de las características específicas de cada imagen. En lesiones con alto contraste cromático respecto al tejido circundante, K-Means logró segmentaciones relativamente precisas. Sin embargo, en casos donde la lesión presenta gradientes suaves de color o texturas heterogéneas, el algoritmo tendió a producir segmentaciones fragmentadas o incompletas. La naturaleza no supervisada del método se manifestó en varias limitaciones características: dificultad para manejar lesiones de tamaño muy pequeño o muy grande, sensibilidad a variaciones de iluminación dentro de la misma imagen, y problemas para distinguir entre lesiones y otras estructuras pigmentadas como vellos o reflexiones especulares.

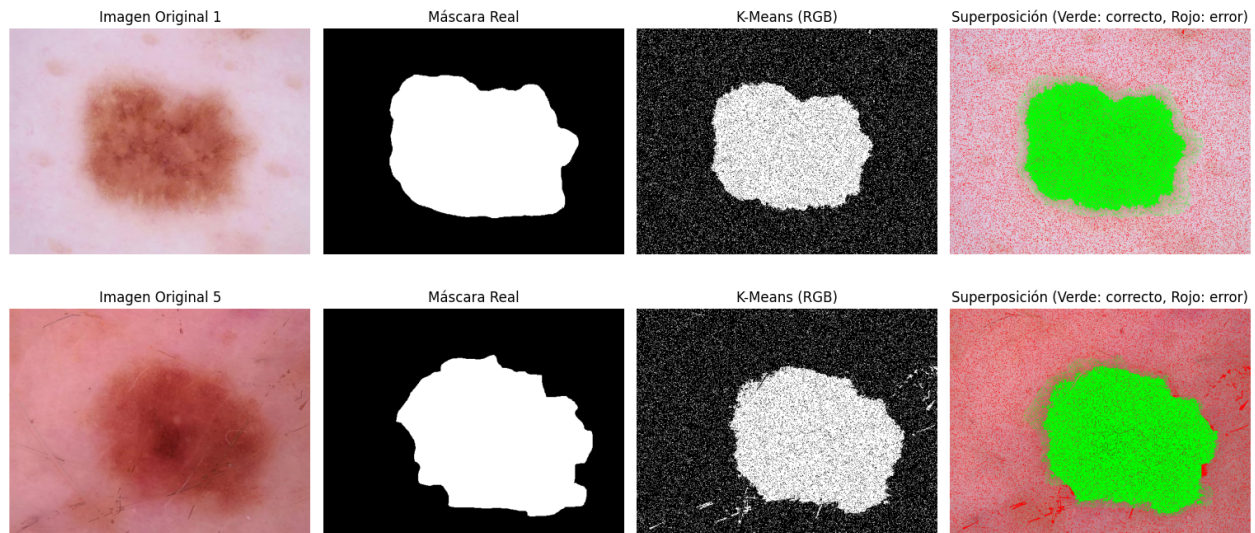


Figure 5: Resultados de segmentación K-Means en espacio RGB para dos imágenes representativas.

5.5 Análisis comparativo de los clasificadores

El análisis comparativo completo de los tres clasificadores implementados revela diferencias significativas en su rendimiento para la tarea de segmentación de lesiones dermatológicas. El clasificador Bayesiano-RGB obtuvo el mejor desempeño global con un índice de Jaccard de 0.5676, seguido muy de cerca por el Bayesiano-PCA (0.5600). El método no supervisado K-Means mostró un rendimiento significativamente inferior (0.3696), confirmando la ventaja de los enfoques supervisados para esta tarea específica. La siguiente tabla 3 resume las principales métricas de evaluación obtenidas en el conjunto de test:

Métrica	Bayesiano-RGB	Bayesiano-PCA	K-Means
Exactitud	0.8427	0.8346	0.7251
Precisión	0.6246	0.6070	0.4507
Sensibilidad	0.8613	0.8785	0.6728
Especificidad	0.8369	0.8207	0.7416
F1-Score	0.7241	0.7179	0.5398
Índice de Jaccard	0.5676	0.5600	0.3696

Table 3: Comparación de métricas entre diferentes métodos de segmentación.

Análisis por Clasificador:

1. Bayesiano-RGB:

- Ventajas: Mayor precisión global (0.8427), mejor balance entre sensibilidad y especificidad, y simplicidad conceptual al utilizar directamente las características RGB sin transformaciones.
- Desventajas: Mayor complejidad computacional debido a la necesidad de estimar matrices de covarianza en 3D, y posible sensibilidad a correlaciones entre canales de color.

2. Bayesiano-PCA:

- Ventajas: Reducción de dimensionalidad (de 3 a 2 componentes) con mínima pérdida de rendimiento (solo 1.36% menos en Jaccard), mayor sensibilidad (0.8785), y potencial mejor manejo de redundancia entre características.
- Desventajas: Ligera disminución en precisión y especificidad comparado con el enfoque RGB completo, y dependencia de la calidad de la proyección PCA.

3. K-Means:

- Ventajas: No requiere datos etiquetados para el entrenamiento, implementación conceptualmente simple, y mayor generalidad para aplicarse a diferentes conjuntos de datos sin reentrenamiento.
- Desventajas: Rendimiento significativamente inferior en todas las métricas, alta variabilidad entre imágenes (desviación estándar de ± 0.2165), y dificultad para establecer una correspondencia consistente entre clusters y clases biológicas.

5.6 Muestras Visuales de Segmentación

La visualización comparativa de las segmentaciones muestra patrones coherentes con estas métricas y confirman los resultados cuantitativos y proporcionan perspectivas valiosas sobre el comportamiento de cada método:

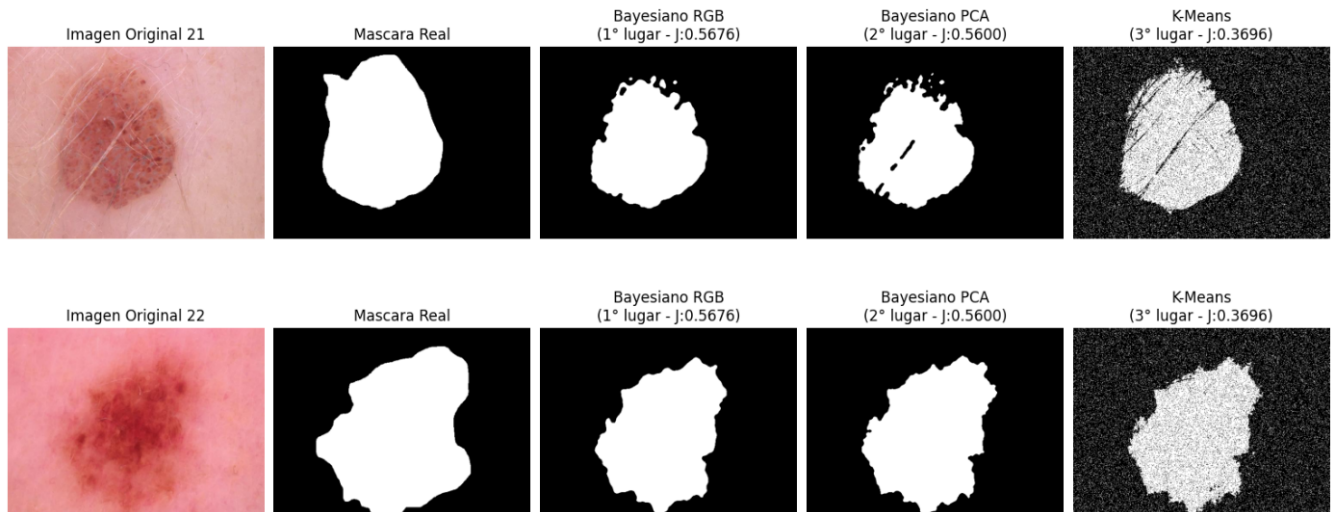


Figure 6: Comparación visual de los tres métodos de segmentación en el conjunto de prueba.

Observaciones Cualitativas:

- **K-Means:** Las segmentaciones de K-Means, si bien capturan la región general de las lesiones, muestran importantes limitaciones: bordes irregulares, tendencia a la fragmentación en lesiones heterogéneas, y dificultad para manejar lesiones de bajo contraste. En particular, K-Means mostró problemas consistentes con lesiones pequeñas o con gradientes suaves de pigmentación.
- **Bayesiano-PCA:** El Bayesiano-PCA produce resultados visualmente similares al Bayesiano-RGB, aunque con ligeras diferencias en la delimitación de bordes que explican su menor precisión.
- **Bayesiano-RGB:** Las segmentaciones del Bayesiano-RGB muestran bordes más definidos y mejor correspondencia con las máscaras de referencia, particularmente en regiones de transición entre lesión y tejido sano

La Figura 6 ilustra claramente la superioridad visual del clasificador Bayesiano-RGB, con segmentaciones que mejor preservan la forma y estructura de las lesiones originales. La comparación visual también muestra que ambos métodos bayesianos mantienen mejor la integridad estructural de las lesiones, mientras que K-Means tiende a subsegmentar o sobresegmentar dependiendo de las características específicas de cada imagen. Estas observaciones cualitativas confirman la superioridad de los enfoques supervisados para la tarea de segmentación de lesiones dermatológicas a nivel de píxel.

6 Conclusiones

El presente estudio permitió desarrollar y comparar diferentes enfoques para la segmentación de lesiones dermatológicas a nivel de píxel, demostrando la efectividad de los métodos supervisados sobre los no supervisados en esta tarea específica. El clasificador Bayesiano en espacio RGB emergió como el método de mejor rendimiento, alcanzando un índice de Jaccard de 0.5676 y mostrando un balance óptimo entre sensibilidad (86.13%) y especificidad (83.69%). Su superioridad se atribuye a la capacidad de aprovechar plenamente la información discriminativa contenida en los canales de color originales sin pérdida de información por transformaciones, resultando en un modelo robusto y efectivo para la delimitación precisa de lesiones cutáneas.

Los resultados obtenidos se alinean consistentemente con el marco teórico de clasificación supervisada, confirmando que la disponibilidad de datos etiquetados permite desarrollar modelos significativamente más precisos que los enfoques no supervisados. El rendimiento competitivo del Bayesiano-PCA, a pesar de su reducción dimensional, valida la utilidad de las técnicas de reducción de dimensionalidad para simplificar modelos manteniendo un rendimiento aceptable. Por otro lado, el desempeño limitado de K-Means refleja los desafíos inherentes de los métodos no supervisados en aplicaciones médicas donde la precisión es crítica.

Entre las limitaciones de este estudio se identifica el desbalance en los datos dermatoscópicos, donde la clase no-lesión representa aproximadamente el 75% de los píxeles, potencialmente sesgando los modelos hacia la clase mayoritaria. Adicionalmente, la asunción de normalidad multivariada en los clasificadores bayesianos puede no capturar completamente la complejidad de las distribuciones de color en lesiones dermatológicas reales, que a menudo presentan heterogeneidad y patrones no Gaussianos. La variabilidad en condiciones de captura entre imágenes también representa un desafío no completamente resuelto en el procesamiento aplicado.

Para trabajos futuros, se recomienda indagar en la incorporación de características de textura junto con las cromáticas, ya que la apariencia visual de las lesiones dermatológicas depende tanto del color como de patrones texturales. La implementación de modelos más avanzados como Support Vector Machines o Redes Neuronales Convolucionales podría capturar relaciones no lineales y patrones complejos no evidentes en los modelos actuales. Asimismo, técnicas de aumento de datos y aprendizaje por transferencia podrían ayudar a mejorar el desbalance de clases y mejorar la generalización. Finalmente, la validación en conjuntos de datos más extensos y diversos, eventualmente en entornos clínicos reales, será crucial para trasladar estos avances a aplicaciones prácticas en el diagnóstico dermatológico asistido por computadora.

References

- DataSource.ai. (2023). *Métricas de evaluación de modelos en el aprendizaje automático*. DataSource.ai. Retrieved from <https://www.datasource.ai/es/data-science-articles/metricas-de-evaluacion-de-modelos-en-el-aprendizaje> (Publicado el 25 de septiembre de 2023; consultado el [2025])
- Jaadi, Z. (2025). *Principal component analysis (pca): A step-by-step explanation*. Built In. Retrieved from <https://builtin.com/data-science/step-step-explanation-principal-component-analysis> (Publicado el 23 de junio de 2025; consultado el [2025])
- Murty, M. N., & Devi, V. S. (2011). Bayes classifier. In *Undergraduate topics in computer science* (pp. 86–102). Springer. doi: 10.1007/978-0-85729-495-1₄
- Ramírez, L. (2024). *Algoritmo k-means: ¿qué es y cómo funciona?* Thinking For Innovation. Retrieved from <https://www.iebschool.com/hub/algoritmo-k-means-que-es-y-como-funciona-big-data/> (Publicado el 30 de octubre de 2024; consultado el [2025])
- Theodoridis, S. (2013). Bayes classifier. In *Encyclopedia of systems biology* (pp. 145–148). Springer. Retrieved 2025-08-27, from <https://www.sciencedirect.com/topics/engineering/bayes-classifier> (Disponible en: ScienceDirect Topics)

Anexos

6.1 Anexo A: Resultados Adicionales

6.1.1 Estadísticos por canal (RGB)

Canal	Clase	Media	Desviación
R	Lesión	0.6917	0.1434
	No-lesión	0.8412	0.0806
G	Lesión	0.3569	0.1480
	No-lesión	0.5813	0.1040
B	Lesión	0.3546	0.1503
	No-lesión	0.6186	0.1217

Table 4: Estadísticos por canal para lesiones y no lesiones