INTELIGENCIA ARTIFICIAL EN LA CLASIFICACIÓN DE lesiones cutáneas

**Artificial intelligence in the classification of skin lesions**

Jordan Plá González1, Manuel Vilas Valiente2, Marcos A. Valdivié Rodríguez3, Claudia Olavarrieta Martínez4, Rigoberto García Gómez5, Rocío Ortiz Gancedo6, Marta L. Baguer Díaz-Romañach7

1 Universidad de La Habana, Cuba, [jordanpg41@gmail.com](mailto:jordanpg41@gmail.com)

2 Universidad de La Habana, Cuba, [mvilasvaliente@gmail.com](mailto:mvilasvaliente@gmail.com)

3 Universidad de La Habana, Cuba, [mvaldivie98@gmail.com](mailto:mvaldivie98@gmail.com)

4 Universidad de La Habana, Cuba, [claudiaomtnez@gmail.com](mailto:claudiaomtnez@gmail.com)

5 Grupo Nacional de Dermatología, Cuba, [rigobertodermatologo@yahoo.es](mailto:rigobertodermatologo@yahoo.es)

6 Universidad de La Habana, Cuba, [rocioog00@gmail.com](mailto:rocioog00@gmail.com)

7 Universidad de La Habana, Cuba, [mlbaguer@gmail.com](mailto:mlbaguer@gmail.com)

Resumen: La incidencia de casos de cáncer de piel aumenta considerablemente todos los años. Cuba no está exenta de este crecimiento siendo este tipo de cáncer el más reportado en el 2020. La detección temprana de esta enfermedad es esencial en el pronóstico favorable de los pacientes. Esta detección se realiza a través de una técnica no invasiva denominada dermatoscopía, la cual permite visualizar estructuras cutáneas profundas no detectables a simple vista. El diagnóstico dermatoscópico precisa de un periodo de aprendizaje y perfeccionamiento por parte del especialista en la clasificación de las lesiones, cuyos patrones resultan complicados de diferenciar. El presente trabajo aborda el problema de la clasificación desde una perspectiva cercana a los procedimientos que utilizan los dermatólogos al observar determinadas características, como, por ejemplo, en los métodos de Menzies, la regla ABCD y la Lista de los 7 puntos. Se analizan hasta ocho tipos de anomalías cutáneas. La clasificación utiliza modelos de aprendizaje automático supervisado basados en vectores de características asociadas principalmente a la apariencia de los bordes, colores y texturas. Se experimentó con un conjunto de imágenes de la Base de Datos ISIC (*International Skin Imaging Collaboration*). Se alcanzó una exactitud de 87.5% y de 88.3% en la clasificación en cuatro y ocho clases respectivamente. Este trabajo constituye un primer acercamiento a un sistema de ayuda al diagnóstico para ser usado por dermatólogos.

Palabras Clave: dermatoscopía, clasificación supervisada, vector de características, procesamiento de imágenes

**ABSTRACT:** There is a yearly increase of skin cancer cases. In the case of Cuba, this was the most reported cancer in 2020. The early detection of this condition influences the positive evolution of patients. Dermatoscopy is a non-invasive technique used for the early detection of deep cutaneous structures not detectable with the naked eye. Dermatoscopic diagnosis requires the specialist to learn and perfect the classification of the lesions based on hard-to-classify visual patterns. This paper approaches the problem of classification from a perspective close to the procedures used by dermatologists when observing certain features, for example, in the methods of Menzies, the ABCD rule and the 7-point list. Up to eight types of skin anomalies are analyzed. The classification uses supervised machine learning models based on feature vectors mainly associated with the appearance of edges, colors and textures. We used a set of images from the ISIC (International Skin Imaging Collaboration) database. An accuracy of 87.5% and 88.3% was achieved in the classification into four and eight classes respectively. This work constitutes a first approach to a diagnostic aid system to be used by dermatologists.

Keywords: dermatoscopy, supervised classification, feature vector, image processing

1. **INTRODUCCIÓN**

El cáncer de piel, caracterizado por el crecimiento descontrolado de células cutáneas anormales es una preocupación global debido a su alta incidencia y riesgo. En el caso específico de Cuba, la Organización mundial de la Salud reportó 9761 casos y

514 muertes en el 2020, colocándolo en el puesto 63 a nivel mundial en términos de mortalidad por esta condición [1].

La mayoría de casos en pacientes con cáncer de piel pueden recuperarse con diagnóstico y tratamiento tempranos [2]. Sin embargo, el diagnóstico temprano resulta complejo dada la existencia de múltiples subtipos y la diversidad en sus manifestaciones clínicas [3]. Esta detección se realiza a través de una técnica no invasiva denominada dermatoscopía, la cual permite visualizar estructuras cutáneas profundas no detectables a simple vista.

La aplicación de modelos de Inteligencia Artificial (IA) ha surgido como una prometedora herramienta para el diagnóstico a través de modelos basados en algoritmos de aprendizaje automático (*machine learning*) en el análisis de imágenes médicas. El presente trabajo busca emplear dichos algoritmos para proporcionar una clasificación precisa de las lesiones cutáneas en imágenes dermatoscópicas.

Para lograr este objetivo, los autores (1) compendiaron los criterios dermatológicos en la clasificación de lesiones cutáneas, (2) procesaron, equilibraron y estandarizaron las imágenes de las bases de datos y (3) evaluaron y analizaron los diferentes modelos en las distintas fases de resolución.

La dermatoscopía es una herramienta de diagnóstico no invasiva que se utiliza en la dermatología para examinar en mayor magnitud las lesiones por debajo de la superficie cutánea. Implica el uso de un instrumento óptico conocido como dermatoscopio, que amplifica *in-vivo* la imagen de la lesión sospechosa, una vez eliminados los fenómenos de refracción y reflexión de la luz sobre la piel [4 - 5]. Esta técnica mejora la precisión en el diagnóstico al proporcionar una imagen amplia y detallada de la piel que permite identificar los signos distintivos que diferencian una lesión maligna de una benigna [6].

Una vez ampliada la imagen de la lesión, se aplican los criterios de diagnóstico para su clasificación. La regla ABCD (Asimetría, Bordes, Color, Diferencias estructurales) o método de Stolz[7], la lista de 7 puntos de Argenziano[8], método de Menzies[9], y el método de análisis de patrones[10], se emplean para la clasificación de imágenes con posibles signos de malignidad. Estos criterios se emplearon en la selección de las características que influyen la competencia del modelo para identificar patrones y relaciones sobre estas.

Una dificultad que presenta el análisis de las imágenes dermatoscópicas es la presencia de artefactos, tales como burbujas de contacto, reflejos de luz, vellos, etcétera; que entorpecen la automatización del diagnóstico. Por tanto, un reto del presente trabajo es procesar, equilibrar y estandarizar las imágenes de las bases de datos, para minimizar los sesgos del modelo, pero conservar la estructura y esencia de las lesiones. El procesamiento de las imágenes comprendió el ajuste del tamaño, aplicación de filtro uniforme, extracción de vellos y la segmentación de la región de interés (ROI). Se utilizó una matriz de confusión para valorar las métricas que se emplearon en la evaluación del modelo. Estás métricas fueron exactitud (*accuracy*), precisión, sensibilidad/recobrado (*recall*), especificidad, curva de precisión-recobrado, F1-medida, macro-media (*macro-average*), y curva ROC (*Receiver Operating Characteristic*).

La implementación se realizó a través de modelos basados en algoritmos de clasificación supervisada.

1. contenido
   1. materiales y métodos

Para asegurar la reproducibilidad de los resultados presentados en esta investigación, el código fuente fue compartido en Github (<https://github.com/jordipynb/feaskles_diagnosis>)

Se empleó Python 3 como lenguaje de programación en un sistema operativo de 64 bits, en entorno Windows. Se emplearon diversas bibliotecas especializadas, entre las cuales se incluyen Numpy, Scipy, Scikit, Opencv, XGBoost, LightGBM, Pandas, Matplotlib, Seaborn y Joblib. Los experimentos fueron ejecutados en un equipo con un procesador Intel(R) Core(TM) i7-8750H de 8ª generación a 2.20GHz, acompañado de 16 GB de RAM y una tarjeta gráfica NVIDIA GeForce GTX 1060 Ti con 4 GB de memoria.

Se utilizaron un conjunto de imágenes que forman parte de la Colaboración Internacional de Imágenes de Piel (*International Skin Imaging Collaboration*), Desafío ISIC 2019, de libre acceso. Especialmente se utilizaron las bases de datos PH2, asociada al Servicio de Dermatología del Hospital Pedro Hispano en Matosinhos, Portugal; y *Human Against Machine* (HAM10000).

Las ocho clases seleccionadas en ISIC fueron carcinoma basocelular (BCC), carcinoma escamocelular (SCC), dermatofibroma (DF), lesión vascular (VASC), melanoma (MEL), nevus melanocítico (NV), queratosis actínica (AK) y queratosis benigna (BKL), como lentigo solar, queratosis seborreica y queratosis liquenoide. Debido al desbalance de clases existente en la base de datos, se efectuaron técnicas de aumento de datos para hacer el entrenamiento de los modelos más uniforme (Figura 1).

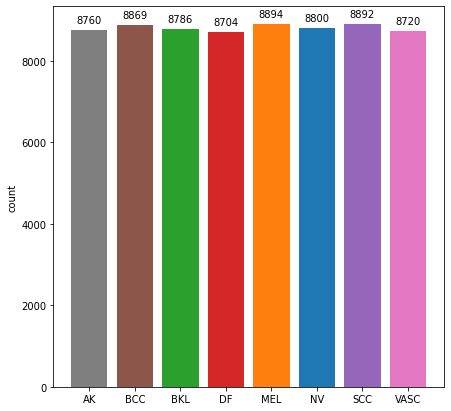
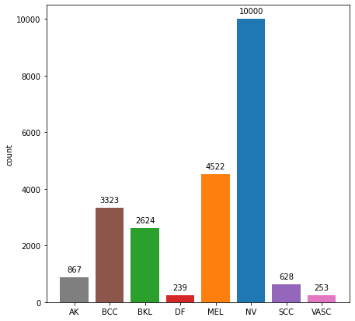


Figura 1. Distribución del número de imágenes por clases luego de eliminar nevus y aumentar los datos de entrenamiento.

2.1 Resultados y discusión

* + 1. *Extracción de características*

La tabla 1 muestra fórmulas empleadas para el cálculo de las características.

Dentro de la característica Contorno se encuentra el área de la región activa en la máscara, el área de la región convexa que encierra la lesión y la solidez, asimetría, perímetro y la irregularidad.

Dentro de la característica Color se calcularon métricas sobre los canales de los modelos de color HSV y RGB. La media o tonalidad predominante y la desviación estándar ofrecen información sobre la uniformidad y dispersión de los colores presentes en una lesión. También se calculó la raíz cuadrática media (RMS) y la raíz cuadrática desviada (RDS).

Dentro de la característica Patrones y texturas se aplicó la Matriz de Co-ocurrencia de Niveles de Gris (GLCM) y el análisis de los Patrones Locales Binarios (LBP) Ambas se realizaron sobre la región de interés en los canales RGB y HSV. Las propiedades incluidas fueron: contraste, disimilitud, homogeneidad, energía, correlación y segundo momento angular (ASM).

Tabla I Fórmulas de cálculo de las características

|  |  |
| --- | --- |
| Características | Fórmula |
| *Contorno* |  |
| Área Lesión |  |
| Área Convexa |  |
| Solidez |  |
| Asimetría |  |
| Perímetro |  |
| Irregularidad |  |
| *Color* |  |
| Media |  |
| Desviación Estándar |  |
| RMS |  |
| RDS |  |
| *Patrones y Textura* |  |
| Contraste |  |
| Disimilitud |  |
| Homogeneidad |  |
| ASM |  |
| Energía |  |
| Correlación |  |

*2.2.2 Preparación de las imágenes*

Las etapas clave de la preparación de la base de datos incluyen redimensionamiento, aplicación de filtro y eliminación de vellos (Figura 2).

Se redimensionaron las imágenes a una resolución de 512 píxeles de ancho, manteniendo su razón de proporción en el largo.

Para solucionar el problema de la distorsión cromática presente de manera recurrente en las imágenes dermatoscópicas se decidió aplicar un filtro que normaliza sus tonalidades. De esta forma se evita que la calidez de la imagen sesgue el diagnóstico cuando se utiliza el modelo de color RGB.

Con el objetivo de realizar un análisis comparativo acerca de la influencia de los vellos a la hora de clasificar, se utilizó el algoritmo de Extracción Digital de Vellos (DHR) para la extracción de estos.

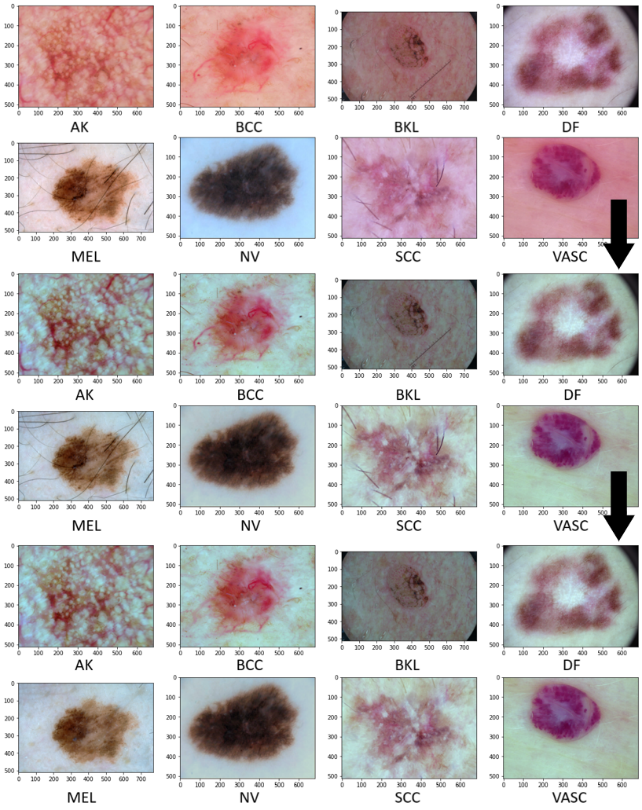


Figura 2. Fases del procesamiento de las imágenes.

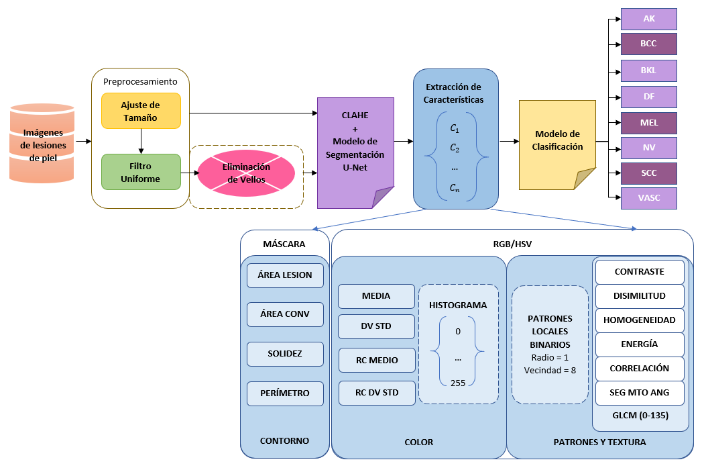
Para la búsqueda de la región de interés (ROI) y extracción de las características se utilizó el segmentador de Valdivié realizado en 2022[11]. Este sistema presenta una arquitectura U-Net que recibe como entrada una imagen dermatoscópica con modelo de color RGB y resolución de 128 de ancho por 128 de largo. Finalmente, devuelve una máscara binaria con la misma resolución. Para mejorar los resultados de la segmentación se aplicó previamente sobre la imagen un filtro de Ecualización de Histograma Adaptativo Limitado por Contraste (CLAHE). El segmentador de imágenes presentó en algunos casos extracciones incompletas de las lesiones o adición de zonas externas a estas, requiriendo correcciones manuales en 8500 de 25331 imágenes, destacando la necesidad de investigar el impacto en los resultados finales (Figura 3).



Figura 3. Resultados en las fases de la segmentación.

*2.2.3 Evaluación y análisis de los diferentes modelos en las distintas fases de resolución*

Después de detallar cada fase de la propuesta, se presenta en la figura 4 el flujo del sistema. La fase siguiente implica la evaluación de varios modelos de clasificación analizados primero con ocho clases (BCC, SCC, DF, VASC, MEL, NV, AK y BKL) y posteriormente con las cuatro clases más relevantes para el diagnóstico dermatológico (BCC, SCC, MEL y misceláneas o MISC).

Figura 4. Esquema de trabajo del sistema propuesto.

En la Tabla II se muestra el rendimiento, teniendo en cuenta la métrica *accuracy,* de los mejores modelos en la evaluación de los datos de prueba para el diagnóstico de las 8 clases. Se incluye la entrada original, que corresponde a las características con el aumento de datos inicial y sin correcciones de segmentación. Se observa una ligera mejoría en la eficacia de los algoritmos al no eliminar los vellos. Se descarta el algoritmo K vecinos cercanos (KNN) por bajo rendimiento.

Tabla II Resultados de fase experimental sobre segmentación y extracción de vellos en 8 clases.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Modelos | Sin Vellos/Multiclases | | Con Vellos/Multiclases | |
| Entrada Original | 8500 correcciones | Entrada Original | 8500 correcciones |
| KNN | 29.9% | 31.7% | 31.2% | 35.6% |
| DT[[1]](#footnote-2) | 52.0% | 53.3% | 53.9% | 55.3% |
| RF | 69.1% | 72.9% | 70.3% | 73.9% |
| XGBoost | 68.5% | 71.1% | 70.1% | 73.4% |

En la Tabla III se muestra el rendimiento de los modelos en la evaluación de los datos de prueba para el diagnóstico de las 8 clases con vellos y a partir de las 8500 correcciones en la segmentación agregando una máquina potenciadora de gradientes de luz (LightGBM). Se observa una mejoría en la eficacia de los algoritmos al incorporar el histograma de colores al vector de características. Nuevamente es utilizado para valorar los resultados el porcentaje de aciertos en la clasificación.

**Tabla III Resultados de fase experimental sobre segmentación y extracción de vellos en 8 clases.**

|  |  |  |  |
| --- | --- | --- | --- |
| Modelos | Con Vellos/8 Clases | | |
| 8500 correcciones | Con Histograma | Datos más Variados |
| DT | 55.3% | 57.0% | 61.0% |
| RF | 73.9% | 76.7% | 84.3% |
| XGBoost | 73.4% | 77.9% | 85.5% |
| LightGBM | 76.2% | 78.8% | 88.3% |

El modelo que arrojó mejores resultados fue el basado en el algoritmo LightGBM. La figura 5 muestra detalles del entrenamiento del modelo.

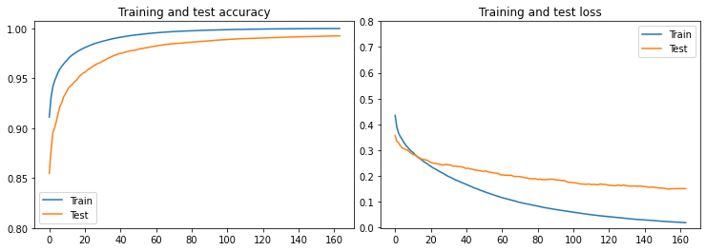


Figura 5. Curvas sobre los datos de entrenamiento y prueba en LightGBM con 8 clases

Con el empleo de cuatro clases se llega a un máximo de 85.5 % aproximadamente de exactitud. Al ser una base de datos más pequeña, el agrego de patrones lineales locales (LBP) no llevó tanto tiempo en la extracción de características, y se percibe un aumento en el rendimiento de todos los modelos. Se puede comprobar que esta adición fue un acierto, aportando más características sobre texturas y en su caso particular, sobre patrones. Con el uso de PCA en 111 componentes [12], se mejora el rendimiento para bosques aleatorios (RF) y el algoritmo de aumento de gradiente extremo (XGboost), no así para LightGBM. Este último, nuevamente resultó ser el más eficaz, por lo que el análisis se hará sobre él.

Tabla IV Resultados de fase experimental con 4 clases y reducción de dimensiones.

|  |  |  |  |
| --- | --- | --- | --- |
| Modelos | Con Vellos/4 Clases | | |
| Datos más Variados | Agrego de LBP | 111 componentes |
| RF | 79.0% | 80.1% | 82.8% |
| XGBoost | 83.7% | 84.1% | 84.2% |
| LightGBM | 85.5% | 87.5% | 84.1% |

Igualmente se obtuvo un desempeño de 98 % de exactitud en entrenamiento y 93 % en la validación, aproximadamente. Este último, nuevamente no está tan lejano al 87.5 % obtenido en el conjunto de prueba, sin embargo, la tasa de error es un poco más dispareja. Se tiene un 0.09 de error para los datos de entrenamiento y 0.21 en los de validación. En la figura 6 se aprecia la matriz de confusión normalizada sobre los datos de prueba, en el que todas las clases tienen buen desempeño según la tasa de FP, pero en la tasa de FN, la clase MEL se ve desfavorecida como contrapartida al modelo de 8 clases. En este caso, el modelo ha clasificado algunos casos de MEL en MISC, sobre todo los casos de nevus melanocíticos.



Figura 6. Matrices de confusión normalizadas en LightGBM con 4 clases

3. conclusiones

Se logró una clasificación precisa de las lesiones cutáneas en imágenes dermatoscópicas a través de modelos basados en algoritmos de aprendizaje automático (machine learning).

Los modelos que mejores resultados presentaron fueron los *ensembles*, en especial LightGBM.

Se obtuvo una mejor clasificación manteniendo los vellos en las imágenes e incorporando el histograma de colores al vector de características.

Las características identificadas como más relevantes después de aplicar PCA al análisis con cuatro clases son mayormente las vinculadas a patrones, textura y color.

referencias

1. **World Health Rankings:** Cuba Skin Cancers. 2020 [www.worldlifeexpectancy.com](https://www.worldlifeexpectancy.com/es/cuba-skin-cancers)
2. [**Cleveland Clinic**. Skin Cancer. 2023 [my.clevelandclinic.org](https://my.clevelandclinic.org/health/diseases/15818-skin-cancer)
3. **World Cancer Research Fund International**: Skin cancer statistics. 2022. [www.wcrf.org](https://www.wcrf.org/cancer-trends/skin-cancer-statistics)
4. **Sarsam, S.:** What is dermatology? DermNet, 2023. [dermnetnz.org](https://dermnetnz.org/topics/what-is-dermatology)
5. **Marghoob, A.** **y Jaimes, N**.: Overview of dermoscopy. UpToDate, 2023. [www.uptodate.com](https://www.uptodate.com/contents/overview-of-dermoscopy)
6. **Wikipedia**: Dermatoscopy. 2023 [en.wikipedia.org](https://en.wikipedia.org/wiki/Dermatoscopy)
7. **Toffoli, L**. ABCD rule. Dermoscopedia, 2023 [dermoscopedia.org](https://dermoscopedia.org/ABCD_rule)
8. **Ludwig, S**.: Seven Point Checklist. Dermoscopedia, 2019. [dermoscopedia.org](https://dermoscopedia.org/Seven_Point_Checklist)
9. **Hoffmeyer, K.:** Menzies Method. Dermoscopedia, 2018. [dermoscopedia.org](https://dermoscopedia.org/Menzies_Method)
10. **Toffoli, L.:** Pattern analysis. Dermoscopedia, 2022. [dermoscopedia.org](https://dermoscopedia.org/Pattern_analysis)
11. **Valdivié, M.:** Segmentación de imágenes dermatoscópicas utilizando la arquitectura U-Net. MATCOM. Universidad de La Habana, 2022.
12. **GeeksforGeeks**: Principal Component Analysis (PCA). 2023. [www.geeksforgeeks.org](https://www.geeksforgeeks.org/ml-principal-component-analysispca)

sintesis curricular de los autores

DrC. Marta Lourdes Baguer Díaz-Romañach, [mlbaguer@gmail.com](mailto:mlbaguer@gmail.com) ; [mbaguer@matcom.uh.cu](mailto:mbaguer@matcom.uh.cu) Nació el 10 de noviembre de 1960 en La Habana, Cuba. Culmina sus estudios universitarios de Licenciatura en Cibernética Matemática en 1983 y defiende su doctorado (Dr. rer. nat. Numerische Mathematik) en 1990 en la Universidad Humboldt de Berlín. Se desempeña como profesora del Departamento de Matemática Aplicada de la Facultad de Matemática y Computación de la Universidad de La Habana desde 1983 y desde 2009 como Profesora Titular. Sus líneas principales de trabajo investigativo han estado relacionadas con aplicaciones de Métodos Numéricos del Álgebra Lineal, Procesamiento de Imágenes médicas y Lingüística Computacional. Ha participado en varios proyectos nacionales e internacionales y en redes académicas. Ha sido miembro o presidente de varios tribunales de doctorado (u oponente), así como tutora de varias tesis de licenciatura, maestría y doctorado. Es autora o coautora de varios artículos publicados en revistas de alto impacto científico, nacionales e internacionales. Actualmente dirige el Programa Nacional de Ciencias Básicas y Naturales, es miembro del Consejo Científico de la facultad, coordina la Mención Matemática Numérica de la Maestría en Ciencias Matemáticas y es miembro de los consejos editoriales de las Revistas Investigación de Operaciones (UH, Cuba) y RMTA (UCR, Costa Rica).

DrC. Rigoberto García Gómez, [rigobertodermatologo@yahoo.es](mailto:rigobertodermatologo@yahoo.es) nació el 6 de junio de 1975. Doctor en Ciencias Médicas, Máster en Enfermedades Infecciosas, Especialista de primer y segundo grados en Dermatología, Especialista de primer grado en MGI. Es Titular de Dermatología, Investigador Titular y jefe del grupo nacional de Dermatología. Ha sido presidente de tribunales de promoción, estatales y de segundo grado de la especialidad, es Miembro del Tribunal Permanente de Grados Científicos y Académico de la Academia de Ciencias de Cuba. El Dr. Rigoberto García Gómez es, además, Académico de Honor de la Academia Española de Dermatología y Venereología y actualmente director de la Clínica Central Cira García. Sus líneas de investigación principales son Dermatoscopía y Cáncer de piel. Tiene en su haber 27 publicaciones en revistas nacionales y extranjeras. Ha obtenido, entre otras, las siguientes distinciones: Orden Carlos J Finlay, 2020; Medalla Jesús Menéndez, 2019; Medalla Tomás Roig, 2023; Medalla Internacionalista: 2005 y 2002; Sello de Oro de la Academia Española de Dermatología y Venerología 2017; Académico de Honor de la Academia Española de Dermatología y Venerología 2017; Premio Anual de la Academia de Ciencias de Cuba 2016; Premio Anual de Salud 2015 en la Categoría Investigación aplicada; Premio Anual de Salud 2015: Mejor Tesis Doctoral y Distinción Abnegado Combatiente Internacionalista por la Salud de los Pueblos: 2002.

1. Árbol de Decisión (*Decision Tree*) [↑](#footnote-ref-2)