

SISTEMA DE DETECCIÓN DE PIE DIABÉTICO A PARTIR DE IMÁGENES EN EL ESPECTRO VISIBLE E INFRARROJO

JHONATAN ESTEBAN RUBIANES REALPE



**UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA
SANTIAGO DE CALI**

2025

**SISTEMA DE DETECCIÓN DE PIE DIABÉTICO A PARTIR DE IMÁGENES EN EL
ESPECTRO VISIBLE E INFRARROJO**

JHONATAN ESTEBAN RUBIANES REALPE

Trabajo de grado para optar por el título de:
Ingeniero Electrónico

Directores:

Sandra Esperanza Nope Rodríguez, Ph.D.

Andrés David Restrepo Girón, Ph.D.

Grupo de Investigación Percepción y Sistemas Inteligentes

Programa:

(3744) Ingeniería Electrónica

**UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA
SANTIAGO DE CALI**

2025

Nota de aceptación:

Firma del director del trabajo

Firma del director del trabajo

Firma del evaluador

Firma del evaluador

Santiago de Cali, 5 de septiembre del 2025.

AGRADECIMIENTOS

Primeramente, agradezco a Dios, por brindarme la fortaleza, salud y sabiduría para culminar este importante proceso académico. Su guía constante ha sido fundamental en cada paso de este camino.

A mi familia, por su amor incondicional y por ser mi mayor fuente de apoyo. A mis padres, Sandra y Alexander, por enseñarme con su ejemplo el valor del esfuerzo y la dedicación; a mi hermano Alejandro, por su compañía y aliento constante; y especialmente a mis abuelos, por su inmenso cariño, enseñanzas y respaldo inquebrantable.

A mis amigos y a mi pareja, por estar presentes en los momentos difíciles y celebrar conmigo los logros alcanzados. Su compañía ha sido vital para mantener el ánimo y la motivación durante todo este proceso.

A los profesores que han sido también amigos desde mis años de colegio, quienes sembraron en mí los valores y conocimientos que hoy me definen tanto en lo personal como en lo profesional. En especial, agradezco a Leonardo, Eduardo y Maryi, por aportar parte de sí mismos a mi formación.

A la Universidad del Valle, por ser el espacio que me acogió y me brindó las herramientas necesarias para crecer como profesional. Mi gratitud al cuerpo docente, cuyo compromiso y dedicación han sido fundamentales en mi proceso formativo, en especial, a los directores de este trabajo de grado, la profesora Sandra y el profesor Andrés, por su orientación, paciencia y valiosas aportaciones. Gracias por guiarme con rigor académico y cercanía humana en la elaboración de esta investigación.

CONTENIDO

	Pág.
1. INTRODUCCIÓN	11
1.1 PLANTEAMIENTO DEL PROBLEMA	12
1.2 JUSTIFICACIÓN	13
1.3 ANTECEDENTES.....	14
1.3.1 Antecedentes a nivel local.....	14
1.3.2 Antecedentes a nivel iberoamericano.....	14
1.3.3 Antecedentes a nivel internacional.....	15
1.4 OBJETIVOS.....	19
1.4.1 General	19
1.4.2 Específicos	19
2. MARCO TEÓRICO	19
2.1. DIABETES, PIE DIABÉTICO Y DIAGNOSTICO	19
2.1.1. Anatomía y fisiología del pie humano	19
2.1.2. Diabetes mellitus.....	20
2.1.3. Pie diabético	20
2.1.4. Neuropatía periférica e isquemia por enfermedad arterial periférica.....	21
2.2. BANCO DE DATOS.....	23
2.2.1. Base de datos STANDUP	23
2.3 VISIÓN ARTIFICIAL	25
2.3.1. Preprocesamiento de datos.....	25
2.3.2. Segmentación de imágenes.	26
2.3.2.1. Región de interés (ROI)	26
2.3.2.2. Histograma de una imagen	26
2.3.2.3. Matiz Saturación Valor (HSV) y YIQ.....	27
2.3.2.4. Coeficiente de Dice	29
2.3.3. Fusión de imágenes	29
2.3.3.1 Métodos de fusión Basados en la Transformada.....	30
2.3.3.2. Métodos de fusión Espaciales	32
2.3.3.3. Métricas evaluativas de fusión de imágenes	33
2.4. RECONOCIMIENTO DE PATRONES.....	34
2.4.1. Extracción y selección de características	35
2.4.1.1. Entropía	35
2.4.1.2. Curtosis y asimetría	36
2.4.1.3 Energía.....	36
2.4.2. Máquinas de Soporte Vectorial (SVM).....	36
2.4.3. Máquinas de aprendizaje superficial	38

2.4.4 Máquinas de aprendizaje profundo	40
2.4.4.1 Yolo	40
2.4.4.2 Épocas	40
2.4.4.3. Lotes	41
2.4.4.4 Transferencia de aprendizaje.....	41
2.4.4.5. Validación cruzada	41
2.5. METRICAS DE DESEMPEÑO DE CLASIFICADORES.....	41
2.5.1. Matriz de confusión.....	41
2.5.2. Métrica de desempeño precisión	42
2.5.3. Métrica de desempeño sensibilidad	42
2.5.4. Métrica de desempeño F1 score	42
2.5.5. Métrica de desempeño exactitud y especificidad.....	42
3. DESARROLLO DEL SISTEMA DE DETECCIÓN DEL PIE DIABÉTICO	43
3.1 SEGMENTACIÓN	44
3.1.1. Detección inicial de la región de interés (ROI).....	44
3.1.2. Aislamiento de la zona plantar mediante detección en cascada.....	46
3.1.3. Segmentación manual y comparación.....	56
3.2. EXTRACCIÓN DE CARACTERÍSTICAS	58
3.2.1. Fusión de imágenes	59
3.2.2. Métricas de desempeño de los métodos de fusión de imágenes	61
3.2.3. Extracción y selección de características	63
3.3. CLASIFICACIÓN ENTRE PIES SANOS Y PIES DIABÉTICOS	69
3.3.1. Máquinas de Vectores de Soporte.....	70
3.3.2. Clasificadores estadísticos	73
3.3.3. Clasificadores basados en redes neuronales.....	75
3.3.4. Proceso de validación cruzada para clasificadores estadísticos y redes neuronales	78
3.3.5. Análisis de resultados.....	80
3.3.6. Resultados adicionales	82
4. CONCLUSIONES	84
5. TRABAJOS FUTUROS	86
6. REFERENCIAS.....	87

LISTA DE TABLAS

Tabla 1. Resumen de las Características Principales de Trabajos Previos.....	18
Tabla 2. Resultados de la detección de pies en imágenes externas y en conjunto de prueba interno.....	51
Tabla 3. Valores de pixeles para algunos casos de color.....	55
Tabla 4. Fragmento de resultados DSC para segmentación de imágenes.	57
Tabla 5. Implementación de métricas para evaluación método de fusión de imágenes.	62
Tabla 6. Puntaje global para los tres mejores métodos de fusión.....	63
Tabla 7. Resultados ANOVA para las 23 características.	67
Tabla 8. Característica diferencia promedio de temperatura entre pies para fragmento de pacientes seleccionados.	68
Tabla 9. Resultados SVM para diferentes kernel.	71
Tabla 10. Promedio de métricas de evaluación para SVM en clasificación binaria.	72
Tabla 11. Mejor caso de clasificación con SVM según kernel y número de características.	72
Tabla 12. Resultados matrices de confusión SVM para el mejor de los resultados.	72
Tabla 13. Resultados clasificadores estadísticos para 23 características con opción de rechazo..	74
Tabla 14. Resultados clasificadores estadísticos para 13 características con opción de rechazo..	74
Tabla 15. Métricas clasificadores estadísticos para 23 características.	74
Tabla 16. Métricas clasificadores estadísticos para 13 características.	74
Tabla 17. Resultados matrices de confusión 23 características.....	75
Tabla 18. Resultados matrices de confusión 13 características.....	75
Tabla 19. Evaluación semilla y función para clasificadores con redes neuronales.....	76
Tabla 20. Resultado métricas evaluativas con clasificadores para 23 características con redes neuronales.....	77
Tabla 21. Resultado métricas evaluativas con clasificadores para 13 características con redes neuronales.....	78
Tabla 22. Promedio de 20 intentos por validación cruzada para clasificadores estadísticos con 23 características.....	78
Tabla 23. Promedio de 20 intentos por validación cruzada para clasificadores estadísticos con 13 características.....	79
Tabla 24. Mejor desempeño entre clasificadores estadísticos.....	79
Tabla 25. Promedio de 20 intentos por validación cruzada para redes neuronales con 23 características.....	79
Tabla 26. Promedio de 20 intentos por validación cruzada para redes neuronales con 13 características.....	80
Tabla 27. Conteo resultados redes neuronales.....	80
Tabla 28. Mejor desempeño entre clasificadores con redes neuronales.....	80
Tabla 29. Resultados clasificación tres clases.....	82

LISTA DE FIGURAS

Figura 1. Secciones del pie. [22]	20
Figura 2. Acumulación de glucosa en la sangre. [26]	20
Figura 3. Grados físicos del pie diabético. [28].....	21
Figura 4. Prueba del monofilamento. [31].....	22
Figura 5. Uso del diapasón para verificar sensibilidad vibratoria. [32]	22
Figura 6. Márgenes utilizadas durante la adquisición de imágenes térmicas del pie plantar. [21]	24
Figura 7. Distribución del conjunto de datos STANDUP	24
Figura 8. Ejemplos de fotografías y termografías de la base de datos STANDUP.....	25
Figura 9. Espacio del color HSV. [41]	27
Figura 10. Procedimiento para fusión de imágenes utilizando Transformada Wavelet.	30
Figura 11. Procedimiento para fusión de imágenes utilizando transformada del coseno discreta.	31
Figura 12. Procedimiento para fusión de imágenes utilizando transformada pirámide laplaciana	31
Figura 13. Primera etapa del reconocimiento de patrones. [56]	34
Figura 14. Segunda etapa del reconocimiento de patrones. [56].....	34
Figura 15. Representación geométrica del principio de las SVM [56]	37
Figura 16. Estructura general de una red neuronal. [66]	39
Figura 17. Diagrama de flujo sistema de detección pie diabético.	43
Figura 18. Diagrama de flujo primera segmentación	44
Figura 19. Resultados delimitación y corte de ROI mediante histogramas.....	45
Figura 20. Algoritmo implementado para segunda segmentación de los pies.	47
Figura 21. Etiquetado pie izquierdo y derecho en Roboflow.	48
Figura 22. Evolución de las métricas de rendimiento durante el entrenamiento de YOLO.....	49
Figura 23. Resultados ejemplo del sistema de detección de pies mediante YOLO.	50
Figura 24. Mascara generada y aplicada caso ejemplo.	51
Figura 25. Sujeto de prueba externo a la base de datos.....	52
Figura 26. Resultados HSV, YIQ y combinación.	54
Figura 27. Resultados del filtrado HSV y YIQ.	54
Figura 28. Resultado filtrado de colores.....	55
Figura 29. Resultado final segmentación planta de los pies.....	56
Figura 30. Pies segmentados a través de la herramienta Image Segmenter de MATLAB.....	57
Figura 31. Imágenes con problemas en el proceso de entrenamiento y clasificación.	58
Figura 32. Flujo general para la obtención del vector de características a partir de imágenes segmentadas.....	58
Figura 33. Resultado ejemplo fusión transformadas para un paciente diabético.	60
Figura 34. Resultados métodos de fusión espacial.	61
Figura 35. Variación del puntaje de fusión de imágenes según el método aplicado.....	62
Figura 36. Puntaje global para cada método de fusión implementado.	63
Figura 37. Diagramas de violín para característica relevante y no relevante	65
Figura 38. Resultados diagramas de violín características significativas.....	66
Figura 39. Resultados diagramas de violín características NO significativas.....	66
Figura 40. Arquitectura final para clasificadores con redes neuronales.....	77
Figura 41. Casos con bajo contraste térmico pie-fondo.	81

RESUMEN

El pie diabético es una de las complicaciones más severas de la diabetes mellitus, caracterizado por infecciones, ulceraciones y riesgo de amputación. Bajo esta problemática, el presente trabajo desarrolla un sistema automático de apoyo al diagnóstico temprano del pie diabético, utilizando imágenes en el espectro visible e infrarrojo de la planta de los pies. A diferencia de otros enfoques dependientes de información clínica o antropométrica, este sistema se basa exclusivamente en el procesamiento de imágenes, permitiendo su aplicación en escenarios prácticos y no controlados.

El sistema aborda el reto de clasificar entre pie diabético y sano mediante una etapa de segmentación automatizada, especialmente desafiante por los fondos no controlados. Esta etapa inicia con un recorte preliminar de ambas plantas del pie a partir de análisis de histogramas y continúa con una estrategia en cascada que combina redes neuronales YOLO, transformaciones de espacios de color (RGB, HSV, YIQ) y operadores morfológicos, sobre imágenes en el espectro visible e infrarrojo.

Tras la segmentación, se aplicaron técnicas de fusión multispectral para la extracción de características, evaluando siete métodos y seleccionando el de mejor rendimiento según métricas cuantitativas como clasificatorias. Posteriormente, se implementaron técnicas de extracción y selección de características, tomando como base aquellas previamente identificadas en trabajos anteriores, con el objetivo de reducir la dimensionalidad del conjunto de datos sin comprometer el rendimiento del sistema en términos de precisión, sensibilidad y F1-Score.

El sistema fue entrenado y validado con una base de datos pública a través de diversos clasificadores: máquinas de vectores de soporte (SVM), modelos estadísticos como Naive Bayes, LDA, QDA y K-NN, así como redes neuronales como el perceptrón multicapa, Patternet y redes de base radial (RBF). En la clasificación binaria, el mejor desempeño se obtuvo con el clasificador SVM, alcanzando una exactitud del 95,0 %, precisión del 96,6 %, sensibilidad del 93,3 % y un F1-Score del 94,9 %, superando el desempeño reportado en los antecedentes, los cuales informaban una exactitud de 89,2 %, precisión de 96,2 %, sensibilidad de 81,9 % y un F1-Score de 88,3 %.

Adicionalmente, se incorporó una opción de rechazo, ausente en trabajos anteriores, permitiendo abstenerse de clasificar cuando la confianza del modelo es baja. En este escenario, el clasificador QDA obtuvo el mejor balance: exactitud del 90 %, precisión del 92,9 %, sensibilidad del 86,7 %, especificidad del 93,3 % y un F1-Score del 89,7 % en la misma clasificación binaria, siendo seleccionado para el sistema implementado.

Aunque el objetivo principal del estudio fue la clasificación entre pie diabético y pie sano, se aprovechó que la base de datos contenía casos etiquetados con neuropatía e isquemia para explorar una clasificación multiclase. Se intentó distinguir entre tres clases: pie sano, pie diabético con neuropatía y otro con isquemia. El mejor desempeño se obtuvo con el clasificador Naive Bayes, logrando una sensibilidad del 91,7 %, precisión del 73,3 % y un F1-score del 81,5 % para la clase sana. Sin embargo, las métricas para las clases afectadas fueron inferiores al 80 %, lo que sugiere que las características empleadas distinguen bien los pies sanos, pero no discriminan con la misma eficacia entre las patologías del pie diabético, lo cual podría explorarse en futuros estudios.

Palabras clave: detección temprana de pie diabético, procesamiento de imágenes, fusión multispectral, imágenes termográficas, imágenes en espectro visible, segmentación automática, clasificación.

ABSTRACT

Diabetic foot is one of the most severe complications of diabetes mellitus, characterized by infections, ulcerations, and risk of amputation. Given this challenge, this study develops an automated system to support the early diagnosis of diabetic foot, using visible and infrared images of the soles of the feet. Unlike other approaches that rely on clinical or anthropometric information, this system relies exclusively on image processing, allowing its application in practical, uncontrolled scenarios.

The system addresses the challenge of classifying diabetic and healthy feet through an automated segmentation stage, which is particularly challenging due to uncontrolled backgrounds. This stage begins with a preliminary cropping of both soles based on histogram analysis and continues with a cascade strategy that combines YOLO neural networks, color space transformations (RGB, HSV, YIQ), and morphological operators on images in the visible and infrared spectrum.

Following segmentation, multispectral fusion techniques were applied for feature extraction, evaluating seven methods and selecting the best performer based on quantitative metrics such as classifiers. Subsequently, feature extraction and selection techniques were implemented, based on those previously identified in previous work, with the goal of reducing the dimensionality of the dataset without compromising system performance in terms of accuracy, sensitivity, and F1 score.

The system was trained and validated with a public database using various classifiers: support vector machines (SVMs), statistical models such as Naive Bayes, LDA, QDA, and K-NN, as well as neural networks such as multilayer perceptron, Patternet, and radial basis networks (RBFs). In binary classification, the best performance was obtained with the SVM classifier, achieving an accuracy of 95.0%, precision of 96.6%, sensitivity of 93.3%, and an F1-Score of 94.9%. This surpassed the performance reported in previous studies, which reported an accuracy of 89.2%, precision of 96.2%, sensitivity of 81.9%, and an F1-Score of 88.3%.

Additionally, a rejection option, absent in previous work, was incorporated, allowing the classifier to refrain from classifying when model confidence is low. In this scenario, the QDA classifier achieved the best balance: accuracy of 90%, precision of 92.9%, sensitivity of 86.7%, specificity of 93.3%, and an F1-Score of 89.7% in the same binary classification, and was selected for the implemented system.

Although the primary objective of the study was the classification between diabetic and healthy feet, the fact that the database contained cases labeled with neuropathy and ischemia was used to explore multiclass classification. An attempt was made to distinguish between three classes: healthy foot, diabetic foot with neuropathy, and another with ischemia. The best performance was obtained with the Naive Bayes classifier, achieving a sensitivity of 91.7%, accuracy of 73.3%, and an F1-score of 81.5% for the healthy class. However, the metrics for the affected classes were below 80%, suggesting that the features used distinguish healthy feet well but do not discriminate as effectively between diabetic foot pathologies, a matter that could be explored in future studies.

Keywords: early detection of diabetic foot, image processing, multispectral fusion, thermographic images, visible spectrum images, automatic segmentation, classification.

1. INTRODUCCIÓN

El pie diabético es una de las complicaciones más graves y frecuentes de la diabetes mellitus, con alto riesgo clínico por su asociación con infecciones, ulceraciones y amputaciones, que pueden presentarse entre los 12 y 18 meses tras el primer diagnóstico si no se detecta a tiempo. Sin cura definitiva, su detección y tratamiento oportunos son esenciales para frenar su progresión.

En este contexto, el presente trabajo propone el desarrollo de un sistema automatizado para la detección del pie diabético mediante el análisis de imágenes en el espectro visible e infrarrojo, iniciando con la segmentación de la planta del pie, tarea especialmente desafiante debido a la presencia de fondos no controlados. Para ello, se realiza un recorte preliminar de ambas plantas del pie mediante análisis de histogramas, lo que permite reducir parte del fondo sin eliminarlo por completo y continua con una estrategia en cascada que combina redes neuronales de detección You Only Look Once (YOLO) [1], transformaciones de color y operadores morfológicos para lograr la discriminación completa de los pies.

Posteriormente, se aplicaron técnicas de fusión multispectral para integrar la información estructural y térmica de los espectros en cuestión, evaluando hasta siete métodos de fusión, tanto por transformadas (wavelet discreta, coseno discreta y pirámide laplaciana) como por técnicas más simples (promedio, ponderado, mínimo y máximo) y seleccionando el más eficaz con base en métricas cuantitativas y de clasificación. Esta etapa generó imágenes enriquecidas que sirvieron como base para la extracción de características, seleccionadas en función de su relevancia y efectividad reportadas en trabajos previos.

Estas características alimentaron diversos clasificadores de manera paralela, incluyendo máquinas de vectores de soporte (SVM), modelos estadísticos como Naive Bayes, análisis discriminante lineal (LDA) [2], análisis discriminante cuadrático (QDA) [2] y K-Nearest Neighbors (K-NN, algoritmo de los k vecinos más cercanos) [3], así como redes neuronales de clasificación como el perceptrón multicapa, Patternet y redes de base radial (RBF). El sistema fue evaluado en la clasificación binaria entre pies sanos y con riesgo de afecciones diabéticas para los propósitos del presente trabajo, y exploratoriamente en una clasificación multiclase que incluyó los subtipos neuropático e isquémico.

Un aporte relevante de este trabajo es la incorporación de una opción de rechazo en el proceso de clasificación, ausente en los estudios revisados. Esta función permite que el sistema se abstenga de emitir un diagnóstico cuando la confianza del modelo es baja, reduciendo así el riesgo de decisiones automáticas incorrectas en contextos clínicos.

A diferencia de otros enfoques más complejos y dependientes de variables clínicas, el sistema desarrollado se basa exclusivamente en el procesamiento y análisis de imágenes de la planta del pie, incluso en entornos con fondos no controlados. Sin embargo, si las imágenes, especialmente las térmicas, no son capturadas con la calidad adecuada o presentan bajo contraste térmico entre el pie y el fondo, el sistema no logra segmentar ni clasificar correctamente. Además, se requiere que el pie esté completamente contenido dentro de una región mínima de 480×384 píxeles para garantizar un recorte adecuado; de lo contrario, partes del pie podrían quedar fuera del análisis después de la primera segmentación, realizada mediante el análisis de histogramas. Por ello, el sistema no pretende reemplazar el juicio clínico, funciona como una herramienta de apoyo que facilita la identificación temprana de riesgos, cuyos resultados deben ser interpretados dentro del contexto clínico integral del paciente.

1.1 PLANTEAMIENTO DEL PROBLEMA

La diabetes mellitus continúa siendo una de las principales preocupaciones de salud pública a nivel mundial, habiendo cobrado la vida de aproximadamente 3,4 millones de personas en 2024, lo que equivale a una muerte cada 9 segundos, según la Federación Internacional de Diabetes [4]. En Colombia, la situación también es alarmante: para el 31 de agosto de 2023, se habían reportado 1.860.370 casos de esta enfermedad crónica, según datos de la Cuenta de Alto Costo [5].

Debido a los desequilibrios metabólicos relacionados con la diabetes mellitus, alrededor del 25 % de las personas que la padecen, presentan síntomas de una complicación grave denominada pie diabético. Esta condición se desarrolla debido a la deficiencia en la secreción o actividad metabólica de la insulina, principalmente por una alteración en el páncreas [6], que afecta el control adecuado de los niveles de glucosa en la sangre, lo cual conlleva a la ulceración y destrucción de tejidos profundos de la extremidad inferior, resultando aproximadamente en amputación el 20% de los casos de infección del pie diabético, según datos de la World Diabetes Foundation reportados por el Ministerio de Salud Pública [7].

La detección temprana del pie diabético resulta fundamental para reducir el riesgo de lesiones graves y amputaciones. Sin embargo, los métodos convencionales presentan diversas limitaciones. Entre los procedimientos más comunes se encuentran: el uso del monofilamento para evaluar la sensibilidad, con una especificidad del 100 %; la exploración neurológica con diapason en el primer metatarsiano, con una precisión del 77 % [8]; y la evaluación vascular mediante la búsqueda de pulsos e hinchazones, además de exámenes de laboratorio complementarios. Aunque ninguna de estas pruebas representa un riesgo para el paciente, en conjunto configuran un proceso prolongado y dependiente de la experiencia del personal de salud.

A lo anterior se suma el hecho de que muchos profesionales en formación carecen de los conocimientos y destrezas suficientes para aplicar e interpretar correctamente estas pruebas. De acuerdo con un estudio realizado en la Universidad de Antioquia [9], solo el 16,9 % de los médicos en último año saben realizar la prueba de monofilamento y apenas el 22,3 % saben interpretarla; además, únicamente el 20,9 % identifica los factores de riesgo asociados a la amputación [10].

En consecuencia, se hace evidente la necesidad de explorar soluciones que permitan una detección del pie diabético más rápida, precisa, accesible y no invasiva, que reduzca la dependencia de múltiples procedimientos y del nivel de experticia del personal médico, y que pueda contribuir a un diagnóstico temprano y a la prevención de complicaciones graves.

En este contexto, surge la pregunta de investigación que orienta este trabajo:

¿Cómo desarrollar un sistema de detección del pie diabético a partir de técnicas de procesamiento de imágenes en el espectro visible e infrarrojo que supere las limitaciones de los métodos convencionales y facilite un diagnóstico temprano?

1.2 JUSTIFICACIÓN

El pie diabético constituye una de las complicaciones más frecuentes y severas de la diabetes mellitus, caracterizada por la ulceración y destrucción progresiva de los tejidos de la extremidad inferior. En muchos casos, esta condición culmina en infecciones graves que pueden derivar en amputaciones parciales o totales, con consecuencias irreversibles para la salud y la calidad de vida del paciente.

Para el pie diabético, estudios han comprobado que el diagnóstico y tratamiento temprano puede reducir hasta en un 45 % el riesgo de presentar lesiones graves o amputaciones [11]; por esta razón, en los últimos años se han desarrollado diferentes prototipos de detección basados en técnicas no invasivas como la termografía. Esta consiste en la implementación de cámaras infrarrojas que permiten observar, mediante escalas de color y sin contacto físico, la temperatura emitida por un cuerpo, facilitando la detección de comportamientos anómalos a través de técnicas de procesamiento de imágenes [12].

Entre estos avances destaca el PINRELL (Prototype for INfraREd analysis of Lower Limbs), un prototipo diseñado para la detección precoz mediante tres sensores en diferentes longitudes de onda (NIR, IR y VIS) [13]. Sin embargo, su costo cercano a los mil euros y la ausencia de reportes clínicos que validen su eficacia limitan su aplicación práctica. De forma similar, en Ciudad de México se desarrolló un procedimiento de análisis de temperatura con escáner IR en la planta del pie de pacientes diabéticos [11], aunque su funcionamiento depende de parámetros externos como hemoglobina glicada, tensión arterial y saturación de oxígeno, lo que genera un proceso más complejo y dependiente de exámenes de laboratorio.

Si bien estas alternativas representan avances relevantes, sus altos costos, complejidad técnica y dependencia de información clínica adicional restringen su implementación en la práctica hospitalaria o de atención primaria.

En este escenario, la propuesta de investigación cobra relevancia, pues busca desarrollar un sistema de detección del pie diabético basado en imágenes en el espectro visible e infrarrojo que no dependa de parámetros externos ni de información antropométrica. Su diseño pretende ofrecer una herramienta rápida, precisa, accesible y no invasiva, capaz de generar resultados inmediatos que faciliten la intervención temprana y reduzcan el riesgo de amputaciones.

Finalmente, este trabajo constituye también una contribución académica y tecnológica, al integrar métodos de procesamiento de imágenes con aplicaciones biomédicas. Así, se aporta al fortalecimiento de soluciones innovadoras en la atención de una complicación de gran impacto social y sanitario, posicionando la ingeniería como un aliado estratégico de la salud pública.

1.3 ANTECEDENTES

En las últimas décadas, el uso de tecnología termográfica para la detección y monitoreo de condiciones médicas relacionadas con el pie diabético ha impulsado el desarrollo de algoritmos de procesamiento de imágenes y técnicas de análisis capaces de identificar patrones específicos en termogramas asociados con esta enfermedad. Diversos autores han documentado estos avances, presentando sistemas de detección basados en análisis térmico y multiespectral, así como metodologías apoyadas en inteligencia artificial. Ejemplos de estos trabajos se describen con mayor detalle en las subsecciones siguientes del apartado correspondiente a la revisión de la literatura, donde se abordan los enfoques, resultados y técnicas empleadas. Estos patrones pueden incluir áreas de alta temperatura que indican inflamación o infección, así como áreas de baja temperatura que pueden indicar problemas de circulación sanguínea. No obstante, las imágenes adquiridas por cámaras comunes de RGB pueden complementar esta información, analizando principalmente el color del pie. Del mismo modo, dada la evolución de los algoritmos y estrategias de inteligencia artificial, estas pueden contribuir a aumentar las tasas de detección no solo de la presencia o ausencia de la condición, sino la asignación de etiquetas o categorías con relación a su avance.

Dada la consideración de las necesidades mencionadas anteriormente, se buscaron, a través de las bases de datos IEEE xplorer, Scopus y el repositorio de la Universidad del Valle, investigaciones pertinentes bajo palabras claves como “thermal”, “thermography”, “diabetic”, “foot”.

1.3.1 Antecedentes a nivel local

En el trabajo de investigación de maestría **“Sistema de Visión Artificial Multiespectral para Análisis de la Huella Plantar”** desarrollada por Ferrin en el 2014 [14], se presenta un sistema hardware-software de visión artificial donde, la parte de adquisición multiespectral de imágenes consiste en dos esquemas separados de captura: visible e infrarrojo. El primero hace uso de métodos de segmentación para un análisis morfológico del pie y, para el análisis de imágenes termográficas, se recurre a la técnica de evolución de contornos mediante niveles de conjunto en compañía de técnicas de registro no rígido.

El sistema, al realizar un análisis morfo-térmico de la zona plantar, demostró su capacidad para identificar riesgos de neuropatía diabética, detectando dos casos potenciales en un grupo de 17 pacientes evaluados. De estos, 7 eran diabéticos; sin embargo, no se especifica la presencia de sintomatologías asociadas en los pacientes evaluados, más allá de su condición de diabéticos. Asimismo, la identificación de los dos posibles casos de neuropatía diabética no contó con verificación clínica por parte de un médico especialista.

En este antecedente, el tamaño reducido de la muestra limita la capacidad de generalizar los resultados, lo que evidencia la necesidad de investigaciones con poblaciones más amplias. Las 14 imágenes empleadas para el análisis morfológico fueron tomadas en un entorno controlado, distinto al utilizado para la adquisición de las imágenes térmicas.

1.3.2 Antecedentes a nivel iberoamericano

Avila D. [11] desarrolló una aplicación clínica del Scanner de imagen IR como indicador de diferencias de temperatura en la planta del pie de pacientes diabéticos, centrando el análisis en tres áreas específicas del pie: falanges distales y medias como una sola área, falanges proximales y metatarsianas. De cada una de las tres áreas se seleccionó 5 puntos térmicos (píxeles) para ambos

pies, luego se obtiene el promedio de temperatura de dichos puntos y se compara su diferencia por zona.

El trabajo se centra en aplicar el análisis termográfico como una herramienta para la identificación temprana de riesgos asociados al pie diabético. Sin embargo, debido al limitado tamaño de la muestra, conformada por 14 pacientes diabéticos de un total de 55 estudiados, el autor no establece un patrón específico para la detección. En su lugar, sugiere aumentar la cantidad de pacientes y desarrollar un clasificador que facilite y optimice el diagnóstico médico.

Este estudio destaca por su propuesta metodológica de calcular promedios térmicos individuales para cada pie y evaluar las diferencias por zona, lo cual constituye una estrategia clave para el diseño e implementación de clasificadores diagnósticos del pie diabético. Sin embargo, el trabajo no presenta métricas cuantitativas ni indicadores de rendimiento que permitan evaluar rigurosamente la capacidad discriminativa del enfoque propuesto, lo que limita sus conclusiones y resalta la necesidad de investigaciones complementarias con métodos estadísticos y clasificadores que incluyan validación cuantitativa.

1.3.3 Antecedentes a nivel internacional

En la tesis titulada **"Liquid crystal thermography in neuropathic assessment of the diabetic foot"** desarrollada por Dr. Munish Bharara en 2007 [15], se presenta un sistema diseñado para evaluar si existe una asociación entre imágenes térmicas plantares anormales y pérdida sensorial en condiciones de carga normal en pacientes diabéticos. Se estudiaron tres grupos: diabéticos con neuropatía, sin neuropatía y sujetos sanos (n=30 por grupo). El proceso de adquisición y calibración de datos implicó una validación exhaustiva de la calibración mediante treinta pruebas in vivo, calculando el tono medio para cada temperatura de ajuste. Posteriormente, se convierten las imágenes LCT (Liquid crystal thermography) a intensidad de 8 bits y se aplica un algoritmo de mapeo tonal en MATLAB. Utilizando coeficientes polinómicos previamente determinados, se calcula la temperatura en cada píxel de la imagen a partir de su valor de tono correspondiente y se opta por el modelo de color HSV por su facilidad de implementación. Con ello, realizaron un análisis comparativo que evidenció diferencias térmicas significativas entre los grupos, confirmando la utilidad de la termografía como herramienta diagnóstica complementaria. Aunque este sistema utiliza una herramienta diferente para la obtención de imágenes termográficas como la lámina de poliéster de cristal líquido termocrómico, el autor presenta un diseño robusto para el análisis de la información, no clasificación, observando diferentes métodos y mostrando una vez más la relevancia de una característica como la diferencia del promedio de los valores de los píxeles para cada pie en este tipo de trabajos.

El **sistema de Termografía Infrarroja PINRELL**, desarrollado por el programa de Tecnología Médica de IACTEC [13], es un prototipo que tiene como objetivo identificar tempranamente posibles patologías como las afecciones del Pie Diabético mediante termografía infrarroja. La aplicación PINRELL, haciendo uso de programas de software libre como 3D Slicer y PLUS, se encarga de adquirir datos, registrar imágenes y realizar un análisis de la información utilizando diversas técnicas, incluyendo segmentación, algoritmos estadísticos y Machine (Deep) Learning. De esta manera, el sistema genera una base de datos de imágenes médicas en múltiples canales (Visible/Infrarrojo cercano/Infrarrojo), permitiendo el desarrollo de algoritmos específicos para el diagnóstico y seguimiento de neuropatías del pie diabético, analizando y monitorizando patrones de temperatura superficial. El instituto de astrofísica de canarias dio a conocer al mundo sobre este sistema en el año 2020 y señaló que el prototipo ya se encontraba en fase de inicialización de los ensayos clínicos, pero no se tiene más información al respecto desde dicho año, no se han publicado

métricas, resultados clínicos ni detalles específicos de los algoritmos o de la base de datos utilizada. No obstante, representa un ejemplo actual de que herramientas avanzadas se están implementando en el diagnóstico del pie diabético, como los clasificados estadísticos y que entran a consideración para el desarrollo del presente trabajo.

En el año 2018, Muhammad Adam Abdul Rahim en “**Detection of Diabetic Foot with Infrared Thermography**” [16], implementó un sistema de detección para el pie diabético con asistencia computacional haciendo uso de parámetros no lineales extraídos de termogramas plantares del pie, obtenidos de 66 personas, de entre ellas, 33 pacientes diabéticos sin neuropatía. El proceso inicia con la segmentación de las regiones plantares a través de un polígono y se deforman hasta conseguir un tamaño uniforme. Luego, se descomponen las imágenes con técnicas como la Transformada Wavelet Discreta (DWT) y los Espectros de Orden Superior (HOS). Además, se realiza la diferencia entre características del pie izquierdo y derecho. La metodología desarrollada logró una precisión de clasificación máxima del 89,39 %. Es uno de los pocos proyectos que ordena características significativas, las cuales están basadas en el cálculo de la entropía, homogeneidad, contraste y correlación, y se introducen en un clasificador de máquina de vectores de soporte (SVM).

Vilcahuaman, en su tesis para obtener el grado de doctor en el año 2013 [17], desarrolló un sistema a partir de la obtención de imágenes térmicas de la zona plantar del pie de un grupo de 80 personas con diabetes tipo 2 y solo 5 personas sanas en un ambiente controlado. Está compuesto por tres pasos: pre-procesamiento; segmentación, para separar los dos pies donde hacen uso del método de segmentación de Chan y Vese y finalmente, registro, haciendo uso de un algoritmo de registro rígido conocido como punto más cercano iterativo (ICP). A partir de estos pasos, se pueden evaluar varios parámetros: la media, la desviación estándar y la diferencia de temperatura entre punto y punto de la superficie plantar del pie derecho e izquierdo. Sin embargo, el mismo autor resalta la falta de un grupo más numeroso conformado por personas sanas que permita una mayor comparación entre casos y mejorar los procesos de segmentación para que las imágenes térmicas puedan ser tomadas en ambientes no controlados.

En el 2019 en el trabajo realizado por **Hernández** [18], quien propuso en su tesis doctoral tres metodologías para la detección temprana de complicaciones relacionadas al pie diabético mediante el análisis termográfico: Índice de cambio térmico basado en la diferencia entre la temperatura del pie y de valores de referencia, identificación de cambios en la distribución de temperatura plantar mediante el uso de medidas como la divergencias Kullback-Leibler y Jensen-Shannon, la distancia cuadrática y la medida Earth mover's distance; y tercero, una caracterización de la distribución de temperatura plantar a partir de técnicas probabilísticas. Antes de efectuar el análisis, realiza un proceso de segmentación por fusión de imágenes RGB y termográficas para la extracción de ambos pies, proceso que se facilita gracias a que las imágenes fueron tomadas en un ambiente controlado.

En el artículo “**Machine Learning-Driven Diabetic Foot Ulcer Detection with YOLOv5**”, aunque su objetivo principal del estudio fue identificar lesiones como úlceras en pies diabéticos, su metodología demuestra la capacidad de YOLO para segmentar regiones anatómicas específicas en imágenes médicas. Se utilizaron imágenes redimensionadas a 640×480 para optimizar el procesamiento, y se demostró que YOLO supera a métodos clásicos como CNN y R-CNN en velocidad y precisión de localización. Esta evidencia respalda su aplicación en segmentación anatómica, como la identificación de la región plantar [1], lo cual motivó su uso en este trabajo.

Finalmente, **Arteaga N., Hernández G., Ortega J. y Ruiz J.** [19], constituyen la principal fuente de referencia para este estudio, quienes utilizaron las bases de datos INAOE [20], STANDUP [21], y un conjunto de datos local. En dicho trabajo, los termogramas infrarrojos fueron procesados

inicialmente mediante una segmentación que permitió aislar la planta completa de ambos pies; posteriormente, cada pie fue separado y subdividido en cuatro regiones específicas. A partir de estas segmentaciones, se extrajeron características relevantes de cada zona, obteniendo un total de 188 características. Aunque no se especificó cuántas fueron finalmente utilizadas en la clasificación, sí se identificaron las diez más significativas para cada base de datos, destacando medidas como curtosis, asimetría, desviación estándar, valores mínimos y máximos, tanto del pie derecho como del izquierdo. Para el proceso de clasificación se empleó un modelo basado en máquinas de vectores de soporte (SVM), pero no profundizan en su arquitectura ni configuración técnica, evaluando su rendimiento con diferentes combinaciones de características y haciendo uso de métricas como exactitud, sensibilidad, especificidad, precisión y F1-score, lo cual permite establecer una comparación directa con los resultados obtenidos en el presente sistema. En la mejor de sus configuraciones alcanzaron una exactitud del 89,21%, precisión del 96,15%, sensibilidad del 81,93% y F1-score de 88,34%.

A lo largo de la investigación, se han explorado diversos sistemas de detección de imágenes aplicados en el contexto de pacientes diabéticos, centrándose en la evaluación tanto en infrarrojo como en imágenes en el espectro visible del pie. Estos sistemas han demostrado ser herramientas cruciales para la detección temprana de posibles complicaciones y la monitorización del estado de la piel en esta población específica. A continuación, se presenta la tabla 1 como resumen comparativo de las características clave de los antecedentes investigativos, destacando sus principales ventajas y enfoques tecnológicos.

REFERENCIA	DETECCIÓN	FUENTE SOMETIDA A ANÁLISIS	SOFTWARE UTILIZADO	TÉCNICAS DE PROCESAMIENTO	MÉTRICAS
[14]	Valoración morfo térmica de la huella plantar	Imagen RGB y termográfica	MATLAB	Segmentación, fusión, técnicas de registro no rígido y de evolución de contornos.	Errores absolutos promedios cercanos a cero. Para parámetros IHC, Tmedia y MFT: 1 desviación estándar. Para otros: 2 desviaciones estándar.
[11]	Diferencia de temperatura entre plantas de los pies en paciente diabético	Imagen termográfica	Visual Basic, Excel, Macros	Filtrado y segmentación	N/A
[15]	Característica neuropática del pie diabético	RGB	MATLAB, Labview.	Mapeo tonal, modelo del color HSV.	N/A

[13]	Posibles patologías como las afecciones del pie Diabético	RGB, infrarrojo cercano e infrarrojo	3D Slicer, PLUS	Segmentación, algoritmos estadísticos y Machine (Deep) Learning	N/A
[16]	Detección del pie diabético	Imagen termográfica	N/A	Segmentación, Transformada Wavelet Discreta, espectros de Orden Superior, clasificador.	Resultado clasificador: Precisión máxima del 89,39 %. Valor Predictivo positivo del 96,43 %. Sensibilidad del 81,81 %. Especificidad del 96,97 %.
[17]	Diagnóstico precoz del pie diabético	Imagen termográfica	MATLAB	Segmentación de Chan y Vese, registro rígido (punto más cercano iterativo)	N/A
[18]	Afectaciones tempranas no visibles del pie diabético	Imagen RGB y termográfica	MATLAB	Segmentación por fusión, corrección de postura.	N/A
[1]	Detección de úlceras de pies diabéticos.	Imagen RGB	Python y OpenCv	Segmentación basada en aprendizaje profundo con YOLO.	N/A
[19]	Detección entre pacientes sanos y diabéticos	Imagen RGB y termográfica	N/A	Segmentación, extracción y selección de características, clasificación con SVM.	Exactitud: 0.8921 Precisión: 0.9615 Exhaustividad: 0.8193 Puntuación F1: 0.8834

Tabla 1. Resumen de las Características Principales de Trabajos Previos.

1.4 OBJETIVOS

1.4.1 General

Desarrollar un sistema de visión artificial que identifique la enfermedad del pie diabético en pacientes, a partir del procesamiento de imágenes en el espectro visible e infrarrojo.

1.4.2 Específicos

1. Implementar un método de segmentación de la planta de los pies a partir de la fusión de las imágenes en espectro visible e infrarrojo.
2. Identificar las características visuales distintivas que indican la presencia de la enfermedad del pie diabético y que puedan ser implementadas en el análisis de imágenes en el espectro visible, infrarrojo o en la imagen fusionada.
3. Implementar los algoritmos para el reconocimiento y diagnóstico automático del pie diabético a partir de las características seleccionadas.
4. Establecer los alcances y limitaciones del sistema a partir del análisis de resultados.

2. MARCO TEÓRICO

En esta sección se presentan los fundamentos teóricos necesarios para el desarrollo del presente trabajo. Se abordan conceptos clave relacionados con la diabetes mellitus, el pie diabético, sus características clínicas y los métodos actuales de diagnóstico. Asimismo, se introducen los principios fundamentales de la visión artificial y el reconocimiento de patrones, tecnologías que constituyen la base metodológica del sistema propuesto.

2.1. DIABETES, PIE DIABÉTICO Y DIAGNOSTICO

2.1.1. Anatomía y fisiología del pie humano

El pie humano es la parte más lejana del cuerpo, una estructura anatómica compuesta por 26 huesos, conectados a través de 33 articulaciones, además de músculos, ligamentos y tendones, y un extenso sistema de vasos sanguíneos y nervios que permiten la sensibilidad y el control [22]. Su función dado a su diseño y ubicación, es soportar el peso corporal, brindar estabilidad y permitir el desplazamiento.

Generalmente, se consideran tres regiones principales del pie: el retropié (talón y tobillo), el mediopié (arco) y el antepié (metatarsos y dedos) [23], zonas que no solo se componen de huesos, sino también del sistema vascular y nervioso que, distribuido en ellas, y que, en el contexto de la diabetes, puede presentar diferentes sintomatologías. Estas regiones pueden observarse esquemáticamente en la Figura 1.

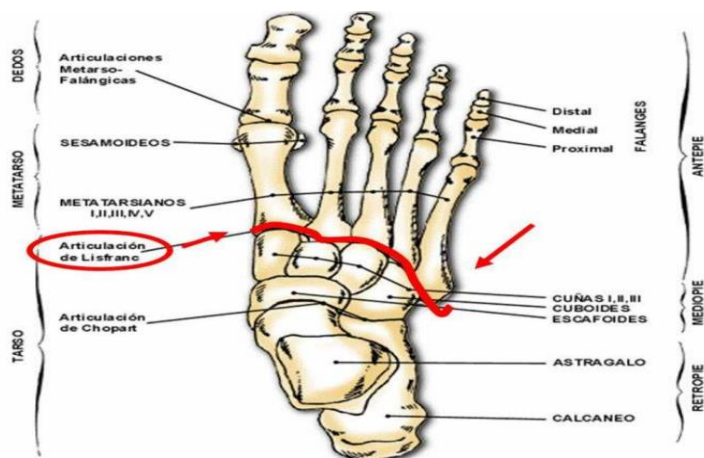


Figura 1. Secciones del pie. [22]

La irrigación sanguínea está principalmente a cargo de las arterias tibial anterior, tibial posterior y peronea, mientras que la innervación (encargado de la sensibilidad) está mediada por los nervios tibial y peroneo, derivados del nervio ciático [24].

2.1.2. Diabetes mellitus

La diabetes mellitus es una enfermedad que se presenta cuando los niveles de glucosa (tipo de azúcar simple que constituye la principal fuente de energía para las células del cuerpo) en la sangre son altos, debido a alteraciones en la secreción de insulina, en su acción, o en ambos procesos, como se ilustra en la figura 2. La insulina es una hormona producida por el páncreas que permite que la glucosa entre en las células para ser utilizada como fuente de energía. Cuando este mecanismo falla, la glucosa se acumula en el torrente sanguíneo y puede causar daños en diferentes órganos y tejidos del cuerpo [25].

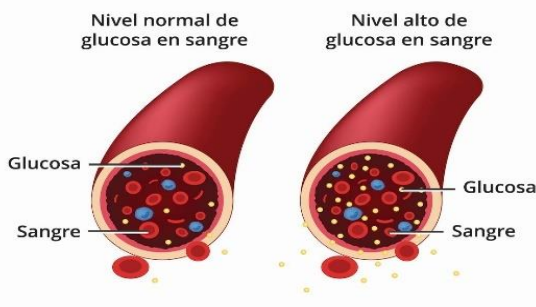


Figura 2. Acumulación de glucosa en la sangre. [26]

Como consecuencia crónica de esta enfermedad aparece el pie diabético, una complicación que, como su nombre lo indica, afecta directamente a los pies, provocando daño a los nervios y reducción del flujo sanguíneo, lo que puede llevar, en los casos más graves, a la formación de úlceras, infecciones severas e incluso a la amputación.

2.1.3. Pie diabético

El pie diabético es una de las complicaciones crónicas más severas de la diabetes. En el ámbito médico, se define como una alteración de origen neuropático y/o vascular, que se desarrolla cuando la diabetes comienza a afectar el sistema nervioso y el flujo sanguíneo de las extremidades inferiores. La neuropatía periférica, caracterizada por la pérdida progresiva de sensibilidad, y la enfermedad arterial periférica, que reduce el riego sanguíneo, son factores determinantes en la aparición del pie diabético. Cualquier daño estructural o funcional en el pie, agravado por traumatismos o infecciones, puede provocar úlceras, necrosis, infecciones profundas, y en los casos más graves, llevar a la amputación parcial o total del pie [25].

La piel plantar, gruesa y rica en glándulas sudoríparas, sirve como una barrera protectora [27], pero en personas con diabetes, puede sufrir alteraciones como resequedad, pérdida de sensibilidad o disminución en la capacidad de cicatrización [25]. Estos factores, combinados con una carga mecánica mal distribuida, favorecen la aparición del pie diabético, una de las complicaciones más graves y comunes de esta enfermedad crónica.

Para la detección temprana del pie diabético, es esencial una evaluación clínica completa que identifique signos de neuropatía y compromiso vascular. Se emplean pruebas como el monofilamento para valorar la sensibilidad táctil, el diapasón para la sensibilidad vibratoria y la palpación de pulsos periféricos (pedio y tibial posterior) para detectar isquemia. Estas evaluaciones, junto con la inspección del estado de la piel, uñas y posibles lesiones, permiten identificar a pacientes en riesgo y aplicar medidas preventivas antes de que surjan complicaciones graves.

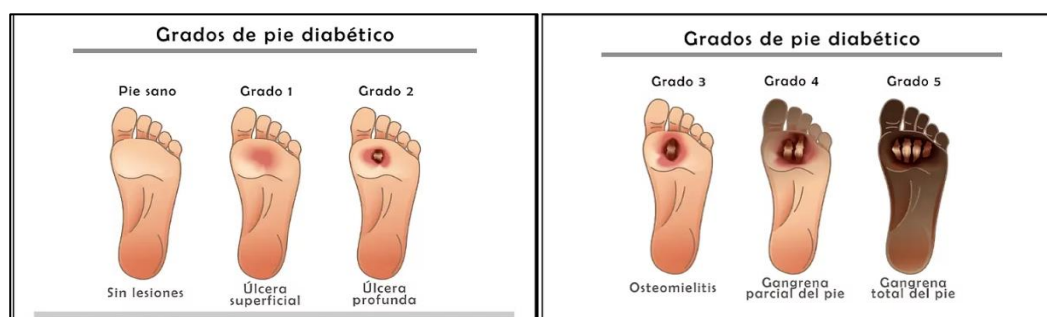


Figura 3. Grados físicos del pie diabético. [28]

La figura 3 ilustra los diferentes grados de evolución del pie diabético, destacando cómo una lesión aparentemente menor puede progresar rápidamente hacia cuadros severos si no se detecta y trata a tiempo. En el grado 1, se observa una úlcera superficial, que con el avance de la infección o la falta de irrigación puede profundizarse (grado 2) y comprometer estructuras más profundas como el hueso (osteomielitis en grado 3). En fases más avanzadas, la infección puede provocar gangrena parcial (grado 4) e incluso gangrena total del pie (grado 5), lo cual conlleva un alto riesgo de amputación.

2.1.4. Neuropatía periférica e isquemia por enfermedad arterial periférica

La neuropatía periférica es una de las complicaciones más comunes de la diabetes, y a causa de los niveles elevados de glucosa en la sangre, se genera daño progresivo a los nervios periféricos. Esta condición genera pérdida de sensibilidad en el pie y mayormente en la piel plantar, lo que impide que el paciente perciba heridas, cambios drásticos de temperatura que puedan generar quemaduras, o zonas de presión excesiva causadas por el calzado o actividades físicas básicas como el caminar [23].

Por otro lado, la isquemia en el pie diabético es consecuencia de la enfermedad arterial periférica (EAP), consiste en la obstaculización del paso de la sangre a través de las arterias que irrigan las extremidades inferiores, reduciendo la oxigenación y el aporte de nutrientes a los tejidos del pie. Como resultado, cualquier úlcera que se forme tendrá una capacidad de sanación y cicatrización leve y una mayor propensión a infectarse. En casos severos, la isquemia puede causar necrosis tisular y requerir amputación [29].

El monofilamento es un dispositivo, generalmente un hilo de nylon delgado, que, al aplicarlo en áreas específicas de la planta del pie, permite evaluar si el paciente es capaz de sentir las diferentes presiones ejercidas por el monofilamento y con ello, evaluar su sensibilidad [30]. De no sentir el monofilamento en estas zonas, es posible asociarlo a la neuropatía. La Figura 4 ilustra la prueba del monofilamento, en la que se señalan los puntos principales de aplicación en la planta del pie.

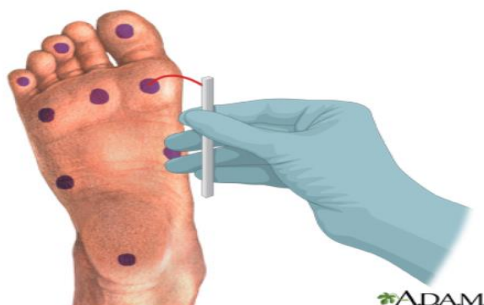


Figura 4. Prueba del monofilamento. [31]

El diapasón es otra herramienta utilizada en la detección de pérdida de sensibilidad de las diferentes zonas del pie en personas con diabetes, la diferencia con el monofilamento es que la prueba con este dispositivo es a través de vibraciones [8]. La Figura 5 muestra un ejemplo del uso del diapasón sobre la falange del pie, evidenciando la evaluación de la sensibilidad vibratoria.

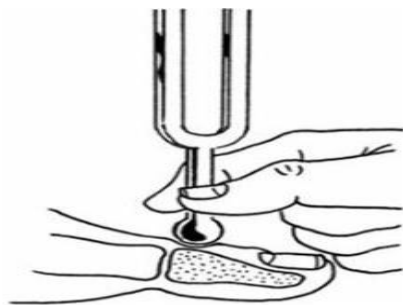


Figura 5. Uso del diapasón para verificar sensibilidad vibratoria. [32]

No obstante, ambos métodos presentan limitaciones: en el caso del monofilamento, su sensibilidad puede variar ampliamente (41 %–93 %), lo que cuestiona su fiabilidad en la detección temprana de neuropatía [33]; mientras que el diapasón, aunque ampliamente empleado, depende en gran medida de la experiencia del profesional y suele tener una precisión moderada, lo que limita su capacidad para identificar alteraciones sutiles en etapas iniciales de la enfermedad [34]. A ello se suma que, en la práctica clínica, el grado de conocimiento y entrenamiento en la aplicación de estas pruebas puede ser insuficiente [10], reduciendo aún más su efectividad como métodos de detección temprana. Por estas razones, suelen emplearse en conjunto con otras valoraciones clínicas para lograr una evaluación más completa del pie en personas con diabetes.

Ambos factores, neuropatía e isquemia, actúan de manera sinérgica. Mientras la neuropatía facilita el desarrollo de lesiones al no percibir el daño, la isquemia limita la respuesta del cuerpo frente a dichas lesiones, complicando el proceso de curación y elevando el riesgo de complicaciones graves como la amputación.

2.2. BANCO DE DATOS

En el desarrollo de sistemas de detección automatizada mediante visión artificial y aprendizaje profundo, el uso de una base de datos adecuada es un componente fundamental. Una base de datos se considera adecuada cuando posee un gran número de sujetos, está balanceada en cuanto a las clases que contiene, las imágenes se capturan bajo condiciones controladas y uniformes, y presenta suficiente variabilidad para que el modelo aprenda correctamente las características relevantes. Además, debe incluir datos de alta calidad, libres de ruido o distorsiones significativas, y contar con una anotación precisa realizada por expertos, garantizando que los modelos entrenados puedan generalizar correctamente a nuevos casos.

Entre las bases de datos más utilizadas en los antecedentes previamente mencionados para la detección del pie diabético mediante visión artificial se encuentra la proporcionada por el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) [20]. Esta base de datos contiene 1,220 imágenes de pacientes con pie diabético y 450 imágenes de personas sanas, todas previamente segmentadas. Además, incluye archivos .csv que registran los valores de temperatura asociados a cada píxel e información complementaria de cada paciente, como peso, altura, edad y género. Sin embargo, esta base de datos fue descartada para este trabajo, ya que solo contiene imágenes térmicas, mientras que el enfoque propuesto requiere la fusión de imágenes en el espectro visible e infrarrojo para mejorar las métricas del sistema.

Dado este enfoque multimodal, se optó por utilizar la base de datos STANDUP [21], la cual ofrece imágenes en ambos espectros (visible e infrarrojo, este último sin formato radiométrico, solo intensidad relativa), proporcionando una base sólida para el desarrollo y evaluación de modelos de fusión de imágenes. A continuación, se presentan las características principales de esta base de datos y su relevancia para el presente trabajo.

2.2.1. Base de datos STANDUP

La base de datos STANDUP, desarrollada por la comisión europea, consta de 415 imágenes en el espectro visible de la planta del pie, con 125 imágenes de sujetos sanos y 290 imágenes de sujetos diabéticos, que, para cada uno de ellos también cuenta con sus correspondientes imágenes térmicas. Aunque todas las imágenes infrarrojas fueron capturadas con la misma cámara FLIR ONE Pro, al realizarse el estudio en dos laboratorios distintos, uno en Francia y otro en Perú, se observaron diferencias en las dimensiones de las imágenes entre las clasificaciones, con tamaños que incluyen 480×640 píxeles y 514×385 píxeles. En la documentación oficial [21], no se especifican las configuraciones de captura ni otros parámetros técnicos relevantes (por ejemplo, emisividad, rango de temperatura, distancia de la cámara al sujeto o condiciones ambientales), por lo que no es posible determinar con certeza la causa de estas variaciones.

El protocolo de reclutamiento y adquisición se basa en adultos sanos, sin diabetes y sin problemas en los pies, siendo esto para el caso de la categoría no diabéticos y para el caso contrario, se

reclutaron pacientes que acudían con frecuencia a la unidad de diabetes diagnosticados con esta enfermedad, excluyendo a pacientes con úlceras o amputaciones.

El campo de visión para la toma de fotografías fue de $40 \times 40 \text{ cm}^2$: 30 cm para incluir la planta del pie, más un margen de 5 cm en la parte superior e inferior. En la otra dirección, 40 cm para abarcar ambos pies a lo ancho, incluyendo el mismo margen de 10 cm para cada pie, tal y como se muestra en la figura 6.

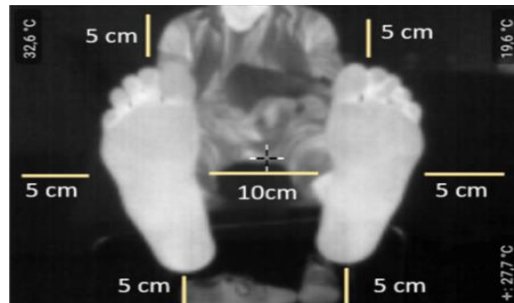
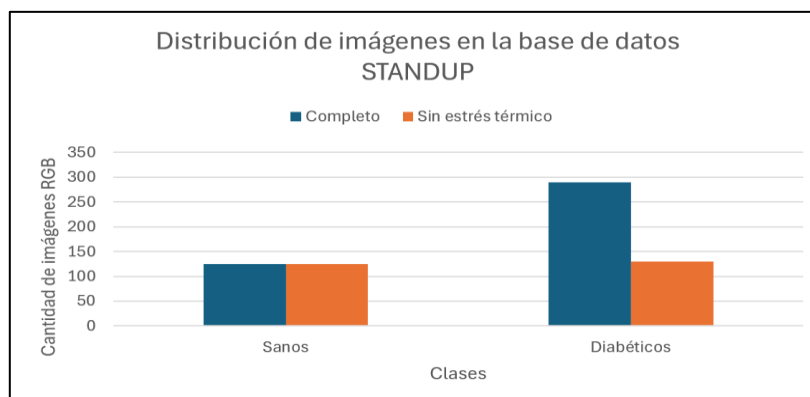


Figura 6. Márgenes utilizadas durante la adquisición de imágenes térmicas del pie plantar. [21]

Para cada sujeto de la categoría diabéticos, se adquirieron imágenes en un tiempo inicial sin alteraciones, y, luego fueron sometidos a una prueba de estrés térmico, sumergiendo los pies en agua fría durante 1 minuto, retirando los pies del recipiente y esperando 10 minutos para una nueva toma fotográfica. Sin embargo, el análisis presente en este trabajo se centra únicamente en las imágenes adquiridas en el periodo inicial, es decir, no se consideraron los termogramas correspondientes a sujetos sometidos a la prueba de estrés por frío. Esto se debe a que, al momento de implementar el sistema de detección, se busca que sea lo menos invasivo posible para los pacientes, evitando replicar situaciones que puedan causarles incomodidad o riesgo, como el estrés térmico. Como consecuencia, la cantidad de sujetos diabéticos considerados para el estudio se reduce aproximadamente a la mitad, quedando en un total de 130.

La figura 7 presenta el histograma del conjunto de datos STANDUP, comparando la cantidad de imágenes disponibles inicialmente y aquellas efectivamente utilizadas tras excluir las capturas posteriores a la prueba de estrés.

Figura 7. Distribución del conjunto de datos STANDUP



Fuente: Elaboración propia.

El grupo diabético se encuentra subclasificado a su vez en tres directorios: R0 correspondió a individuos no neuropáticos y no isquémicos; R1 incluyó individuos neuropáticos sin isquemia; y

R2 estaba compuesto por individuos con isquemia. Para efectos comparativos con antecedentes, se dedicó la mayor parte del tiempo a desarrollar la clasificación en solo dos categorías: sano y pie diabético.

A continuación, en la figura 8 se ilustra ejemplos de las imágenes pertenecientes a la base de datos STANDUP.

Figura 8. Ejemplos de fotografías y termografías de la base de datos STANDUP.



Fuente: Elaboración propia.

En este estudio no se aplicaron técnicas de *data augmentation*. La decisión responde, en primer lugar, a la necesidad de mantener comparabilidad con estudios de referencia, como el de Arteaga et al. [19], quienes trabajaron con la base de datos STANDUP [21] sin aplicar dicho proceso. Además, tras la exclusión de las imágenes con estrés térmico, el conjunto resultante presentó un balance adecuado entre categorías, por lo que no se consideró necesario generar datos adicionales para compensar desbalances. Finalmente, dado que el objetivo del sistema es evaluar la detección en condiciones lo más cercanas posible a la práctica clínica, se evitó introducir transformaciones artificiales que pudieran alterar la distribución natural de los patrones térmicos y morfológicos.

2.3 VISIÓN ARTIFICIAL

La visión artificial se centra en el desarrollo de métodos y aplicaciones para entender, interpretar, interactuar y reconstruir escenas 3D estáticas o dinámicas a partir de una o múltiples imágenes en términos de las propiedades de los objetos presentes en la escena y la tarea para la cual se concibe [35]. Además, permite su implementación con aplicaciones que van desde el control de calidad hasta el guiado de robots gracias a la captura, procesamiento y entendimiento de imágenes estáticas o dinámicas. Su desarrollo se nutre de áreas como la computación, electrónica, matemáticas y ciencias cognitivas[35], [36].

A raíz de su complejidad y amplio campo de aplicación, la visión artificial está rodeada de conceptos fundamentales para el desarrollo de este trabajo, tales como:

2.3.1. Preprocesamiento de datos

Consiste en aplicar una serie de técnicas destinadas a mejorar la calidad de un conjunto de datos, o en el caso de este trabajo, la imagen, facilitando la extracción de información relevante en etapas

posteriores como la segmentación, el reconocimiento o la clasificación, asegurando que los datos sean consistentes, limpios y adecuados para el análisis computacional [37]. Esta etapa previa puede incluir procesos como la conversión a escala de grises, la normalización, la eliminación de ruido, el ajuste de contraste y el realce de bordes, entre otros que dependen netamente de la aplicación.

2.3.2. Segmentación de imágenes.

La segmentación de imágenes es una etapa fundamental en la visión artificial, cuyo objetivo es dividir una imagen en regiones homogéneas que faciliten la identificación y análisis de objetos o áreas de interés [38]. Es especialmente relevante en aplicaciones médicas, donde permite aislar estructuras anatómicas específicas para su posterior estudio o diagnóstico.

Existen diversas técnicas de segmentación. Entre las más clásicas se encuentran la umbralización, que separa regiones con base en valores de intensidad de los píxeles [39]; la detección de bordes, que identifica límites abruptos entre regiones usando operadores como Sobel o Canny; y los métodos basados en regiones, que agrupan píxeles contiguos con características similares. También hay herramientas más avanzadas relacionadas con el aprendizaje profundo que han permitido desarrollar arquitecturas especializadas para la segmentación, como U-Net, Mask R-CNN o modelos de segmentación semántica. Incluso algoritmos originalmente diseñados para detección, como YOLO (You Only Look Once) [1].

En el contexto del presente trabajo, algunos conceptos complementarios resultan especialmente relevantes dentro del proceso de segmentación:

2.3.2.1. Región de interés (ROI)

La Región de Interés hace referencia a un área específica dentro de una imagen que se selecciona por contener la información más relevante para la tarea de análisis. Esta delimitación permite focalizar el procesamiento en zonas clave, excluyendo elementos irrelevantes que podrían introducir ruido o aumentar el costo computacional, lo cual resulta especialmente útil en aplicaciones médicas, industriales o de visión por computadora.

La identificación de la ROI puede apoyarse en diversas herramientas analíticas, como el uso de histogramas de intensidad, que permiten reconocer rangos de valores significativos en la imagen y orientar el recorte de la zona más informativa.

2.3.2.2. Histograma de una imagen

En imágenes, consiste en calcular la distribución de frecuencias de los niveles de intensidad de los píxeles. Es una herramienta fundamental en procesamiento digital de imágenes, ya que permite visualizar cuántos píxeles tienen cada nivel de brillo o color y describiendo la apariencia global de la imagen, aunque no necesariamente proporciona información sobre el contenido de la imagen [36].

En este trabajo, se empleó un enfoque basado en histogramas de distribución de intensidad. Inicialmente, se calcula un umbral adaptativo para binarizar la imagen térmica en función de su media (μ) y su desviación estándar (σ), según la expresión:

$$T = \mu + 0.15 \times \sigma \quad (1) \quad [40]$$

Donde T es el umbral utilizado para separar las áreas de mayor temperatura (píxeles de mayor intensidad). Esto permite obtener una imagen binaria, donde los píxeles con intensidad superior a T son considerados como regiones de interés.

A partir de esta imagen binaria, se generan histogramas de distribución de calor en las direcciones horizontal y vertical. Estos histogramas se calculan sumando los valores de los píxeles en cada columna y fila, respectivamente:

- Histograma vertical:

$$H_v(j) = \sum_{i=1}^M B(i, j) \quad (2)$$

- Histograma horizontal:

$$H_h(i) = \sum_{j=1}^M B(i, j) \quad (3)$$

Donde B(i, j) representa la imagen binaria, M y N son las dimensiones de la imagen. Estos histogramas permiten identificar las regiones donde se concentra el calor, correspondientes a los pies sobre el resto de la imagen, y establecer límites óptimos que garantizan una correcta segmentación de la ROI.

2.3.2.3. Matiz Saturación Valor (HSV) y YIQ

El espacio de color HSV (Hue, Saturation, Value) es un modelo ampliamente utilizado en visión artificial y procesamiento de imágenes, debido a que representa el color de una forma más cercana a la percepción humana que el modelo tradicional RGB. Mientras que RGB describe el color como una combinación aditiva de los componentes rojo, verde y azul, HSV descompone el color en tres parámetros: matiz (Hue), saturación (Saturation) y valor (Value). El matiz hace referencia al tipo de color, medido en grados dentro de un círculo cromático como se observa en la figura 9, que abarca desde el rojo (0°), pasando por el verde (120°), hasta el azul (240°) y retornando al rojo en 360°. La saturación indica qué tan puro o intenso es un color, donde una saturación baja corresponde a colores apagados o cercanos al gris, y un valor alto representa colores vivos. Finalmente, el valor se asocia con el brillo o la luminosidad del color, permitiendo distinguir colores claros de oscuros, independientemente de su tonalidad [36].

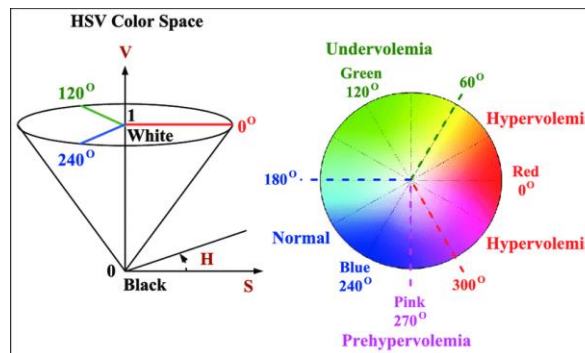


Figura 9. Espacio del color HSV. [41]

Una de las principales ventajas del espacio HSV en aplicaciones de visión artificial es la posibilidad de aislar la información del color del componente de iluminación, lo que facilita tareas como la segmentación, la detección de objetos y el seguimiento por color en condiciones de iluminación variables. En este sentido, el canal Hue es especialmente útil para identificar regiones específicas

de color, como es de nuestro interés separar el color piel del resto de la imagen, mientras que los canales de Saturación y Valor pueden ser utilizados para filtrar zonas con colores poco definidos o mal iluminadas. Esto proporciona una mayor robustez frente a sombras, reflejos y cambios de intensidad lumínica, lo que resulta esencial en entornos no controlados.

A continuación, se muestran las expresiones matemáticas empleadas para calcular los componentes HSV de [36].

$$H = \begin{cases} \text{Indefinido} & \text{si } \max(R, G, B) = 0 \\ \frac{\pi}{3} \times \frac{G-B}{\max(R, G, B) - \min(R, G, B)} & \text{si } R = \max(R, G, B) \text{ y } G \geq B \\ \left(\frac{\pi}{3} \times \frac{G-B}{\max(R, G, B) - \min(R, G, B)} \right) + 2\pi & \text{si } R = \max(R, G, B) \text{ y } G < B \\ \left(\frac{\pi}{3} \times \frac{G-B}{\max(R, G, B) - \min(R, G, B)} \right) + \frac{2\pi}{3} & \text{si } G = \max(R, G, B) \\ \left(\frac{\pi}{3} \times \frac{G-B}{\max(R, G, B) - \min(R, G, B)} \right) + \frac{4\pi}{3} & \text{si } B = \max(R, G, B) \end{cases} \quad (4)$$

$$S = \begin{cases} 0 & \text{si } \max(R, G, B) = 0 \\ 1 - \frac{\min(R, G, B)}{\max(R, G, B)} & \text{en otro caso} \end{cases} \quad (5)$$

$$V = \max(R, G, B) \quad (6)$$

Por otro lado, el propósito de YIQ es separar la información de luminancia (brillo) de la crominancia (color), permitiendo así que las señales de televisión a color sean compatibles con televisores en blanco y negro, que solo interpretan la información de brillo. La conversión del espacio RGB al espacio YIQ se realiza mediante una transformación lineal, que permite expresar cada componente de YIQ como una combinación ponderada de los valores rojo (R), verde (G) y azul (B) de un píxel [36]. La fórmula para esta conversión es la siguiente:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (7) \quad [42]$$

Donde los valores de entrada (R, G, B) están normalizados en el rango de 0 a 1. Este modelo está compuesto por tres componentes. La primera, Y, corresponde a la luminancia, es decir, la intensidad o brillo de la imagen, lo que percibiría un dispositivo monocromático. Las otras dos componentes, I (In-phase) y Q (Quadrature), codifican la información de color: la componente I representa principalmente variaciones entre tonos anaranjados y azulados, mientras que Q representa diferencias entre colores violeta y verde [42]. El ojo humano es más sensible a los detalles en la luminancia que a los de crominancia, por lo que I y Q se pueden transmitir con menor resolución, lo cual reduce el uso de ancho de banda sin afectar significativamente la percepción visual.

Diversos trabajos han demostrado la utilidad de los espacios de color HSV y YIQ en tareas de segmentación anatómica, especialmente cuando se requiere identificar regiones con tonalidades propias de la piel humana. En particular, Marcial Basilio et al. [43], [44] desarrollaron una técnica

para la detección de imágenes con contenido explícito basada en los modelos de color HSV y YCbCr, con el objetivo de extraer regiones cutáneas en contextos complejos y con variaciones lumínicas, resaltando la eficacia del modelo HSV para preservar la información del tono (matiz) separándola de la iluminación. Por su parte, Ceballos et al. [44] propusieron un clasificador de páginas web con contenido explícito basado en el análisis del contenido visual, utilizando también la segmentación del color de piel como etapa clave, apoyándose en modelos como YIQ, el cual separa la información de luminancia (Y) de la cromática (I y Q), permitiendo una mayor robustez frente a cambios de iluminación y facilitando el aislamiento de tonalidades características de la piel. Aunque ambos estudios están enfocados en contextos distintos al clínico, su metodología demuestra el potencial de estos espacios de color para segmentar con precisión zonas de color piel.

2.3.2.4. Coeficiente de Dice

Este coeficiente es ampliamente utilizado en tareas de segmentación médica debido a su sensibilidad a la superposición entre las regiones segmentadas. Sirve para cuantificar el grado de similitud entre dos segmentaciones, evaluando cuán bien una segmentación automática reproduce una segmentación de referencia (habitualmente manual) [45]. Matemáticamente, el coeficiente de Dice se define como:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (8)$$

donde A y B representan las dos máscaras binarias a comparar, siendo A la máscara de referencia (en este estudio, generada manualmente) y B la máscara generada por el algoritmo propuesto para este trabajo. El numerador corresponde al doble de la cantidad de píxeles que coinciden entre ambas segmentaciones, mientras que el denominador representa la suma total de píxeles en cada una de las máscaras [45]. Un valor de 1 indica una coincidencia perfecta, mientras que un valor de 0 representa una ausencia total de superposición.

2.3.3. Fusión de imágenes

La fusión de imágenes es una herramienta importante en el área de la visión artificial que combina información de múltiples imágenes, usualmente de diferentes espectros, para obtener una representación más completa y detallada de la escena en análisis [46]. Esta técnica ha encontrado aplicaciones en diferentes campos, como la medicina, la vigilancia, la agricultura de precisión y la exploración geológica. Por ejemplo, en la imagenología médica, la fusión de imágenes de resonancia magnética (MRI) y tomografía por emisión de positrones (PET) permite una visualización combinada de la anatomía y el metabolismo [47], mejorando el diagnóstico de enfermedades como el cáncer. En la vigilancia, la fusión de imágenes ópticas y térmicas se implementa cada vez con mayor frecuencia para detectar fallas en infraestructuras como puentes y edificios [48], permitiendo realizar mantenimientos preventivos para evitar catástrofes.

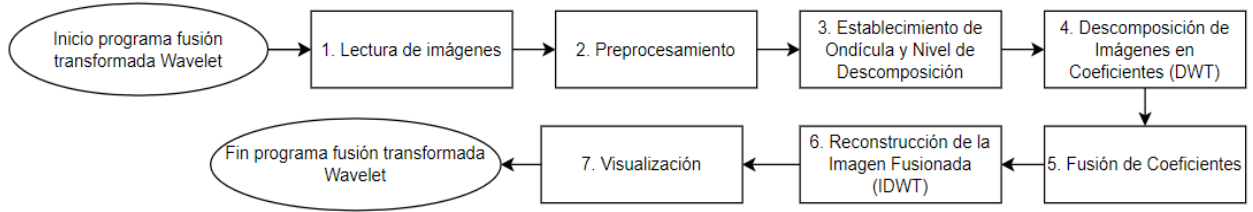
En el presente estudio, se propone explorar los beneficios de la fusión de imágenes como técnica clave para mejorar la detección del pie diabético, partiendo del principio de que la combinación de distintas representaciones de una misma escena puede enriquecer la información disponible y facilitar un análisis más preciso. Para ello, se decidió consultar, implementar y evaluar diversos métodos de fusión en base a los antecedentes descritos previamente, con el fin de identificar cuál de ellos ofrece un mejor desempeño en la integración de características relevantes para la delimitación de la región plantar.

2.3.3.1 Métodos de fusión Basados en la Transformada

Los métodos basados en la transformada son más avanzados y utilizan transformadas matemáticas para dividir las imágenes en componentes de diferente índole antes de fusionarlas, como frecuencia, tiempo-frecuencia o espacio-escala.

Un ejemplo prominente es la transformada de ondícula discreta o de su inglés conocido, transformada Wavelet Discreta (DWT), que, a diferencia de la transformada de Fourier, donde solo provee información de frecuencia, la transformada de ondículas mantiene la información de tiempo o posición espacial, lo que posibilita la fusión de características a diferentes niveles de resolución [49], permitiendo conservar detalles, minimizar errores y mejorar la calidad de la imagen fusionada, lo que la convierte en una herramienta apropiada para aplicaciones médicas donde los detalles finos son críticos. El procedimiento se ilustra en la figura 10:

Figura 10. Procedimiento para fusión de imágenes utilizando Transformada Wavelet.



Fuente: Elaboración propia.

En la DWT, las imágenes se dividen en diferentes niveles de escala y posición mediante filtros pasa altas y pasa bajas, creando coeficientes de detalle y aproximación [49]. Estos parámetros son discretos, dados generalmente en potencias de dos, lo que permite una descomposición por niveles de la imagen con un número determinado de coeficientes en cada jerarquía. Además, el proceso de la transformada incluye elaborar submuestreo, reduciendo a la mitad el número de coeficientes en cada nivel de descomposición, lo que permite una representación más compacta y detallada de la imagen [47].

Para fusionar las imágenes desarticuladas, los coeficientes se combinan utilizando métodos matemáticos como máximo, mínimo o promedio, realizando este proceso en cada sub-banda para obtener los coeficientes fusionados, que, por último, y utilizando la transformada inversa de ondícula discreta (IDWT) permite reconstruir la imagen fusionada [49]. La fórmula de la DWT está dada como:

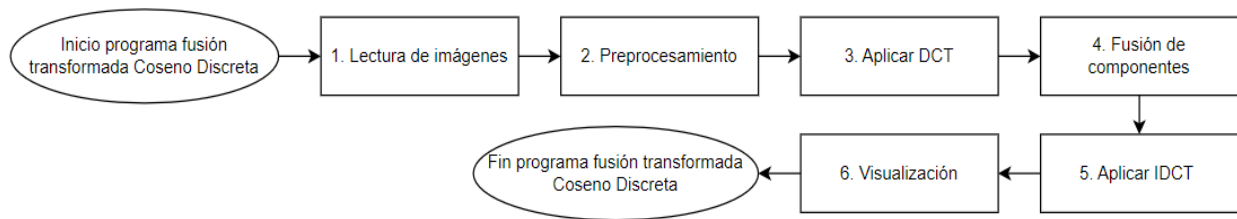
$$W_{j,k} = \sum_{n=0}^{N-1} x[n] \psi_{j,k}^*[n] \quad (9) \quad [49]$$

Donde j es el nivel de descomposición, k es el índice de la posición y $\psi_{j,k}^*[n]$ es la función de ondícula escalada y trasladada $\psi_{j,k}[n] = 2^{-j/2} \psi(2^{-j}n - k)$ [49].

Un método similar al anterior es la transformada de coseno discreta (DCT). La DCT es un método matemático ampliamente utilizado en el procesamiento de señales e imágenes donde convierte la variable de entrada desde el dominio espacial al dominio frecuencial. Su diferencia con la DWT radica en la utilización única de cosenos, otorgando la propiedad de concentrar la mayor parte de

la energía de una señal en unos pocos coeficientes de frecuencia baja, lo que permite una representación eficiente y una compresión efectiva [50].

Figura 11. Procedimiento para fusión de imágenes utilizando transformada del coseno discreta.



Fuente: Elaboración propia.

Siguiendo el flujo de la figura 11, para el trabajo en cuestión se requiere hacer uso de la DCT bidimensional, su fórmula está dada en la ecuación 10:

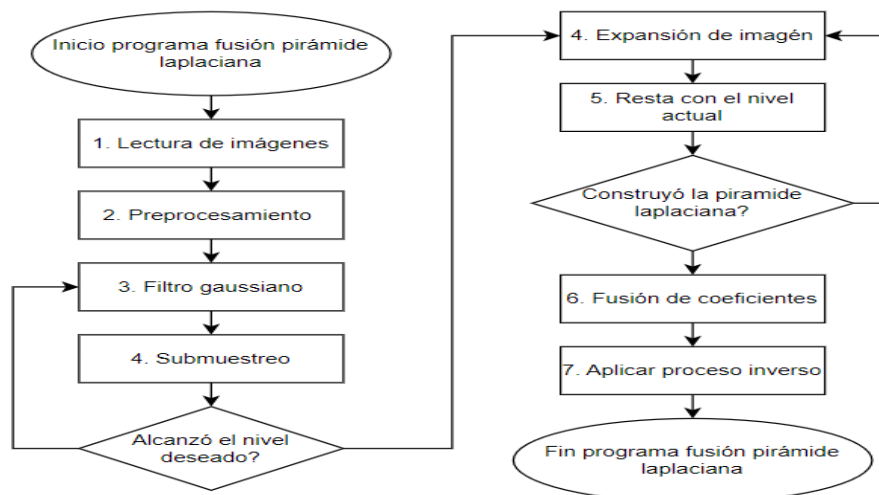
$$X_{k_1, k_2} = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1, n_2} \cos \left[\frac{\pi}{N_1} \left(n_1 + \frac{1}{2} \right) k_1 \right] \cos \left[\frac{\pi}{N_2} \left(n_2 + \frac{1}{2} \right) k_2 \right] \quad (10)$$

Donde X_{k_1, k_2} es el coeficiente DCT correspondiente a los índices de frecuencia k_1, k_2 . El valor del píxel en la posición n_1 y n_2 se considera en x_{n_1, n_2} y las dimensiones de la imagen N_1 y N_2 [50]. Asimismo, una vez fusionados los coeficientes obtenidos mediante la DCT bidimensional, se hace necesario aplicar la transformada inversa de coseno discreta para reconstruir la imagen en el dominio espacial. La fórmula correspondiente para la IDCT bidimensional está dada por la ecuación 11:

$$x(n_1, n_2) = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \alpha(k_1) \alpha(k_2) X(k_1, k_2) \cos \left[\frac{\pi}{N_1} \left(n_1 + \frac{1}{2} \right) k_1 \right] \cos \left[\frac{\pi}{N_2} \left(n_2 + \frac{1}{2} \right) k_2 \right] \quad (11)$$

Otro método basado en la descomposición de imágenes es el propuesto por Burt y Adelson conocida como transformada de la pirámide laplaciana (TPL), el cual sigue un flujo indicado en la figura 12:

Figura 12. Procedimiento para fusión de imágenes utilizando transformada pirámide laplaciana



Fuente: Elaboración propia.

Previamente a la construcción de la pirámide Laplaciana se debe crear una pirámide Gaussiana, para ello, se parte de la imagen original, la cual se denota como G . Esta imagen se suaviza mediante un filtro Gaussiano, lo que atenúa las altas frecuencias de la imagen. Una vez filtrada se submuestra por n hacia abajo, reduciendo su tamaño para obtener el siguiente nivel k de la pirámide [51]. El proceso previo se repite hasta que la imagen deja de aportar información significativa en niveles más profundos, es decir, cuando la variación entre niveles consecutivos se vuelve despreciable. Por consiguiente, se produce una serie solida de niveles suavizados y submuestreados G_k que forman la pirámide Gaussiana.

$$G_k = [w * G_{k-1}]_{\downarrow n} \quad (12)$$

A cada G_k de la pirámide Gaussiana le es restado el nivel original correspondiente a la pirámide, considerando la convolución por el núcleo generador w y un submuestreo hacia arriba, obteniendo así los niveles de la pirámide Laplaciana L_k [52]. Este procedimiento de expansión y diferencia permite mantener los detalles más relevantes en cada nivel de la imagen original.

$$L_k = G_k - 4w * [G_{k+1}]_{\uparrow n} \quad (13)$$

En la fusión de imágenes, la pirámide Laplaciana se construye para cada imagen a fusionar, y luego se combinan las pirámides resultantes utilizando diversas estrategias como las mencionadas para el caso de DWT (máximo, mínimo, promedio, etc.). Realizada la fusión, para reconstruir la imagen resultante F_K se requiere invertir el proceso hasta obtener F a partir de la ecuación 14:

$$F_K = L_k + 4w * [F_{k+1}]_{\uparrow n} \quad (14)$$

2.3.3.2. Métodos de fusión Espaciales

Los métodos espaciales son los más simples y directos, operando directamente sobre los píxeles de las imágenes. Ejemplos de estos métodos incluyen la fusión por promedio, fusión ponderada y la fusión basada en la máxima o mínima intensidad [53]. Estos métodos son rápidos y fáciles de implementar, pero pueden no preservar bien los detalles importantes ni las características específicas de las imágenes fusionadas [47].

Para trabajar con el método promedio, se debe realizar una suma de los valores de cada píxel $I_k(i, j)$ correspondiente a la k -ésima imagen, seguida de una división por el número total de imágenes n involucradas tal como se muestra en la ecuación 15:

$$F(i, j) = \frac{1}{n} \sum_{k=1}^n I_k(i, j) \quad (15)$$

donde $F(i, j)$ representa el valor del píxel fusionado en la posición espacial (i, j) .

La fusión ponderada es similar a la ecuación 15, consiste en seleccionar diferentes pesos w_k a cada imagen antes de fusionarlas. Estos pesos pueden basarse en la importancia o calidad de cada imagen. Luego, se multiplica cada píxel $I_k(i, j)$ por su peso correspondiente y se suman los resultados como se ilustra en la ecuación 16 para el caso de dos espectros:

$$F(i, j) = w_1 * I_1(i, j) + w_2 * I_2(i, j) \quad (16)$$

Para los casos de máxima y mínima intensidad basta con seleccionar dichos valores de cada una de las imágenes a fusionar:

$$F(i, j) = \max\{I_1(i, j), I_2(i, j), \dots, I_n(i, j)\} \quad (17)$$

2.3.3.3. Métricas evaluativas de fusión de imágenes

La evaluación de los métodos de fusión de imágenes en el espectro visible e infrarrojo implementados, se llevó a cabo utilizando una combinación de métricas cuantitativas y evaluaciones visuales con el fin de determinar la efectividad de cada método al realzar las características relevantes para la detección del pie diabético.

Autores como Orellana, Rodríguez [47] y Marín [49] concuerdan en el uso de métricas de evaluación como la relación señal a ruido (PSNR), error cuadrático medio normalizado (NMSE) y coeficiente de correlación (CC), siendo estas implementadas en la fusión de imágenes médicas permitiendo evaluar la calidad de la imagen fusionada, abarcando aspectos como la preservación de la información original, la similitud con las imágenes fuente y la cantidad de ruido introducido.

Para evaluar cuantitativamente el desempeño de los métodos de fusión, se seleccionaron varias métricas estándar en el campo de la visión artificial. Estas métricas permiten comparar objetivamente la calidad de las imágenes fusionadas con respecto a ciertos criterios predefinidos.

- Error Cuadrático Medio Normalizado (NMSE)

Es una versión normalizada del MSE el cual cuantifica la diferencia promedio al cuadrado entre los valores de los píxeles de la imagen original G y la imagen fusionada F , dividiendo por la varianza de los valores de los píxeles de la imagen original [49]. Esta normalización reduce la dependencia a la escala de los valores de los píxeles, permitiendo conocer cuánta información original se ha conservado después del proceso de fusión:

$$MSE = \sum_{i=1}^X \sum_{j=1}^Y (G(i, j) - F(i, j))^2 \quad (18)$$

$$NMSE = \frac{\sum_{i=1}^X \sum_{j=1}^Y (G(i, j) - F(i, j))^2}{\sum_{i=1}^X \sum_{j=1}^Y (G(i, j))^2} \quad (19) \quad [49]$$

Donde X y Y son las dimensiones de las imágenes.

- Relación Señal a Ruido de Pico (PSNR)

En el contexto de la fusión de imágenes en el espectro visible e infrarrojo, un valor alto de PSNR indica menor distorsión, lo cual es fundamental para garantizar que la imagen fusionada conserve adecuadamente la información útil. Esto resulta especialmente importante en tareas posteriores como la extracción de características, donde una imagen con bajo ruido permite obtener descriptores más fiables, y en la clasificación, donde se minimiza el riesgo de entrenar modelos con datos erróneos o degradados [54]. La PSNR se calcula mediante la siguiente expresión:

$$PSNR = 20 \cdot \log_{10} \frac{MAX}{\sqrt{MSE}} \quad (20) \quad [54]$$

Donde MAX representa el valor máximo posible de un píxel en la imagen.

- Coeficiente de correlación (CC)

Permite la medición de la similitud estructural entre las imágenes originales y fusionada, ayudando a evaluar si las características clave o relevantes de la fuente original se han mantenido en la imagen fusionada [49], [55]. Su rango de valores está comprendido de cero a uno, donde 1 indica una alta similitud, mientras que un valor cercano a 0 indica poca o ninguna correlación.

$$CC = \frac{\sum_{i=1}^X \sum_{j=1}^Y (G(i,j) - \mu_G)(F(i,j) - \mu_F)}{\sqrt{\sum_{i=1}^X \sum_{j=1}^Y (G(i,j) - \mu_G)^2 \sum_{i=1}^X \sum_{j=1}^Y (F(i,j) - \mu_F)^2}} \quad (21) \quad [55]$$

El numerador calcula la covarianza entre las dos imágenes a partir de las medias μ calculadas de cada imagen. Esto mide cómo varían conjuntamente las intensidades de los píxeles en ambas imágenes, mientras que el denominador calcula el producto de las varianzas de las dos imágenes [55].

2.4. RECONOCIMIENTO DE PATRONES

El reconocimiento de patrones es una rama de la inteligencia artificial que busca que las máquinas puedan identificar estructuras o regularidades en los datos. Su objetivo es clasificar automáticamente objetos en diferentes categorías, basándose en mediciones realizadas sobre ellos, para que puedan interpretar y organizar información del mundo real de forma precisa, y, de acuerdo con el Ph.D. Loaiza, H [56] este proceso se desarrolla en dos etapas básicas: entrenamiento y predicción.

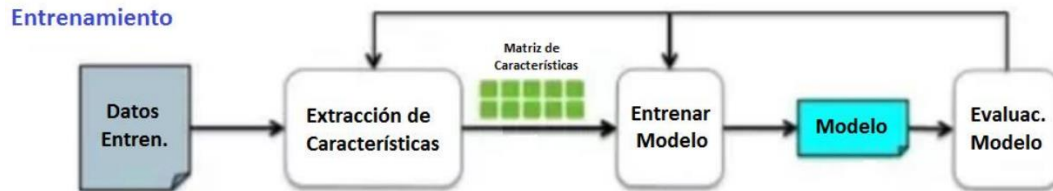


Figura 13. Primera etapa del reconocimiento de patrones. [56]

En la etapa de entrenamiento como se ilustra en la figura 13, el sistema aprende a partir de un conjunto de datos conocidos mediante la extracción de características relevantes, la construcción de una matriz de características y el ajuste de un modelo.

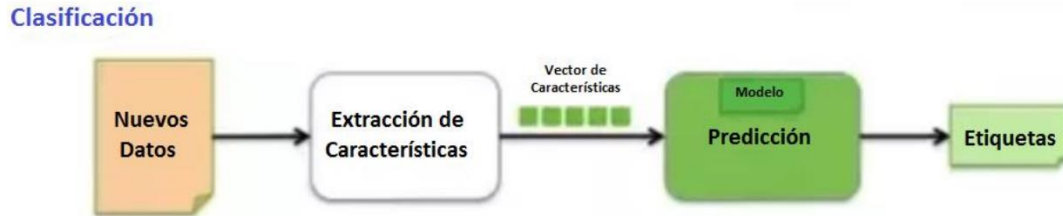


Figura 14. Segunda etapa del reconocimiento de patrones. [56]

Una vez entrenado y evaluado, este modelo se utiliza en la etapa de predicción indicada en la figura 14, donde se recibe nueva información, se extraen sus características y se clasifica automáticamente asignando etiquetas según lo aprendido.

2.4.1. Extracción y selección de características

La extracción consiste en obtener de los datos originales aquellos valores que permiten describir de manera significativa el objeto de interés, formando patrones para el análisis. Dentro de este proceso son importantes aspectos como la detección de valores que se alejan significativamente del comportamiento general de los datos, conocidos generalmente como outliers [57] o para facilitar la identificación de estos valores atípicos, se utilizan herramientas gráficas como los diagramas de cajas, que permiten visualizar la distribución de los datos, identificar la mediana, los cuartiles y detectar posibles valores extremos, siendo esta una etapa previa a la selección de características.

Por otro lado, la selección de características no es simplemente un filtro básico, sino un proceso más elaborado que busca identificar, dentro del conjunto de características extraídas, aquellas que aportan mayor relevancia para el problema a resolver. Este proceso puede involucrar estrategias de búsqueda exhaustiva o heurística, como los métodos wrapper, que evalúan distintas combinaciones de características en función del desempeño del modelo, o los métodos filter, que se basan en criterios estadísticos independientes del clasificador [58]. El objetivo es reducir la dimensionalidad del problema, eliminando datos redundantes o irrelevantes, lo cual no solo mejora la eficiencia del modelo y facilita su interpretación.

En base a los antecedentes evaluados en este trabajo para la detección del pie diabético como [16] y [19], se identifica el uso recurrente de características relacionadas con propiedades estadísticas y de textura de los datos. Entre las más comunes se encuentran medidas como la entropía, curtosis, asimetría y energía, así como operaciones estadísticas clásicas como el máximo, mínimo, promedio y la desviación estándar.

2.4.1.1. Entropía

La entropía en procesamiento de imágenes es una medida de la cantidad de información o complejidad presente en una imagen. En términos simples, cuantifica cuán impredecible o caótica es la distribución de los niveles de intensidad de los píxeles.

$$H = - \sum_{i=0}^{L-1} p_i * \log_2(p_i) \quad (22) \quad [59]$$

Donde:

- H: es la entropía de la imagen.
- L: es el número de niveles de intensidad posibles (por ejemplo, 256 para una imagen de 8 bits).
- p_i : es la probabilidad de que el nivel de intensidad i ocurra en la imagen. Se calcula el número de píxeles con intensidad i , dividido el número total de píxeles en la imagen.
- El logaritmo se toma en base 2 para expresar la entropía en bits.

2.4.1.2. Curtosis y asimetría

La asimetría y la curtosis son indicadores estadísticos implementados en el análisis de imágenes para describir la forma de la distribución de intensidades, pero se enfocan en aspectos distintos. La asimetría, como su nombre lo indica, se centra en determinar si la distribución es simétrica o tiene tendencia hacia valores más bajos o altos, es decir, si la imagen tiende a ser más clara u oscura dependiendo de la concentración de píxeles en el extremo, con exponencial impar para poder valorar números negativos. Por otro lado, la curtosis es una medida estadística que ilustra la forma en la que está distribuida una variable, en este caso, que tan concentrados están los niveles de intensidad de una imagen de acuerdo con la media y qué tan pronunciados son los picos y colas de la distribución.

$$K = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4 \quad (23) \quad \text{y} \quad S = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3 \quad (24) \quad [60]$$

Donde:

- K: curtosis.
- S: Asimetría.
- n: número total de datos (píxeles).
- x_i : valor de cada píxel.
- μ : media de los valores de intensidad.
- σ : desviación estándar.

2.4.1.3 Energía

La energía es una medida que permite analizar la intensidad global de la información contenida en una imagen a través de $P_{(i,j)}$ que es la probabilidad de ocurrencia de una combinación de niveles de gris. Su fórmula viene dada como la suma de los cuadrados de los valores de intensidad, y guarda relación con la textura y la variación de tonos dentro de una zona en análisis o en toda la imagen. Una imagen con alta energía suele tener muchos detalles o contrastes, mientras que una con baja energía es más uniforme o suave.

$$ASM = \sum_{i,j=0}^N P_{(i,j)}^2 \quad (25) \quad [61]$$

2.4.2. Máquinas de Soporte Vectorial (SVM)

Una vez extraídas las características relevantes de las imágenes, se requiere un clasificador que permita distinguir entre las distintas condiciones evaluadas. Para este propósito, uno de los métodos utilizados es el de Máquinas de Soporte Vectorial (Support Vector Machines, SVM) debido a su amplia implementación en estudios previos relacionados con la detección del pie diabético, como se pudo evidenciar en el apartado 1.2 Antecedentes.

Las SVM son máquinas de aprendizaje que buscan construir un hiperplano óptimo que separe los datos de diferentes clases, maximizando el margen entre ellos. Este enfoque permite una mejor capacidad de generalización, lo que significa que el modelo puede clasificar adecuadamente datos nuevos no vistos durante el entrenamiento [62].

De acuerdo con [62] y [56], el concepto clave detrás de las SVM es encontrar un hiperplano que no solo separe las clases, sino que lo haga con la mayor distancia posible a los ejemplos (vectores de soporte) más cercanos de cada clase. Este principio puede apreciarse en la figura 15, donde se

muestra la representación geométrica de un hiperplano separando dos clases y el margen definido por los vectores de soporte.

Para este caso, al ser dos clases ω_1 y ω_2 , correspondientes a pacientes sanos y diabéticos, y un conjunto de datos $\{(X_k, Z_k)\}_{k=1}^n$, donde $z_k \in \{+1, -1\}$ denota la clase, se realiza un mapeo no lineal $\phi(\cdot)$ a un espacio de características de mayor dimensión:

$$y_k = \phi(X_k) \quad (26)$$

La función discriminante lineal aumentada se define como:

$$g(y) = a^T y \quad (27)$$

Con una regla de decisión:

$$\begin{cases} \text{Clase } \omega_1 \text{ si } a^T y_k > 0 \\ \text{clase } \omega_2 \text{ si } a^T y_k < 0 \end{cases} \quad (28)$$

Para asegurar la separación correcta entre clases garantizando que todos los datos de entrenamiento estén correctamente clasificados y suficientemente alejados del hiperplano, se impone la restricción:

$$z_k \cdot g(y_k) \geq 1, \forall k \quad (29)$$

La distancia r de un patrón transformado y al hiperplano definido por a está dada por:

$$r = \frac{g(y)}{\|a\|} \quad (30)$$

Asumiendo que existe un margen b positivo, el hiperplano se define de forma que, al normalizar por $\|a\|$, el producto $z_k g(y_k)$ sea al menos igual a b ; es decir, que el objetivo es maximizar el margen b , lo cual se traduce en minimizar $\|a\|^2$ bajo la condición anterior. Para evitar soluciones arbitrariamente escaladas, se impone la normalización $\|a\| = 1$.

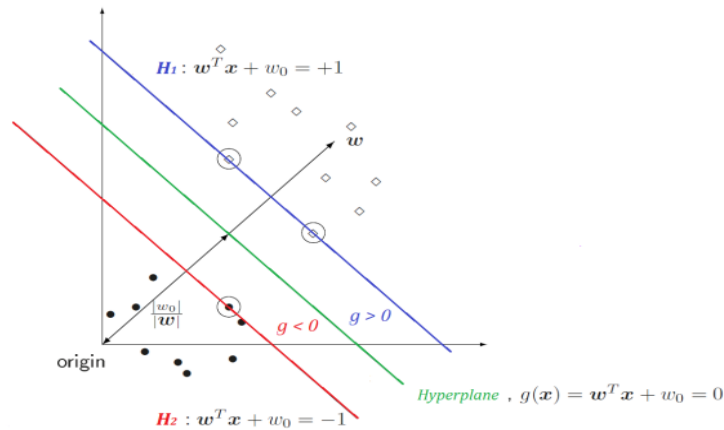


Figura 15. Representación geométrica del principio de las SVM [56]

Cuando los datos no son separables en el espacio original, esta formulación se extiende mediante una función núcleo $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, que permite realizar el cálculo del producto interno sin necesidad de aplicar explícitamente Los núcleos más utilizados tanto en los antecedentes como en la literatura son:

$$\text{Lineal: } K(x_i, x_j) = x_i^T x_j \quad (31) \quad [62]$$

$$\text{Polinomial: } K(x_i, x_j) = (x_i^T x_j + c)^d \quad (32) \quad [16]$$

$$\text{RBF (Radial Basis Function): } K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (33) \quad [56]$$

Aunque la función discriminante $g(y) = a^T y$ permite clasificar correctamente los patrones, su salida no representa una probabilidad. Para solventar esta limitación, se aplicó un ajuste posterior de probabilidades sobre la salida del clasificador SVM mediante un modelo logístico, técnica propuesta por Platt (1999) [63] y ampliamente utilizada en la literatura. El modelo consiste en transformar la distancia al hiperplano de decisión, $f(x)$, en una probabilidad estimada $P(y=1|x)$ utilizando la función logística:

$$P(y = 1 | x) = \frac{1}{1 + \exp(Af(x) + B)} \quad (34) \quad [63]$$

Donde los parámetros A y B se ajustan minimizando el error logarítmico negativo (log-loss) sobre un conjunto de validación. Esta probabilidad se utilizó posteriormente para implementar una estrategia de decisión con opción de rechazo, asignando a la clase "rechazada".

2.4.3. Máquinas de aprendizaje superficial

En este trabajo se implementaron varios clasificadores estadísticos y redes neuronales artificiales en paralelo con las Máquinas de Vectores de Soporte (SVM) con el fin de comparar su desempeño al clasificar entre sujetos sanos y pacientes con indicios de pie diabético. Al aplicar diferentes métodos sobre las mismas características extraídas de imágenes térmicas y morfológicas, se buscó establecer una comparativa clara que permitiera identificar cuál clasificador ofrecía mejores resultados.

El clasificador Naive Bayes es uno de los más conocidos y utilizados, se basa en el teorema de Bayes y parte de la suposición de independencia condicional entre las variables predictoras. Este modelo estima la probabilidad de pertenencia a una clase dada una muestra de entrada, considerando que las características son estadísticamente independientes entre sí. Su fórmula general es:

$$P(C_k | x) = \frac{P(x|C_k) \cdot P(C_k)}{P(x)} \quad (35)$$

Donde $P(C_k | x)$ es la probabilidad posterior de la clase C_k dado el vector de características x , $P(x | C_k)$ es la verosimilitud, $P(x)$ es la probabilidad total [64]. Este clasificador resulta útil en contextos donde se requiere rapidez en el entrenamiento y toma de decisiones, como en sistemas de diagnóstico temprano con restricciones computacionales.

El Análisis Discriminante Lineal (LDA) busca proyectar los datos en un subespacio que maximice la separación entre clases. Parte del supuesto de que cada clase sigue una distribución gaussiana con igual matriz de covarianza, pero diferente media. Su función discriminante está dada por:

$$\delta_k(x) = x^T \Sigma^{-1} u_k - \frac{1}{2} u_k^T \Sigma^{-1} u_k + \log(\pi_k) \quad (36)$$

donde u_k es el vector de medias de la clase k , Σ es la matriz de covarianza común y π_k es la probabilidad a priori de la clase. Este método es adecuado para problemas en los que las clases se distribuyen linealmente en el espacio de características, como puede ocurrir al analizar ciertas medidas térmicas o morfológicas estandarizadas.

El Análisis Discriminante Cuadrático (QDA) es una generalización de LDA que permite que cada clase tenga su propia matriz de covarianza, lo cual le otorga mayor flexibilidad para modelar fronteras de decisión no lineales [2]. Su función discriminante es:

$$\delta_k(x) = -\frac{1}{2} \log|\Sigma_k| - \frac{1}{2}(x - u_k)^T \Sigma_k^{-1}(x - u_k) + \log(\pi_k) \quad (37)$$

Este modelo es útil cuando se espera que las clases tengan formas de distribución distintas, como es el caso de las respuestas fisiológicas al calor en pies sanos frente a pies con daño neurológico o vascular.

El clasificador K-Nearest Neighbors (KNN) es un modelo no paramétrico que asigna una muestra a la clase más frecuente entre sus K vecinos más cercanos, en función de una métrica de distancia. Es especialmente útil cuando no se desea asumir una distribución previa de los datos. Es importante tener en cuenta que k no debe ser un múltiplo del número M de clases [3], [36].

En conjunto, estos clasificadores ofrecen diferentes perspectivas para abordar el problema de detección del pie diabético, permitiendo analizar tanto la linealidad, la dispersión, como la estructura de correlación entre características, y comparar su desempeño bajo criterios estadísticos.

Por otro lado, las redes neuronales artificiales son modelos inspirados en el funcionamiento del cerebro humano, diseñadas para aprender patrones complejos a partir de los datos. En este trabajo se utilizaron dos tipos de arquitecturas: redes neuronales feedforward (MLP y Patternet) y redes de base radial (RBF), cada una con características particulares que las hacen útiles en tareas de clasificación como la detección de pie diabético.

Las redes neuronales feedforward, entre las cuales se encuentran el Perceptrón Multicapa (MLP) y la Pattern Recognition Network (Patternet), están compuestas por una capa de entrada, una o más capas ocultas y una capa de salida. Cada neurona en una capa está conectada con todas las neuronas de la siguiente, transmitiendo señales en una sola dirección: de entrada, a salida [65]. Este esquema puede observarse en la Figura 16, donde se muestra la estructura general de una red neuronal. Estas redes utilizan funciones de activación no lineales como la sigmoidea tangencial (tansig) para permitir la modelación de relaciones no lineales entre variables.

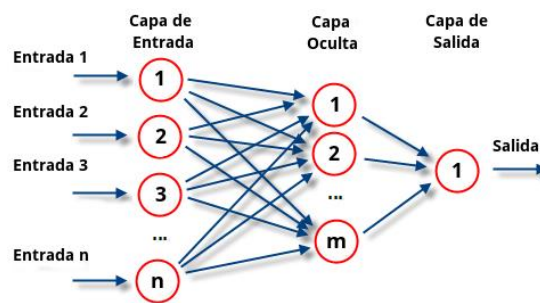


Figura 16. Estructura general de una red neuronal. [66]

Su entrenamiento se realiza mediante el algoritmo de retropropagación del error (backpropagation), ajustando los pesos de conexión para minimizar el error entre la salida de la red y el valor esperado [65]. Estas redes son particularmente útiles cuando se dispone de una gran cantidad de datos etiquetados y se requiere un modelo flexible capaz de aprender patrones complejos.

Por otro lado, la Red de Base Radial (RBF) presenta una estructura distinta. Aunque también cuenta con una capa de entrada, una capa oculta y una de salida, su principio de funcionamiento se basa en la activación de funciones radiales, comúnmente gaussianas, en la capa oculta. Cada neurona de

esta capa responde de manera significativa solo a entradas cercanas a su centro, lo que permite una partición del espacio de características en regiones localizadas [67]. A diferencia de las redes feedforward tradicionales, las RBF son redes con una arquitectura más interpretable y entrenamiento más rápido, ya que los pesos de salida suelen calcularse mediante métodos lineales una vez definidos los centros y radios de las funciones radiales. Esta propiedad las hace especialmente útiles cuando se espera que las clases estén bien separadas en el espacio de características.

2.4.4 Máquinas de aprendizaje profundo

El aprendizaje profundo (*Deep Learning*) es una subárea del aprendizaje automático que emplea redes neuronales artificiales con múltiples capas ocultas para modelar representaciones complejas a partir de grandes volúmenes de datos y su capacidad para aprender características jerárquicas ha permitido avances significativos en tareas como clasificación, segmentación, reconocimiento de patrones e interpretación de imágenes [68].

2.4.4.1 Yolo

YOLO (*You Only Look Once*) es una arquitectura de aprendizaje profundo diseñada para la detección de objetos en imágenes. Se caracteriza principalmente por su alta velocidad y precisión, lo que la hace especialmente adecuada para aplicaciones en tiempo real, como la detección de objetos en video [68].

A diferencia de otros métodos que dividen el proceso en múltiples etapas (como la propuesta de regiones y su posterior clasificación), YOLO aborda el problema como una única tarea de regresión. La red procesa la imagen completa de una sola vez y predice directamente los cuadros delimitadores (*bounding boxes*) y las clases de los objetos presentes. El antecedente descrito en la Sección 1.2.3 [1] demuestra la efectividad de este enfoque en el reconocimiento de estructuras anatómicas, lo que respalda su aplicación en este trabajo para la detección automática de la planta del pie, con el fin de identificar características asociadas al pie diabético.

El método consiste en dividir la imagen de entrada en una cuadrícula de tamaño $S \times S$. Cada celda de esta cuadrícula es responsable de detectar los objetos cuyo centro se encuentra dentro de sus límites. Por cada celda, se predice un número fijo de cuadros, incluyendo las coordenadas del cuadro, una medida de confianza y las probabilidades de pertenecer a cada clase [69]

La red se implementa como una red neuronal convolucional profunda que permite extraer características relevantes de la imagen. En su estructura, se combinan múltiples capas convolucionales para extraer patrones visuales y capas completamente conectadas que generan la salida final, en la cual se encuentran codificadas las predicciones de ubicación y clase para cada región de la imagen [70]

Gracias a este enfoque directo y global, YOLO logra reducir significativamente los errores de detección asociados a falsas positivas y duplicidad de predicciones. Además, su diseño permite alcanzar tasas de procesamiento muy elevadas, lo que la convierte en una solución eficaz para contextos donde el tiempo de respuesta es crítico.

2.4.4.2 Épocas

Una época (epoch) corresponde a un ciclo completo en el que el algoritmo de entrenamiento ha procesado todos los datos disponibles en el conjunto de entrenamiento una vez. Durante cada

época, la red ajusta sus pesos internos mediante el algoritmo de retropropagación para minimizar el error [71]. En la práctica, se utilizan múltiples épocas para garantizar que la red converja hacia una solución óptima. Sin embargo, un número excesivo de épocas puede llevar al sobreajuste (overfitting), donde el modelo aprende demasiado bien los datos de entrenamiento, perdiendo capacidad de generalización.

2.4.4.3. Lotes

El entrenamiento de redes neuronales se realiza comúnmente utilizando mini-lotes (mini-batches). En lugar de actualizar los pesos después de cada muestra (como en el aprendizaje estocástico) o después de procesar todo el conjunto de entrenamiento (como en el aprendizaje por lotes completo), el uso de lotes permite un equilibrio entre la eficiencia computacional y la estabilidad del entrenamiento. El tamaño del lote (batch size) influye directamente en la calidad del gradiente calculado y en la rapidez con que el modelo converge [67].

2.4.4.4 Transferencia de aprendizaje

La transferencia de aprendizaje consiste en utilizar una red previamente entrenada en una gran base de datos (como ImageNet o COCO) y adaptarla a una nueva tarea con una base de datos más pequeña o especializada. Esta técnica permite aprovechar el conocimiento general aprendido por el modelo original, como la detección de bordes, formas o texturas, y aplicar dichas representaciones a un nuevo dominio con menor esfuerzo computacional [72]. Es ampliamente utilizada en problemas de clasificación de imágenes, detección de objetos y segmentación, reduciendo significativamente los tiempos de entrenamiento y mejorando la precisión en contextos con datos limitados. En este trabajo, se utilizó esta técnica al emplear YOLO con pesos preentrenados, lo que facilitó su adaptación al reconocimiento de la planta del pie en imágenes médicas.

2.4.4.5. Validación cruzada

En este trabajo, la validación cruzada se implementó como una etapa posterior a la comparativa entre clasificadores, permitiendo evaluar el rendimiento de cada modelo de manera más robusta y generalizable. Consiste en dividir el conjunto de datos en varias particiones o "folds", entrenando el modelo varias veces y rotando la sección destinada para entrenamiento y la sección para prueba. El objetivo es obtener una estimación más robusta y generalizable del rendimiento del modelo, evitando el sobreajuste a una sola partición de datos [2].

2.5. METRICAS DE DESEMPEÑO DE CLASIFICADORES.

El sistema desarrollado en este trabajo tiene como finalidad detectar automáticamente la presencia de pie diabético mediante el análisis de imágenes en el espectro visible, infrarrojo y su fusión. Dado que se trata de un problema con implicaciones clínicas, las métricas de evaluación deben priorizar no solo la precisión global del modelo, sino especialmente su capacidad para identificar correctamente los casos positivos (pies diabéticos), minimizando los errores de diagnóstico.

2.5.1. Matriz de confusión

La matriz de confusión es una herramienta fundamental para evaluar el rendimiento de un

clasificador binario, ya que permite analizar de forma detallada los aciertos y errores cometidos por el modelo. Esta matriz se compone de cuatro elementos: los verdaderos positivos (TP) representan los casos en los que el modelo predice correctamente la clase positiva; los verdaderos negativos (TN) son los casos en los que el modelo identifica correctamente la clase negativa; los falsos positivos (FP) corresponden a las instancias en las que el modelo predice la clase positiva cuando en realidad pertenecen a la clase negativa; y los falsos negativos (FN) son aquellos casos en los que el modelo predice erróneamente la clase negativa cuando la muestra realmente pertenece a la clase positiva [73].

2.5.2. Métrica de desempeño precisión

Indica qué proporción de las muestras clasificadas como positivas realmente lo son. Es una medida de la calidad de las predicciones positivas [74]. Esta métrica es importante para evitar falsos positivos, que pueden generar ansiedad innecesaria o tratamientos erróneos.

$$\text{Precisión} = \frac{TP}{FP+TP} \quad (38)$$

2.5.3. Métrica de desempeño sensibilidad

Mide la capacidad del clasificador para detectar correctamente las muestras positivas. También se conoce como recall o true positive rate [2]. En el contexto clínico, es fundamental maximizar esta métrica, ya que un falso negativo implica no detectar un caso de pie diabético, lo cual puede retrasar intervenciones médicas oportunas.

$$\text{Sensibilidad} = \frac{TP}{FN+TP} \quad (39)$$

2.5.4. Métrica de desempeño F1 score

Es la media armónica entre precisión y sensibilidad. Es útil cuando se desea un equilibrio entre ambas métricas, especialmente en contextos con clases desbalanceadas [74]. Su uso es adecuado cuando se requiere un balance entre ambas métricas, como ocurre en este trabajo. Además, permite comparar directamente con resultados de trabajos anteriores que también la reportan.

$$F1 \text{ Score} = 2 * \frac{\text{Precisión} * \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \quad (40)$$

2.5.5. Métrica de desempeño exactitud y especificidad.

Representa la proporción de predicciones correctas sobre el total de predicciones [2].

$$\text{Exactitud} = \frac{TP+TN}{TP+TN+FP+FN} \quad (41)$$

Evalúa la capacidad del modelo para identificar correctamente las muestras negativas [2].

$$\text{Especificidad} = \frac{TN}{FP+TN} \quad (42)$$

Aunque la exactitud y la especificidad son métricas ampliamente utilizadas en clasificación, su papel en este trabajo es secundario, para temas comparativos con los antecedentes. La exactitud, definida como la proporción total de predicciones correctas, puede ser engañosa en contextos donde existe un desbalance de clases, como suele ocurrir en problemas médicos donde la cantidad de

casos positivos (pies diabéticos) es menor respecto a los negativos (pies sanos). Un modelo que clasifique correctamente la mayoría de los negativos, pero falle al identificar los positivos, puede mostrar una alta exactitud sin tener un valor clínico real.

Por otro lado, la especificidad, que mide la proporción de negativos correctamente identificados, también es menos prioritaria en este contexto. Si bien es deseable minimizar los falsos positivos, en el ámbito clínico es generalmente más crítico evitar los falsos negativos, es decir, no detectar a un paciente que sí presenta pie diabético.

3. DESARROLLO DEL SISTEMA DE DETECCIÓN DEL PIE DIABÉTICO

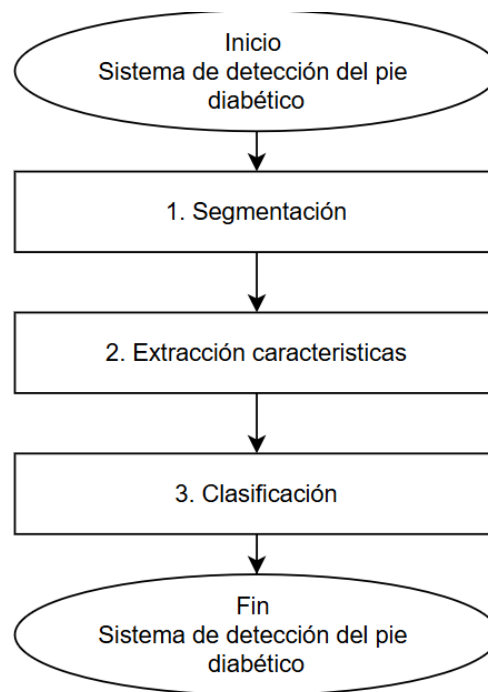


Figura 17. Diagrama de flujo sistema de detección pie diabético.

La Figura 17 muestra la estructura general del sistema desarrollado para la detección del pie diabético, compuesto por tres etapas principales: segmentación, extracción de características y clasificación. En primer lugar, la segmentación se llevó a cabo en dos fases secuenciales. La primera consistió en una detección automática de la región de interés (ROI) que abarca ambos pies, a partir de la información térmica para localizar las zonas con mayor actividad calórica. Esta etapa permitió obtener un recuadro centrado y de tamaño fijo, común tanto para la imagen infrarroja como para su par en el mundo visible. En la segunda fase, se empleó un enfoque en cascada para refinar esta región y aislar específicamente la zona plantar de los pies, utilizando una red YOLO entrenada mediante aprendizaje transferido y técnicas de eliminación de fondo basadas en distintos espacios de color HSV, YIQ y RGB. Este procedimiento permitió estandarizar el tamaño de las imágenes y aislar las plantas de los pies, eliminando el fondo.

Una vez finalizada la segmentación, se realizó una fusión multiespectral de las regiones infrarrojas y visibles, evaluando diferentes métodos tanto mediante inspección visual como con métricas objetivas ampliamente utilizadas en el ámbito médico, como NMSE, PSNR, CC y entropía. Esta

fusión permitió integrar la información de textura y temperatura en una única representación enriquecida. A partir de esta imagen fusionada, así como de los espectros individuales (infrarrojo y visible), se extrajeron más de veinte características cuantitativas relevantes, seleccionadas en función de los antecedentes revisados y del valor informativo que pudieran aportar en la etapa de clasificación.

La etapa final del sistema correspondió a la clasificación, donde se evaluaron diversos modelos con el objetivo de identificar la presencia de pie diabético. Para ello, se entrenaron clasificadores de diferentes enfoques, incluyendo Máquinas de Vectores de Soporte (SVM), modelos estadísticos y algoritmos de aprendizaje superficial. El rendimiento de cada modelo fue analizado mediante métricas de validación como precisión, sensibilidad, exactitud, F1 score y especificidad, permitiendo seleccionar el clasificador con mejor desempeño para el sistema propuesto.

A continuación, se describen en detalle cada una de las etapas que conforman el sistema.

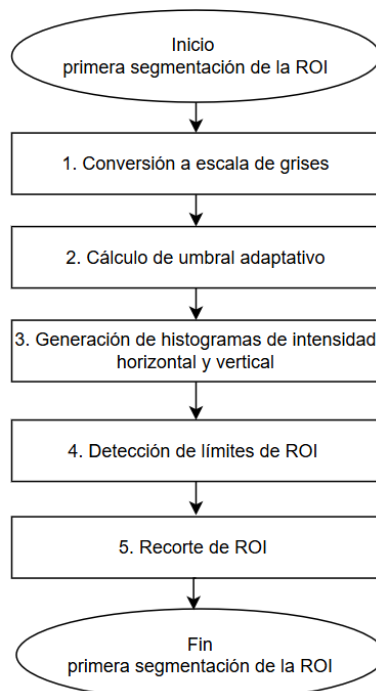
3.1 SEGMENTACIÓN

La segmentación de los pies constituyó una de las fases más críticas del proyecto, debido a la complejidad impuesta por diversos factores: fondos no controlados, variaciones en la resolución y tamaño de las imágenes, así como diferencias en la posición y morfología de los pies en las capturas térmicas. Para abordar estos desafíos, el proceso de segmentación se estructuró en dos fases consecutivas, diseñadas para localizar con precisión la región de interés y aislar específicamente la zona plantar de los pies.

3.1.1. Detección inicial de la región de interés (ROI)

La primera segmentación de las imágenes se lleva a cabo en varias etapas, como se ilustra en la Figura 18:

Figura 18. Diagrama de flujo primera segmentación



Fuente: Elaboración propia.

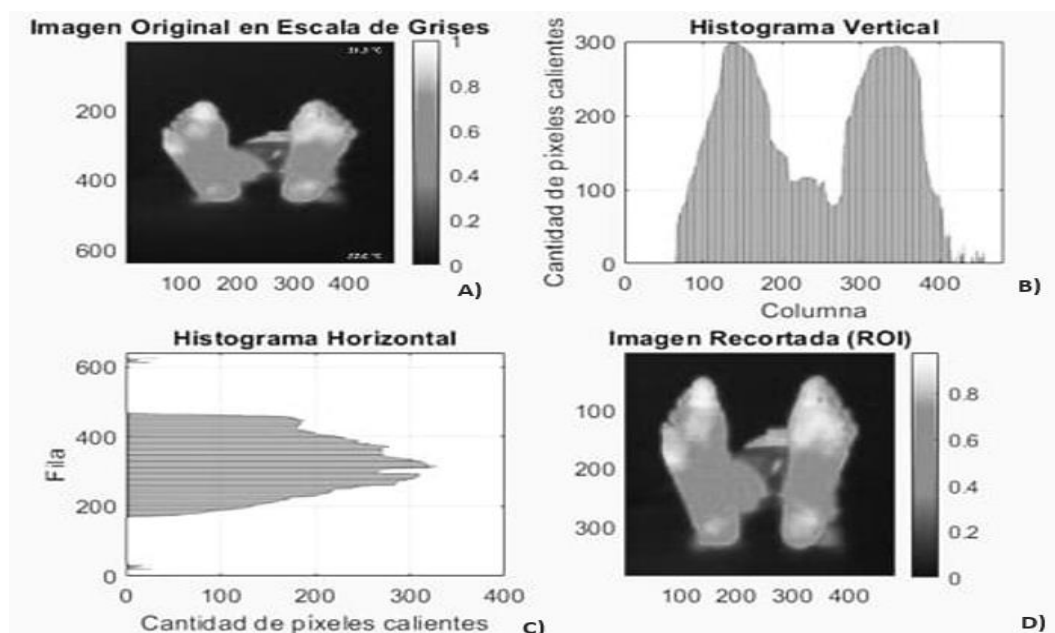
Primero, se realizó la conversión a escala de grises, dado que las imágenes térmicas del conjunto de datos están en formato RGB y no en formato radiométrico. En este tipo de imágenes, los valores de color corresponden a intensidades relativas representadas en una escala pseudocolor incorporada en la captura, más no a temperaturas absolutas en grados Celsius. Al convertirlas a escala de grises, se facilita el procesamiento en el software utilizado, ya que muchas funciones de análisis de intensidad, segmentación y filtrado están diseñadas para operar sobre imágenes monocanal.

Dado que las imágenes en la base de datos presentan diferentes dimensiones, fue necesario establecer un tamaño estándar para todas con el objetivo de descartar información innecesaria y facilitar futuros procesos de entrenamiento a redes neuronales. Se eligió un tamaño de 480x384 píxeles como el área común entre las resoluciones encontradas (como 480x640 y 514x385), área que aún no descarta información de los pies gracias a los protocolos de toma fotográfica discutidos en la sección 2.2.1. Todo esto con el fin de evitar el uso de interpolaciones o extrapolaciones que puedan alterar la información relativa de la imagen térmica.

Para segmentar la región de interés según el tamaño elegido (480x384), el análisis se centró exclusivamente en las imágenes térmicas, dado que estas proporcionan un contraste más marcado entre las zonas activas (pies) y el fondo. Como paso inicial, se aplicó un umbral adaptativo para generar una imagen binaria. Este umbral se calculó a partir de la media y desviación estándar de la intensidad térmica (Ecuación 1), permitiendo destacar las zonas con mayor actividad calórica.

Una vez obtenida la imagen binaria, se construyeron histogramas verticales y horizontales mediante la suma de los valores de píxeles por columnas y filas, respectivamente (Ecuaciones 2 y 3). Estos histogramas permitieron identificar las zonas de mayor concentración térmica dentro de la imagen, facilitando la detección automática de la región de interés. Con ello, se determinaron los primeros y últimos puntos donde la intensidad supera un umbral del 5 % del valor máximo en cada dirección, lo que permitió enmarcar ambos pies en un primer rectángulo. Este valor fue definido heurísticamente, tras realizar pruebas sistemáticas en incrementos del 1 % dentro del rango de 1 % a 10 % sobre el conjunto de imágenes. El umbral del 5 % demostró ser eficaz para descartar el fondo sin comprometer la información relevante de la zona plantar. Finalmente, esta región fue recortada y centrada para ajustarse al tamaño estándar previamente definido.

Figura 19. Resultados delimitación y corte de ROI mediante histogramas.



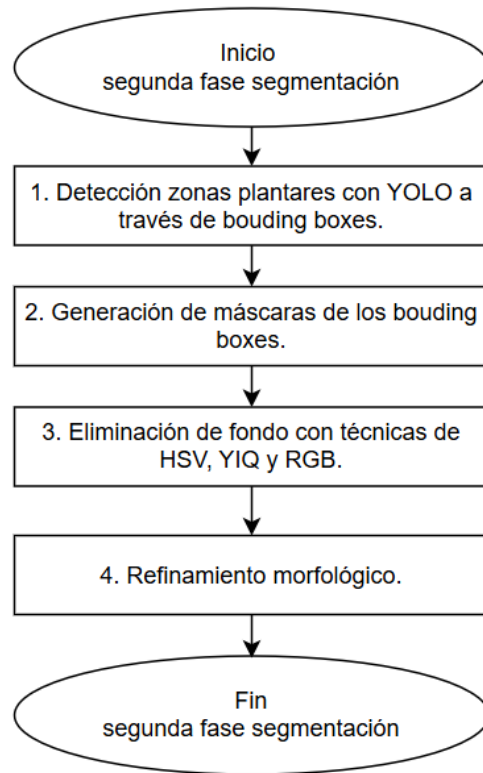
Fuente: Elaboración propia.

La Figura 19 muestra un ejemplo representativo correspondiente a la imagen térmica de un paciente con diabetes, con una resolución de 480×640 píxeles. En este caso, se observa en el Histograma Vertical (Cuadrante B) el conteo de píxeles que superan el umbral térmico establecido, agrupados por columnas, donde la mayor concentración se ubica entre las columnas 75 y 410. Por otro lado, el Histograma Horizontal (Cuadrante C) presenta la distribución equivalente por filas, evidenciando una mayor acumulación de píxeles sobre el umbral entre las filas 182 y 450. De esta forma, una vez definidos los límites de la ROI, la región recortada se centra y ajusta al tamaño estandarizado, tal y como se ilustra en el cuadrante D, donde la reducción es apreciable en mayor medida en el eje Y en comparación al cuadrante A.

3.1.2. Aislamiento de la zona plantar mediante detección en cascada

Una vez establecida la región de interés (ROI) en la imagen térmica mediante el análisis de histogramas, y tras almacenar sus medidas y coordenadas, se procede a aplicar el mismo recorte sobre las imágenes en el espectro visible correspondientes a cada caso. Este paso garantiza la sincronización espacial entre las distintas modalidades de imagen y permite concentrar el análisis únicamente en las áreas relevantes. A partir de esta etapa, se da inicio a la segunda fase del proceso de segmentación, la cual se encuentra representada en la Figura 20.

Figura 20. Algoritmo implementado para segunda segmentación de los pies.



Fuente: Elaboración propia.

La Figura 20 muestra el algoritmo desarrollado para la segunda fase de segmentación de los pies, el cual consta de cuatro etapas secuenciales. Como primer paso, se realiza la detección de las zonas plantares utilizando el modelo YOLO mediante cajas delimitadoras (*bounding boxes*).

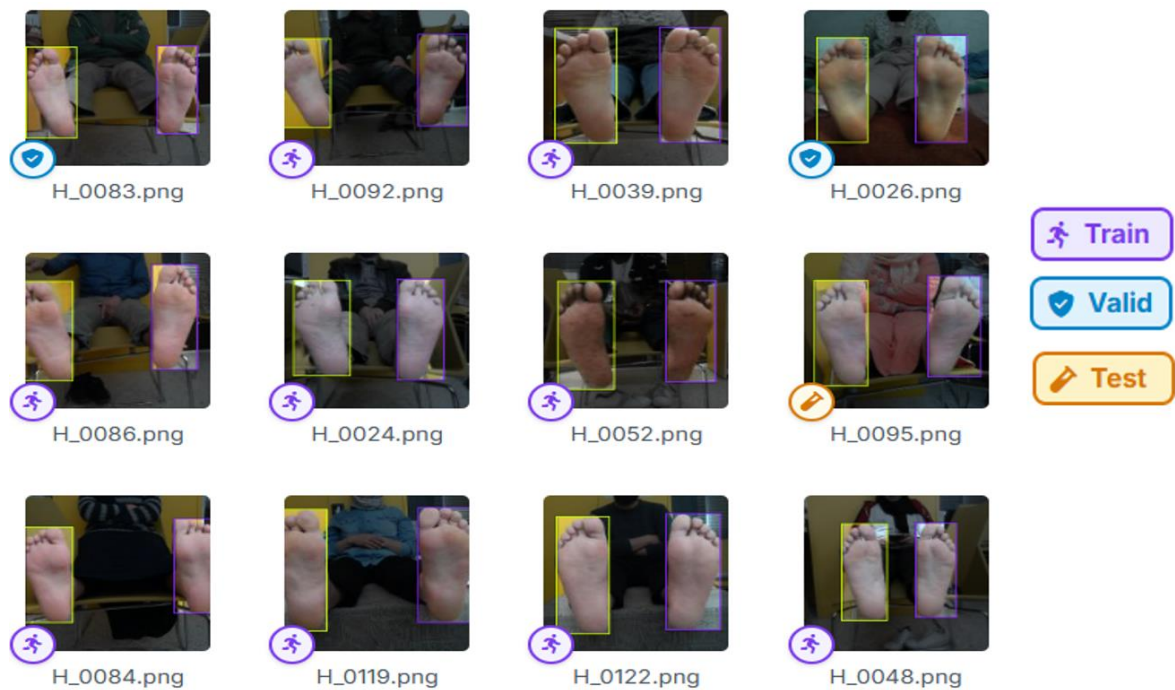
Para llevar a cabo esta detección, de las 415 imágenes en color, se seleccionaron 247 imágenes para el entrenamiento del modelo de detección. Estas fueron divididas en tres subconjuntos: 70% imágenes para entrenamiento (174 imágenes), 20% para validación (49 imágenes) y 10% para prueba (24 imágenes).

Para cada una de estas imágenes, se generaron etiquetas que indican la ubicación de los pies. Este proceso de etiquetado se llevó a cabo utilizando la plataforma Roboflow [75], la cual no solo facilita la tarea de etiquetado manual, sino que además permite exportar los conjuntos de datos en formatos compatibles con distintos modelos de detección, entre ellos YOLO. Esta integración directa con Roboflow agiliza la preparación del dataset y garantiza la compatibilidad estructural con el modelo seleccionado.

Dado que el objetivo de esta etapa era identificar de manera eficiente las regiones correspondientes a los pies dentro de cada imagen, se optó por utilizar el modelo YOLO (You Only Look Once), el cual basa su funcionamiento precisamente en la detección mediante cajas delimitadoras. Esta elección se justifica principalmente por su velocidad y simplicidad, ya que YOLO ha demostrado ser considerablemente más rápido que otros enfoques clásicos como R-CNN y Fast R-CNN, logrando detección en tiempo real sin requerir una arquitectura computacional compleja [76].

En el caso de este trabajo, se señalaron los pies bajo dos distintivos: pie derecho y pie izquierdo, con color amarillo y morado respectivamente como se ilustra en la figura 21.

Figura 21. Etiquetado pie izquierdo y derecho en Roboflow.



Fuente: Elaboración propia.

Otra de las razones por las que YOLO resulta tan eficiente es porque realiza la predicción de clases y la localización de objetos en una sola pasada sobre la imagen, lo que reduce significativamente el tiempo de procesamiento y es gracias al haber sido previamente entrenado con grandes bases de datos como COCO [77], lo cual le proporciona una comprensión generalizada de formas, contornos y estructuras, facilitando su adaptación a nuevas tareas, como la de las plantas de los pies.

El entrenamiento del modelo se llevó a cabo configurando un total de 50 épocas. Este valor se determinó tras realizar una serie de pruebas exploratorias, en las cuales se entrenó el modelo con diferentes cantidades de épocas en incrementos de 10. Durante este proceso, se observó que el rendimiento del modelo mejoraba progresivamente hasta alcanzar un punto de estabilidad en torno a las 50 épocas, sin indicios de sobreajuste. Esta decisión se valida con los resultados obtenidos durante el entrenamiento, los cuales se presentan más adelante.

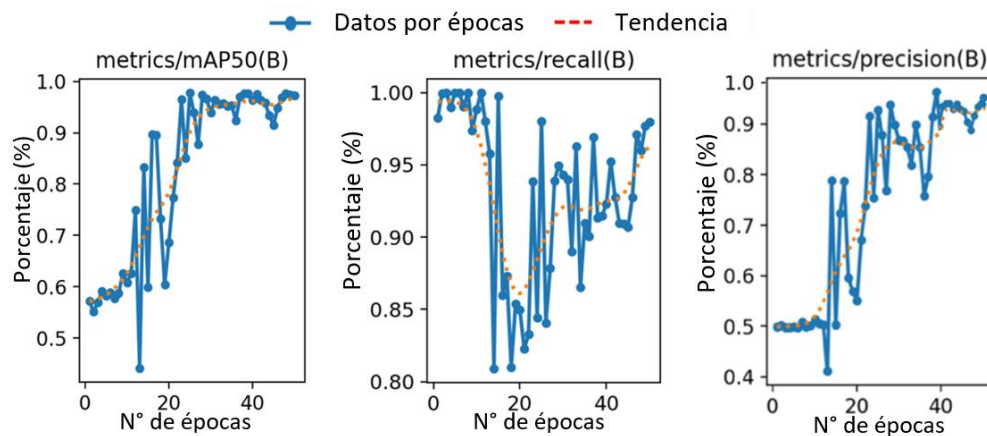
Se configuró un tamaño de lote en 8 muestras para no comprometer la capacidad de procesamiento del entorno utilizado, y el tamaño de imagen fue el establecido en la etapa de estandarización de la imagen, con 480x384 píxeles, siendo un valor que no presentó gran costo computacional a la hora de entrenar y que además es una medida inferior al antecedente mencionado en la sección 1.2.3.

El entrenamiento del modelo YOLO se completó en un tiempo aproximado de 30 a 50 minutos. Al finalizar, el desempeño del modelo alcanzó valores satisfactorios para detectar correctamente la región plantar del pie, con indicadores presentados a continuación:

- Precisión (97.83%): Indica que, de todas las detecciones realizadas por el modelo, el 97.83% correspondieron efectivamente a pies reales. Un valor cercano al 100% refleja una baja tasa de falsos positivos.
- Sensibilidad (97.13%): Mide la proporción de pies correctamente detectados respecto al total real. Este valor sugiere que el modelo rara vez deja de detectar un pie cuando está presente (baja tasa de falsos negativos).
- mAP@50 (98.19%): La media de precisión promedio para un umbral de Intersección sobre Unión (IoU) del 50%, lo cual representa qué tan bien coinciden las predicciones con las verdaderas regiones de los pies. Un valor superior al 98% es indicativo de una detección precisa [78].

Estos resultados se respaldan en la Figura 22, donde se muestra la evolución de las métricas de rendimiento durante el entrenamiento. En la gráfica se aprecia cómo, a lo largo de las 50 épocas, las curvas de precisión, sensibilidad y mAP@50 alcanzan valores cercanos a los obtenidos en la evaluación final.

Figura 22. Evolución de las métricas de rendimiento durante el entrenamiento de YOLO.



Fuente: Elaboración propia.

Los valores reportados a continuación se obtuvieron durante la fase de prueba del modelo YOLO entrenado para la detección de plantas de pies. Se evaluó utilizando un conjunto de imágenes con resolución fija y uniforme.

- Preprocesamiento: 2.87 ms.
- Inferencia (detección): 167.68 ms.
- Postprocesamiento: 2.21 ms.

Dando un tiempo total aproximado de 173 ms por imagen.

Estos valores corresponden al procesamiento de imágenes individuales, todas con la misma resolución, y los resultados varían si se modifican las dimensiones de entrada (menor tiempo a resoluciones por debajo de la indicada) o si se procesan múltiples imágenes simultáneamente.

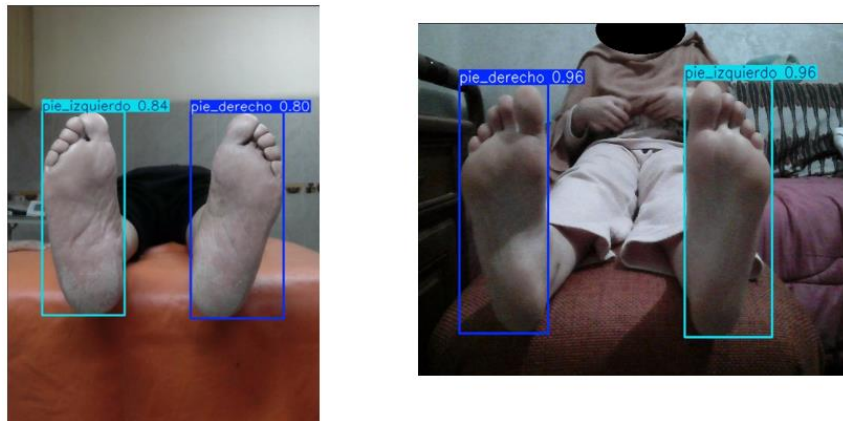
Un aspecto a considerar es que los tiempos de ejecución dependen del hardware utilizado. En este caso, las pruebas se realizaron en un equipo Dell Inspiron 15 3520 con procesador Intel® Core™ i5-1135G7 de 11.^a generación (2.40 GHz, 8 núcleos lógicos), 8 GB de memoria RAM, sistema

operativo Windows 11 Pro de 64 bits y sin el uso de GPU dedicada para la aceleración del entrenamiento o la inferencia.

El rendimiento observado es adecuado para aplicaciones de análisis por lotes con tiempos de respuesta rápidos o escenarios de semi-tiempo real, siempre que se mantenga la misma estandarización del tamaño de la región a través de los histogramas.

Para comprobar el desempeño del modelo en condiciones más generales, se implementó la detección de pies en las 168 imágenes en el dominio visible restantes, que no fueron implementadas en ninguna de las etapas previas. Los resultados confirmaron que el modelo mantuvo un alto rendimiento. No obstante, se identificaron 9 casos en las que, además de detectar los pies, también se identificaron objetos ajenos como rostros o camas, lo que equivale a un 2,17% de error en la detección. Además, se puede observar varios casos en los que el modelo a pesar de haber logrado la identificación de ambos pies no acertó al momento de clasificarlos en pie izquierdo o derecho, pero al no ser interés de este trabajo hacer dicha discriminación sino, la obtención de ubicación de ambos pies indiferente de cuál sea, se continua con la evaluación del modelo ya entrenado.

Figura 23. Resultados ejemplo del sistema de detección de pies mediante YOLO.



Fuente: Elaboración propia.

Con la detección de los pies completada mediante YOLO, el siguiente paso consiste en la segmentación a través de sus respectivas máscaras. El modelo YOLO proporciona etiquetas (labels) con las coordenadas de los rectángulos delimitadores (bounding boxes) de los pies en las imágenes; la Figura 23 ilustra este resultado al mostrar los cuadros de detección sobre las regiones de interés. Como se menciona anteriormente, no es relevante si el modelo identificó correctamente el pie derecho o izquierdo; lo esencial es la ubicación de ambos pies en la imagen.

A partir de las etiquetas proporcionadas por YOLO, se define una máscara binaria en la que:

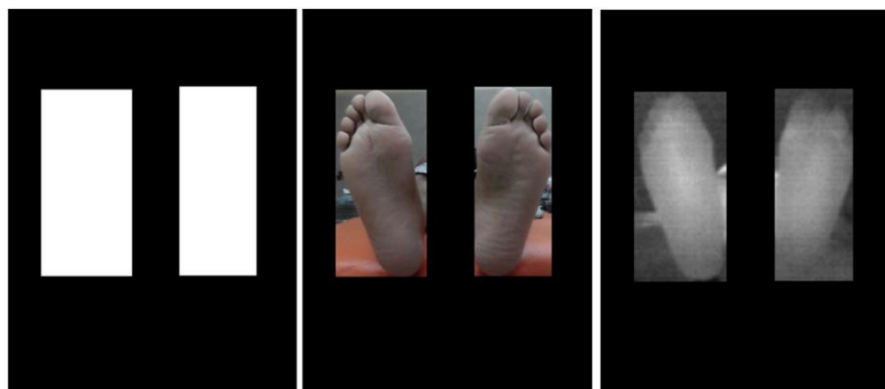
- El área interna del rectángulo delimitado por YOLO se establece con valores de 1, indicando la región correspondiente a los pies.
- El área externa al rectángulo se asigna con valores de 0, representando el fondo y eliminando elementos no deseados.

Una vez generada la máscara, esta se aplica de manera equivalente a las imágenes en modalidad visible e infrarroja. Aunque originalmente las cámaras utilizadas presentan resoluciones distintas, en la etapa de preprocesamiento descrita en la sección anterior todas las imágenes fueron

estandarizadas a una resolución fija de 480×384 píxeles. De esta manera, se asegura la correspondencia espacial entre ambas modalidades: el rectángulo delimitador detectado por YOLO en la imagen visible puede ser proyectado sobre la imagen infrarroja, quedando ambas recortadas en la misma región y con idénticas dimensiones.

Gracias a este procedimiento, cuando se superponen las imágenes recortadas, la región plantar de los pies coincide correctamente en ambas modalidades, lo cual permite garantizar una segmentación coherente y uniforme. La figura 24 muestra un ejemplo de la máscara generada y aplicada en imágenes infrarrojas y visibles, evidenciando esta correspondencia.

Figura 24. Mascara generada y aplicada caso ejemplo.



Fuente: Elaboración propia.

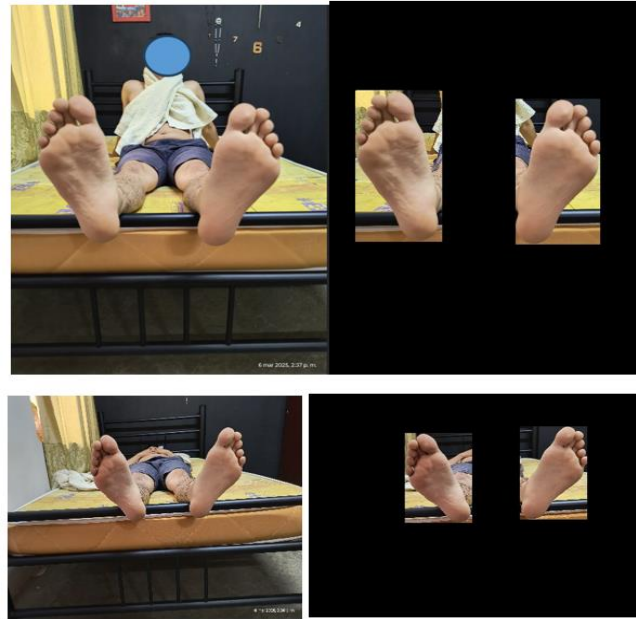
Adicionalmente, para evaluar la capacidad del sistema en escenarios distintos al de la base de datos, se tomaron fotografías de la planta de los pies de cuatro personas de manera independiente. En este conjunto externo, el modelo logró detectar correctamente todos los pies, alcanzando una precisión y sensibilidad del 100%, como se resume en la Tabla 2.

Conjunto de prueba	N° imágenes	Detecciones correctas	Falsos positivos	Falsos negativos	Precisión (%)	Sensibilidad (%)
Subset interno (168 imágenes no vistas)	168	159	9	0	97.83	97.13
Sujetos externos (4 personas)	4	8	0	0	100.0	100.0

Tabla 2. Resultados de la detección de pies en imágenes externas y en conjunto de prueba interno.

La Figura 25 ilustra un ejemplo de este proceso en uno de los sujetos externos, lo cual refuerza la robustez del sistema en la generalización de la tarea de detección.

Figura 25. Sujeto de prueba externo a la base de datos.



Fuente: Elaboración propia.

Una vez obtenidos los segmentos o regiones donde se ubican los pies mediante YOLO, se realizaron esfuerzos adicionales para mejorar la segmentación, ya que los recuadros de detección aún incluían parte del fondo. Esto motivó la necesidad de implementar el paso tres, aplicando técnicas de análisis de color y filtrado para eliminar regiones no deseadas y aislar de manera más precisa la zona correspondiente a los pies.

Inicialmente, en las imágenes térmicas se aplicaron métodos tradicionales como contornos activos y umbralización automática mediante el método de Otsu para la eliminación del fondo restante. Sin embargo, estos enfoques presentaron limitaciones importantes:

- La umbralización tendía a eliminar regiones internas de los pies, que presentaban zonas naturalmente más oscuras.
- En las imágenes capturadas con la cámara FLIR ONE Pro, cuyo sensor posee una resolución nativa de 160×120 píxeles, se evidenció una zona térmicamente activa alrededor de los pies, posiblemente relacionada con el procesamiento de las imágenes en la base de datos. Si bien la cámara original registra a baja resolución, en el conjunto de datos se encontraron imágenes con dimensiones mayores (480×640 y 514×385 píxeles), lo que sugiere que pudieron haber sido interpoladas o reescaladas. Este proceso puede haber generado bordes térmicos difusos que dificultaron una separación clara entre el objeto y el fondo.
- Si bien una segmentación precisa píxel a píxel podría ser más adecuada para ciertos análisis, entrenar modelos con segmentación semántica o instancia requiere mayor capacidad computacional, mayor cantidad de datos anotados a nivel de máscara, y más tiempo de ajuste para estabilizar los hiperparámetros.

Debido a estas dificultades, se optó por complementar la segmentación utilizando también la información de las imágenes en espectro visible explorando distintas técnicas sobre las imágenes RGB.

Se realizó una investigación preliminar enfocada en técnicas de filtrado de color de piel, en la cual se identificaron diversos estudios relacionados con la detección de contenido explícito [43], [44], de donde se seleccionaron los espacios de color HSV y YIQ, que dieron buenos resultados por su capacidad para identificar diferentes tonos de piel.

Como punto de partida, se tomaron los rangos propuestos en el trabajo de Camargo y dos Santos [43], donde se utilizó el modelo HSV para identificar tonos de piel en imágenes RGB. En dicho estudio, se establecieron los siguientes umbrales:

- $0 < H \leq 0.12$
- $0.5 < S < 0.9$
- $0.2 < V < 0.95$

Estos valores sirvieron como referencia inicial para el filtrado de tonos de piel en las imágenes RGB asociadas al mundo del color utilizadas en este trabajo, centradas específicamente en la detección de las plantas de los pies.

Posteriormente, se realizó una exploración sobre los resultados obtenidos, examinando el comportamiento de los píxeles no deseados en este intervalo. A partir de esta revisión, se ajustaron los umbrales a:

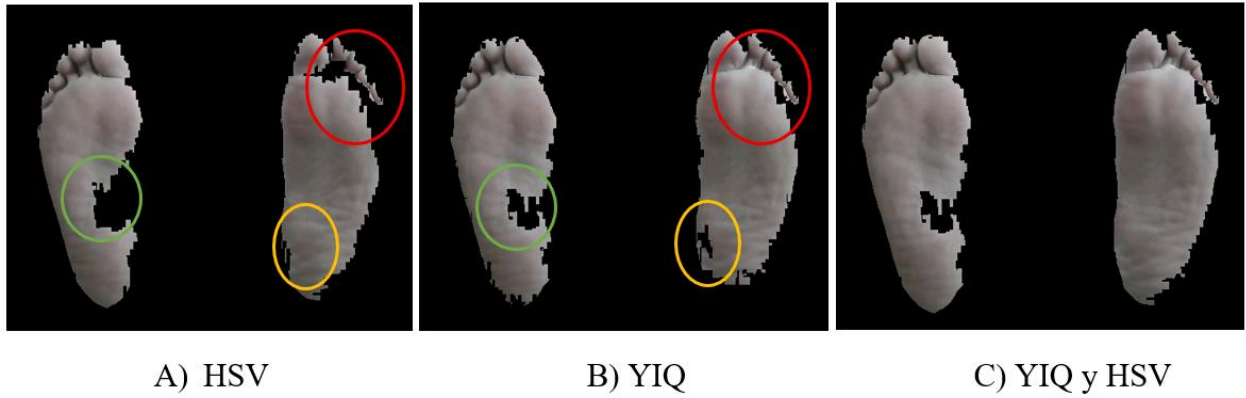
- $H \leq 0.12$ o $H \geq 0.80$, para abarcar tonalidades rojizas y magentas propias de la piel.
- $S \leq 0$ o $S \geq 0.8$, eliminando regiones excesivamente apagadas o saturadas.
- V entre 0.1 y 0.74, excluyendo zonas demasiado oscuras o con reflejos intensos que pudieran deberse a malas condiciones de iluminación.

Paralelamente, la imagen fue convertida al espacio YIQ, el cual es utilizado en procesamiento de imágenes debido a su capacidad para separar la información de luminancia (Y) de la crominancia (I y Q). La transformación se realizó mediante la ecuación 7.

A partir del análisis de la distribución de valores en este espacio, y considerando criterios académicos utilizados para la detección de regiones de piel en el material de clase de visión artificial de la Universidad del Valle [36], como $(I > 25) \& (Q > -10) \& (Q < 15)$, se definieron los siguientes parámetros de segmentación:

- Se seleccionaron las regiones con valores de $I > 1$, identificando áreas con una predominancia de tonos similares a la piel.
- Se filtraron valores de Q entre -10 y 15, delimitando mejor los colores cercanos a la piel y evitando la inclusión de zonas de fondo con tonalidades similares como colores naranja o amarillo.

Figura 26. Resultados HSV, YIQ y combinación.



Fuente: Elaboración propia.

Para mejorar la segmentación de las regiones correspondientes a los pies en imágenes a color, se diseñó un procedimiento que aprovecha las ventajas complementarias de los espacios de color HSV y YIQ. En particular, como se observa en la figura 26, hay ciertos píxeles correctamente identificados en uno de los espacios, pero no en el otro. Como se muestra en A y B, el espacio HSV permite detectar regiones con ciertas tonalidades de piel que el espacio YIQ omite, mientras que YIQ revela detalles que el HSV no logra capturar debido a su sensibilidad a variaciones de iluminación. Por esta razón, se propuso una estrategia que integra la información segmentada en ambos espacios (C), permitiendo así compensar las limitaciones individuales de cada uno. Para lograr dicha integración, se aplicó una operación lógica OR donde para cada píxel (x,y) , se definió la imagen fusionada $F(x,y)$ como:

$$F(x,y) = \begin{cases} I_{YIQ}(x,y) & \text{si } I_{YIQ}(x,y) \neq [0,0,0] \\ I_{HSV}(x,y) & \text{si } I_{HSV}(x,y) \neq [0,0,0] \\ [0,0,0] & \text{en otro caso} \end{cases} \quad (43)$$

La ecuación 43 describe que si un píxel contiene información válida (distinta de cero) en al menos uno de los espacios, ese valor es conservado en la imagen resultante. Aunque esta estrategia logró reducir en gran medida el fondo no deseado, persistían errores debido a objetos o zonas del fondo que presentaban colores similares a los de la piel (tonos de café, naranja o amarillo) como se ilustra en la figura 27.

Figura 27. Resultados del filtrado HSV y YIQ.



Fuente: Elaboración propia.

Para enfrentar este desafío, se realizó un análisis manual y detallado de los colores que generaban falsas detecciones, utilizando la herramienta *impixelinfo* en MATLAB, la cual permite identificar los valores individuales de las componentes R, G y B de cada píxel. Durante este proceso, se examinaron de forma directa los valores de los píxeles que persistían tras la segmentación, y con base en los patrones observados se ajustaron de manera iterativa los parámetros de los filtros. Este procedimiento permitió afinar progresivamente los criterios de eliminación, hasta dejar establecidos los filtros finales empleados en el procesamiento. Algunos valores recopilados durante el análisis se presentan en la tabla 3.

Capa:	R	R-G	G	G-B	B
Cafe_1	85	15	70	12	58
	89	25	64	21	43
	98	20	78	20	58
	98	18	80	19	61
Cafe_2	23	8	15	6	9
	33	12	21	5	16
	21	2	19	6	13
	23	6	17	5	12
Naranja	159	78	81	23	58
	145	65	80	11	69
	153	75	78	20	58
	160	71	89	25	64

Tabla 3. Valores de píxeles para algunos casos de color.

Con base en este análisis, se diseñaron una serie de filtros específicos para detectar y eliminar colores particulares del fondo, tales como:

- Color naranja.
- Diferentes tonos de amarillo.
- Tonos de café.

Cada uno de estos filtros fue expresado mediante condiciones lógicas sobre los valores de las componentes RGB, generando máscaras binarias específicas que permitieron eliminar progresivamente los elementos no deseados. Aunque este enfoque es manual e intensivo, permitió mejorar significativamente la calidad de la segmentación final, tal y como se muestra en el ejemplo de la figura 28, adaptándose a las particularidades del conjunto de imágenes utilizadas en el estudio.

Figura 28. Resultado filtrado de colores.

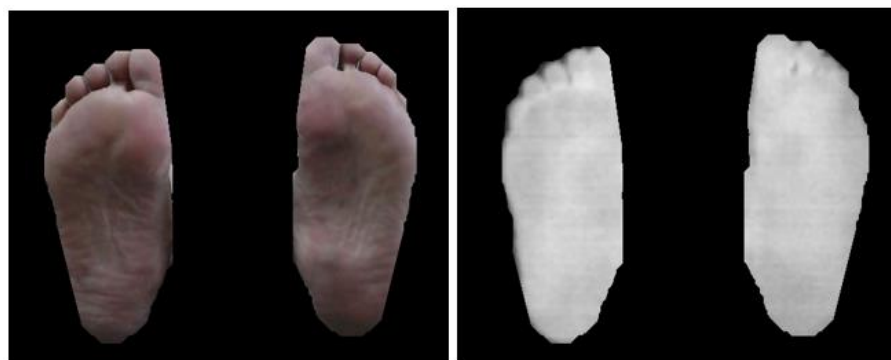


Fuente: Elaboración propia.

Por último, para la limpieza final de residuos no pertenecientes a los pies, se aplicó una apertura morfológica utilizando un elemento estructurante circular de 20 píxeles de radio tras evaluar el efecto de distintos tamaños sobre la morfología de las regiones segmentadas. Valores menores resultaban insuficientes para eliminar completamente los residuos del fondo, mientras que valores mayores comprometían la integridad de las regiones plantares, eliminando zonas relevantes como los dedos o los bordes del talón. El valor de 20 píxeles ofreció un balance adecuado entre limpieza del ruido y conservación de la estructura del pie.

Esta operación consiste en una erosión seguida de una dilatación, permitiendo eliminar pequeñas regiones de ruido mientras se preserva la estructura de los pies. Posteriormente, esta máscara se utilizó para depurar tanto las imágenes en el espectro visible como las térmicas, suprimiendo completamente las áreas de fondo, como se observa en la figura 29.

Figura 29. Resultado final segmentación planta de los pies.



Fuente: Elaboración propia

3.1.3. Segmentación manual y comparación

Con el objetivo de establecer una base de comparación objetiva y evaluar el desempeño de los métodos automáticos implementados, se realizó una segmentación manual de las imágenes mediante la herramienta Image Segmenter de MATLAB. En este proceso, se delimitaron cuidadosamente, una a una, las regiones correspondientes a las plantas de los pies.

Como se observa en la figura 30, cada recorte abarcó la totalidad de la base plantar, incluyendo los dedos, ya que el enfoque de este estudio está orientado al análisis temprano de condiciones asociadas al pie diabético. Enfermedades como la neuropatía periférica y la isquemia no afectan de manera uniforme toda la superficie del pie, por lo que era fundamental conservar la mayor cantidad de información térmica y morfológica posible en las regiones críticas. La base de datos utilizada incluye únicamente pacientes sin úlceras ni amputaciones visibles, permitiendo centrar el análisis en las posibles alteraciones previas al desarrollo de lesiones avanzadas.

Estas segmentaciones manuales sirvieron como referencia para evaluar la calidad de la segmentación automática y también como fuente de comparación para los resultados obtenidos en la etapa de clasificación, aportando una perspectiva adicional sobre el impacto que tiene la precisión de las máscaras en el rendimiento general del sistema.

Figura 30. Pies segmentados a través de la herramienta Image Segmenter de MATLAB.



Fuente: Elaboración propia.

Para la comparación cuantitativa entre las segmentaciones manuales y automáticas, se implementó el coeficiente de Dice como métrica de evaluación de acuerdo con la ecuación 8. El análisis se realizó sobre un conjunto de 200 imágenes para cada tipo de segmentación, donde cada par de segmentaciones fue comparado individualmente. Posteriormente, se calculó el porcentaje de similitud correspondiente a cada imagen y se obtuvo un promedio global para evaluar el desempeño general de los métodos automáticos.

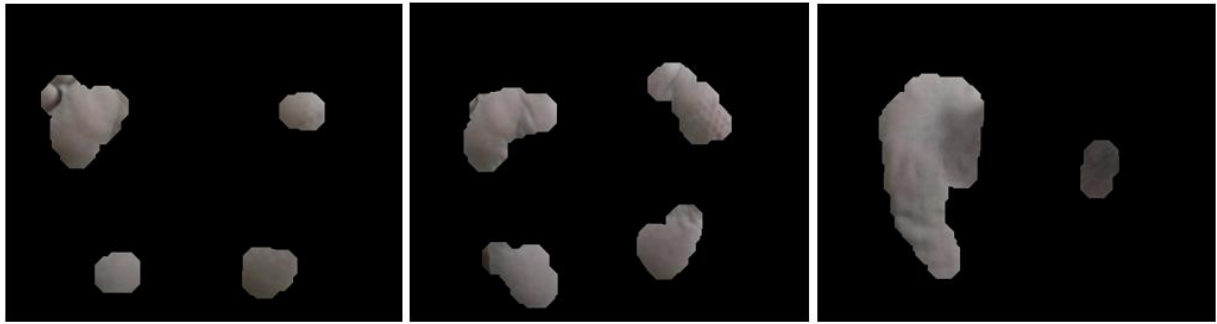
Como resultado de esta evaluación, se obtuvo un Promedio global de similitud de 90.01%, indicando un alto grado de concordancia entre las segmentaciones automáticas y las referencias manuales. Algunos de los resultados se ilustran en la tabla 4.

Imagen	Similitud Porcentaje
segmentado_recortada_R0_0001.png	93.3630
segmentado_recortada_R0_0003.png	90.9849
segmentado_recortada_R0_0005.png	93.5030
segmentado_recortada_R0_0007.png	73.0276
...	...
segmentado_recortada_R0_0013.png	91.6942
segmentado_recortada_R0_0015.png	81.5567
segmentado_recortada_R0_0017.png	91.6810
segmentado_recortada_R0_0019.png	93.1155

Tabla 4. Fragmento de resultados DSC para segmentación de imágenes.

Para entender la causa de los porcentajes por debajo del 80%, se realizó un análisis visual de las imágenes mal clasificadas. Se observó que varias de ellas presentaban una tonalidad pálida en los pies, lo que ocasionó una segmentación deficiente tal y como se observa en la figura 29. Este efecto no estuvo necesariamente relacionado con el fondo, sino con la falta de coincidencia entre los tonos presentes en los pies y los umbrales definidos para los filtros en el espacio de color, los cuales se diseñaron para resaltar colores típicos de piel y estos, contaban con valores más altos en las componentes H, V e I, pero incluir estos valores dentro de los rangos utilizados, significaba la aparición de píxeles cercanos al color blanco, aumentando significativamente valores erróneos. Esto resultó en una segmentación incompleta o inexacta como se ilustra en la figura 31, afectando negativamente la extracción de características y, por consiguiente, el desempeño del clasificador.

Figura 31. Imágenes con problemas en el proceso de entrenamiento y clasificación.



Fuente: Elaboración propia.

3.2. EXTRACCIÓN DE CARACTERÍSTICAS

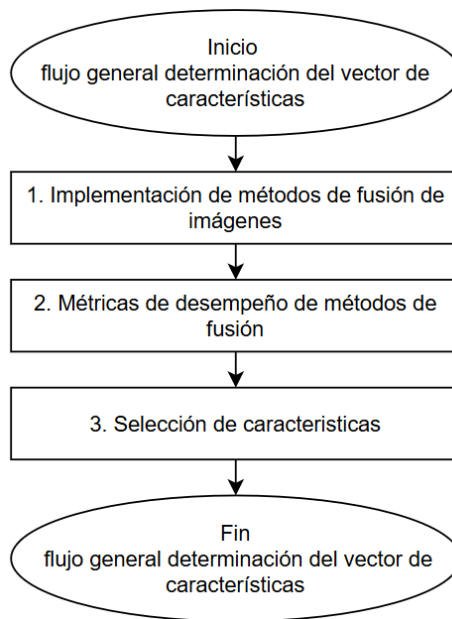
Finalizada la segmentación, se dio inicio al proceso de extracción de características, tomando como referencia aquellas descritas en trabajos previos. Para cada región segmentada del pie, las métricas estadísticas (máximo, mínimo, mediana, promedio, desviación estándar, curtosis y asimetría) se calcularon sobre el conjunto de píxeles correspondientes a ambos pies.

Adicionalmente, se calcularon características de textura a partir de la matriz de coocurrencia de niveles de gris (GLCM). A partir de esta matriz se obtuvieron los descriptores de contraste, energía, homogeneidad y entropía.

Finalmente, la diferencia promedio de temperatura se obtuvo comparando el valor medio de todos los píxeles segmentados del pie izquierdo frente al derecho.

No obstante, para llegar a esta configuración se realizó un proceso metodológico que involucró la evaluación de distintos métodos de fusión, seguido de la extracción y análisis de características sobre las imágenes resultantes. Este procedimiento se resume en la Figura 32, que muestra el flujo de fusión, evaluación y consolidación del vector final de características.

Figura 32. Flujo general para la obtención del vector de características a partir de imágenes segmentadas.



Fuente: Elaboración propia.

Los resultados presentados en este capítulo corresponden a imágenes segmentadas manualmente, utilizadas como referencia para llevar a cabo una evaluación detallada y guiada del método de fusión y selección de características.

3.2.1. Fusión de imágenes

En la literatura se destacan tres enfoques ampliamente utilizados y reconocidos por su efectividad: la Transformada Wavelet Discreta (TWD), la Transformada Coseno Discreta (DCT) y la Pirámide Laplaciana (TPL). La TWD, estudiada por Rodríguez Esparragón et al. [79], ha sido empleada en fusión multiresolución por su capacidad para descomponer señales en diferentes escalas y conservar estructuras espaciales relevantes. La DCT, utilizada por Fidalgo et al. [80] en contextos de clasificación y fusión de imágenes, ha mostrado buenos resultados en la compresión y en la integración de información por su eficiencia energética y bajo costo computacional. Por su parte, la Pirámide Laplaciana, aplicada por Rojas Carrasco [81] en sistemas de reconocimiento facial basados en la fusión de imágenes visibles y térmicas, ha demostrado ser eficaz en la preservación de bordes y detalles finos en tareas de análisis clínico y biométrico.

En el caso de la DWT, es un método que permite la extracción eficaz de detalles relevantes mediante la obtención de coeficientes de aproximación y detalle en diferentes niveles jerárquicos (ecuación 9), tal como se explicó en la sección 2.3.3.1. Basta con seleccionar el tipo de ondícula y su nivel de descomposición a trabajar. En varios trabajos, como el de Bin, Shutao y Fengmei [82] o Mondragón y Solen [83], se utilizan ondículas Daubechies a diferentes niveles comprendidos entre 4 y 8, destacando mejores resultados para el nivel 4, el cual también es seleccionado en este trabajo.

La transformada de coseno discreta es, entre las mencionadas, la más sencilla de implementar, se toman los coeficientes de cada imagen a través de la ecuación 10 y son fusionados a través de cualquier operación matemática o estadística deseada (Multiplicación, suma, resta, promedio), usualmente y en el caso de este estudio, se escoge un valor promedio.

En un caso más complejo se encuentra la transformada de la pirámide laplaciana, la cual requiere, además de la etapa de preprocesamiento, construir la pirámide gaussiana a partir de la implementación de un filtro gaussiano, con el objetivo de conseguir suavidad. Posteriormente, se submuestra la imagen suavizada, como se describe en la ecuación 12. Luego, se suma otro proceso de iteración para la construcción de la pirámide laplaciana, donde se expande la imagen suavizada del siguiente nivel y se resta del nivel actual, según lo definido en la ecuación 13. Para finalizar, se fusionan los coeficientes de las dos imágenes y se reconstruye la imagen a través de un proceso inverso iterativo, que implica la expansión de la imagen fusionada desde el nivel más bajo de la pirámide y su suma con la imagen del nivel superior, hasta alcanzar el nivel original, conforme a la ecuación 14. Los resultados de la fusión haciendo uso de las transformadas se presentan a continuación en la figura 33.

Figura 33. Resultado ejemplo fusión transformadas para un paciente diabético.

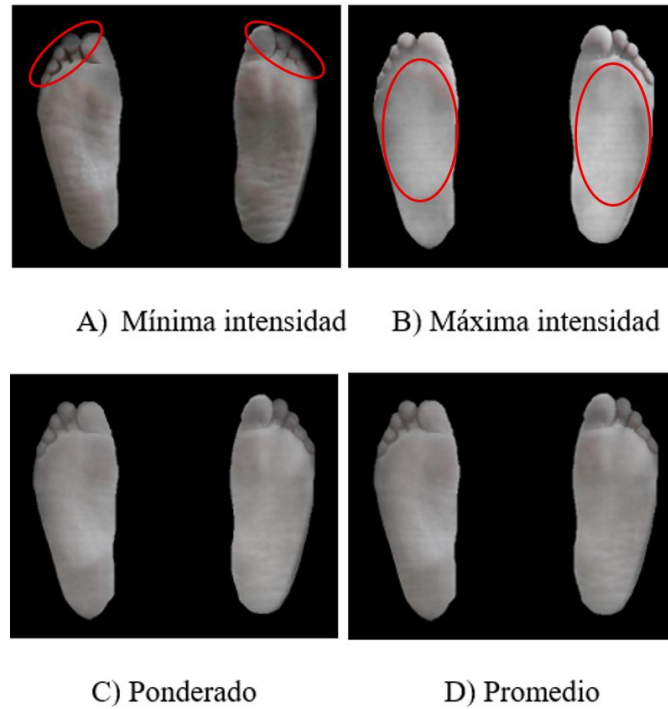


Fuente: Elaboración propia.

A simple vista, no se evidencian diferencias visuales significativas entre los resultados obtenidos con cada una de las transformadas aplicadas. No obstante, estas aparentes similitudes serán contrastadas y evaluadas cuantitativamente a través de métricas objetivas, las cuales se presentan y analizan en la sección posterior 3.2.2.

Además de las técnicas de fusión basadas en transformadas, se decidió implementar métodos de fusión espacial, caracterizados por su simplicidad y facilidad de implementación, al estar compuestos principalmente por operaciones elementales como el promedio, la ponderación lineal, y la selección de valores mínimos o máximos entre los píxeles correspondientes de las imágenes fuente. El objetivo de incluir estos enfoques fue evaluar si alguno de ellos, a pesar de su simplicidad, podría ofrecer resultados comparables en calidad a los obtenidos mediante las transformadas, y así considerar su viabilidad como alternativa eficiente en contextos donde los recursos computacionales sean limitados. Los resultados se presentan a continuación en la figura 34:

Figura 34. Resultados métodos de fusión espacial.



Fuente: Elaboración propia.

La opción A de la figura 32 permite observar que las zonas de menor valor (temperatura) se localizan principalmente en los bordes del pie. En contraste, la opción B evidencia una mayor acumulación de temperatura en la región interna del pie (zona clara). Por su parte, las imágenes C y D presentan resultados visualmente más balanceados, y muestran una similitud notable con los obtenidos mediante técnicas basadas en transformadas.

3.2.2. Métricas de desempeño de los métodos de fusión de imágenes

Las técnicas de fusión fueron evaluadas cuantitativamente a través de un conjunto de métricas objetivas ampliamente utilizadas en la literatura, entre las que se encuentran: el error cuadrático medio normalizado (NMSE, ecuación 19), la relación señal/ruido pico (PSNR, ecuación 20), el coeficiente de correlación (CC, ecuación 21) y la entropía (ecuación 22).

Con el fin de facilitar la comparación global entre los distintos métodos aplicados y considerando la multiplicidad de métricas involucradas, se propuso un puntaje compuesto que las integra en una sola expresión, como se muestra en la ecuación 44.

Dado que cada métrica posee una escala y naturaleza distintas, todos los valores fueron previamente normalizados en el rango de 0 a 1, garantizando así una contribución equitativa a la evaluación global. En este contexto, se consideró que un valor bajo de NMSE indica una mejor similitud entre la imagen fusionada y las originales, por lo que se transformó esta métrica en $1 - \text{NMSE}$, permitiendo así alinearla con el criterio de que un valor más alto indica un mejor desempeño, tal como ocurre con PSNR, CC y la entropía. Finalmente, el puntaje compuesto se definió como:

$$\text{Puntaje} = \text{PSNR} + \text{CC} + \text{Entropía} + (1 - \text{NMSE}) \quad (44)$$

Este puntaje tiene un valor máximo teórico de 4 por imagen, correspondiente al caso ideal en el que cada una de las métricas alcanza su valor óptimo (es decir, 1 tras la normalización). Dado que la evaluación se realizó sobre un conjunto de 100 imágenes, el puntaje total máximo esperado para cada método de fusión es de 400. Por tanto, valores cercanos a 400 indican un desempeño global elevado.

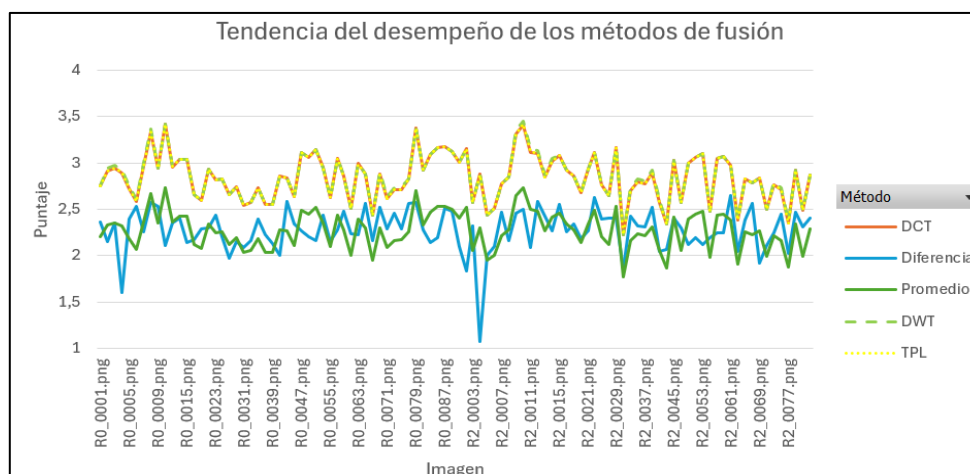
Tras la implementación de las métricas a cada una de las imágenes fusionadas con sus correspondientes RGB e infrarrojas, se evidenció que el método de fusión basado en DWT obtuvo el puntaje total más alto, posicionándose como la mejor técnica dentro del conjunto evaluado y siendo este, el método a implementar durante las siguientes secciones. Un fragmento de los resultados y el puntaje total por método de fusión se encuentran en la tabla 5 y 6.

Imagen	NMSE	PSNR	CC	Entropía	Método	Puntaje
R0_0001.png	0,043495	0,513486	0,955061	0,3309413	DCT	2,755993
R0_0001.png	0,043541	0,512863	0,955263	0,3310455	DWT	2,755631
R0_0001.png	0,043674	0,512103	0,955054	0,3310872	TPL	2,75457
R0_0001.png	0,119053	0,331771	0,760056	0,389627	Promedio	2,362401
...	...	...	...	...	...	...
R2_0083.png	0,053824	0,611173	0,922659	0,3886844	DCT	2,868693
R2_0083.png	0,053874	0,610412	0,922906	0,3889957	DWT	2,86844
R2_0083.png	0,054006	0,609665	0,922715	0,388295	TPL	2,866669
R2_0083.png	0,165574	0,310294	0,656881	0,6088772	promedio	2,410479

Tabla 5. Implementación de métricas para evaluación método de fusión de imágenes.

Para complementar este análisis, en la Figura 35 se ilustra la variación del puntaje en función de las imágenes evaluadas. En ella se aprecia que los métodos DCT, DWT y TPL presentan un comportamiento muy similar y consistentemente superior frente a los demás, mientras que Promedio y Diferencia se mantienen en niveles inferiores a lo largo de todo el conjunto de pruebas.

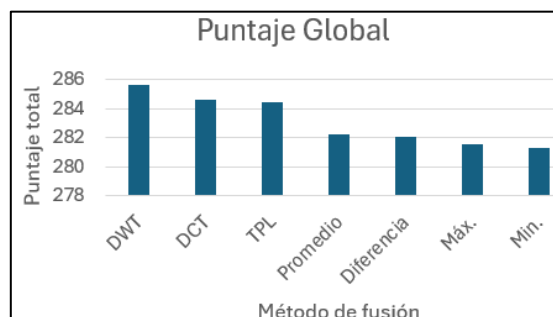
Figura 35. Variación del puntaje de fusión de imágenes según el método aplicado.



Fuente: Elaboración propia.

La Figura 36 resume el desempeño global de cada método de fusión. Allí se observa que el método DWT alcanzó el puntaje total más alto.

Figura 36. Puntaje global para cada método de fusión implementado.



Fuente: Elaboración propia.

Método	Total
DWT	285,597
DCT	284,574
TPL	284,465

Tabla 6. Puntaje global para los tres mejores métodos de fusión.

3.2.3. Extracción y selección de características

Ya contando con los espectros RGB e infrarrojo así como la imagen resultante del proceso de fusión, se procede con la etapa de extracción y selección de características. Esta etapa se fundamenta en atributos previamente reportados en los antecedentes, considerados de mayor relevancia diagnóstica y con capacidad de diferenciación entre clases. Se calcularon métricas estadísticas: máximo, mínimo, mediana, promedio, desviación estándar, curtosis y asimetría a partir del conjunto de píxeles correspondientes a ambos pies. De manera complementaria, se obtuvieron características de textura empleando la matriz de coocurrencia de niveles de gris (GLCM), de la cual se derivaron los descriptores de contraste, energía y homogeneidad. Una característica particularmente relevante fue la diferencia en la intensidad promedio entre el pie derecho y pie izquierdo, la cual guarda relación con indicadores clínicos ampliamente documentados en la literatura. Diversos estudios, entre ellos Rutkove [84], han mostrado que una diferencia térmica superior a 2 °C entre zonas homólogas puede constituir un signo temprano de complicaciones asociadas al pie diabético, como úlceras o laceraciones. Dichos hallazgos evidencian que variaciones relativamente pequeñas en la temperatura son suficientes para alterar la fisiología neuromuscular, lo que resalta la sensibilidad de este parámetro como marcador de procesos patológicos. En este trabajo, aunque las imágenes térmicas utilizadas no fueron calibradas en valores absolutos de grados Celsius, el análisis se enfocó en la detección de diferencias relativas de intensidad entre ambos pies, considerando que estas diferencias poseen un sustento clínico reconocido y resultan informativas para la identificación temprana de alteraciones asociadas al pie diabético.

Las demás características fueron seleccionadas por su capacidad para capturar distintos aspectos de la distribución estadística e información de las regiones segmentadas de acuerdo con los trabajos previos. Por ejemplo:

- El promedio y la mediana de las intensidades permiten capturar el comportamiento central de la imagen, y son útiles para identificar elevaciones o reducciones generales de temperatura que podrían estar asociadas a procesos inflamatorios o isquémicos.
- La desviación estándar, curtosis y asimetría brindan información sobre la forma y dispersión de la distribución térmica. Estas métricas ayudan a identificar si la imagen presenta zonas localizadas de mayor o menor temperatura, lo cual es típico en pies diabéticos, donde hay regiones con alteraciones circulatorias o daño nervioso que modifican el patrón térmico.
- En cuanto al análisis de textura, características como el contraste, energía y homogeneidad, extraídas a partir de matrices de co-ocurrencia, permiten evaluar la organización espacial de los valores de intensidad [16]. Este tipo de análisis no solo es útil en imágenes térmicas, sino que gana valor cuando se aplica sobre imágenes fusionadas entre espectro infrarrojo y visible ya que la fusión aporta detalles estructurales del espectro a color que refuerzan la localización de patrones térmicos irregulares, haciendo más evidente cualquier alteración en la vascularización o en la integridad de los tejidos.
- Por último, la entropía cuantifica el nivel de desorden en la imagen, siendo un buen indicador de complejidad térmica. Un pie sano tiende a presentar una distribución térmica ordenada, mientras que un pie con indicios de neuropatía o problemas circulatorios muestra mayor entropía debido a la presencia de zonas anómalas dispersas. Además, se tomó como referencia el estudio “State-of-the-Art Features for Early-Stage Detection of Diabetic Foot Ulcers Based on Thermograms”, en el que la combinación de la mayoría de estas variables demostró una alta capacidad discriminativa para detectar alteraciones asociadas al pie diabético.

Inicialmente, se planteó la extracción de 12 características por cada espectro (visible, infrarrojo) y fusión, para un total de 36 características por paciente. Sin embargo, tras un análisis más detallado, se descartó el espectro visible, ya que los pacientes no presentaban lesiones visibles como úlceras o amputaciones, y las imágenes en el espectro visible no mostraban diferencias físicas claras entre clases. Además, considerando antecedentes como el trabajo de Arteaga et al. [19], donde únicamente se implementó el espectro térmico de la base de datos STANDUP, se ratifica que este enfoque resulta más adecuado para capturar las alteraciones internas relevantes para la detección del pie diabético.

Adicionalmente, para la característica de la diferencia en la intensidad promedio entre ambos pies, se decidió manejar el mismo esquema que los trabajos previos, donde solo es aplicada al espectro térmico, ya que es allí donde se representan de manera más precisa los cambios de distribución térmica relacionados con el pie diabético. Por este motivo, el conjunto final de características utilizadas para la etapa de clasificación fue de 23 variables.

Para garantizar una correcta comparación entre características de distinta escala y naturaleza, todas las variables fueron normalizadas en un rango de 0 a 1 mediante una transformación lineal mínimo - máximo. Esta normalización se aplicó utilizando la siguiente expresión:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (45)$$

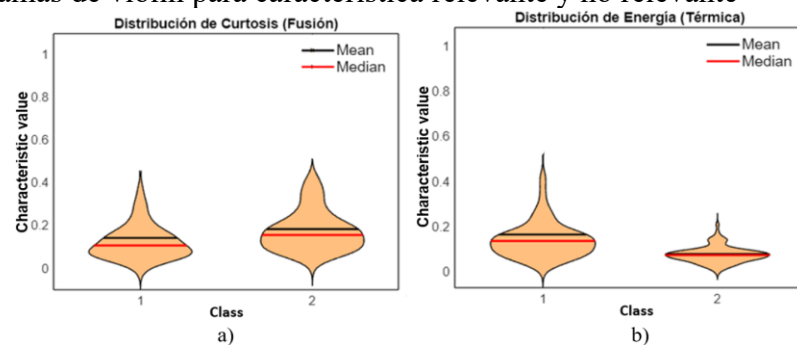
Esta normalización es fundamental para evitar que características con valores numéricos mayores dominen el comportamiento del modelo de clasificación y facilita la convergencia de algoritmos de aprendizaje automático y mejora la estabilidad numérica del procesamiento posterior.

Tras la extracción y normalización de las 23 características por paciente, se procedió a evaluar su relevancia mediante diferentes métodos estadísticos y algoritmos de selección de características.

Inicialmente, se generaron diagramas de violín para visualizar la distribución de cada característica entre pacientes sanos y diabéticos. Posteriormente, se aplicó la prueba estadística ANOVA (Análisis de Varianza) con un nivel de significancia (p) de 0.01. Un valor p menor a 0.05 indica que la característica presenta diferencias relevantes entre los grupos, y, por tanto, se considera significativa. Al usar un nivel de significancia más estricto como 0.01, se reduce la probabilidad de falsos positivos [56].

Se presenta el caso ejemplo en la figura 37, donde la imagen a) corresponde a la característica curtosis de la imagen fusionada y que el ANOVA considera no significativa. Aunque visualmente podría sugerir cierta diferencia en la posición de la media y la mediana entre clases, las distribuciones presentan un alto solapamiento, es decir, ambas clases se concentran en zonas muy parecidas, con un rango de valores que rondan desde 0 hasta 0.3. Caso contrario la imagen b) donde si se visualiza una diferencia significativa en la distribución y ubicación de la media y mediana para la característica energía de la imagen térmica, indicando que los pacientes sanos (Clase 1) cuentan con valores mayores concentrados en el rango de 0 a 0.2 mientras que los pacientes con pie diabético (Clase 2) abarcan valores en mayor profundidad para un rango entre 0.05 y 0.15.

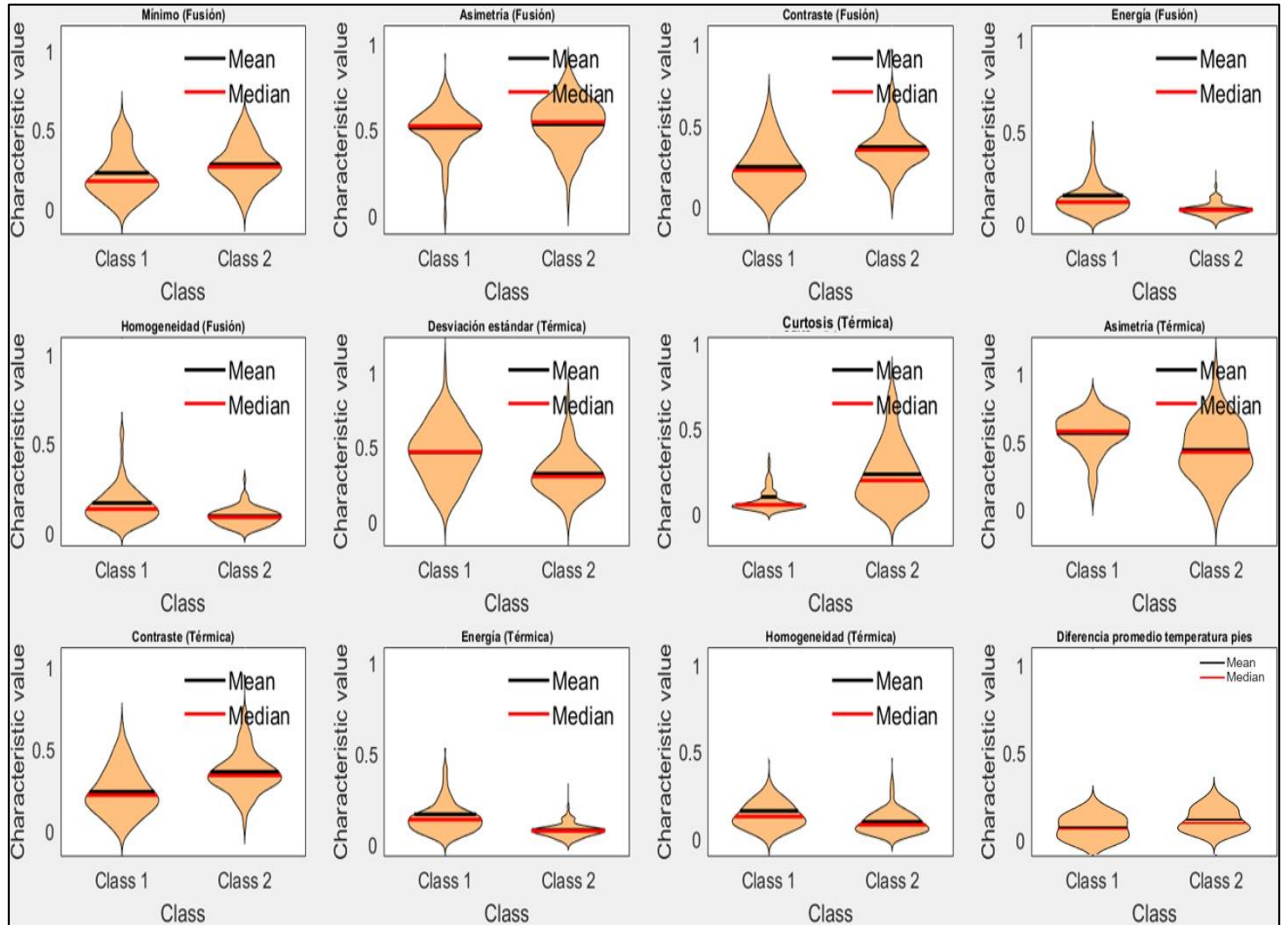
Figura 37. Diagramas de violín para característica relevante y no relevante



Fuente: Elaboración propia.

Asimismo, en la siguiente serie de diagramas de violín contenida en la figura 38 se presentan el grupo restante de características donde se puede observar visualmente una diferencia clara en la distribución de valores entre las clases, lo cual respalda estadísticamente su capacidad discriminativa para la clasificación entre pies sanos y pies diabéticos.

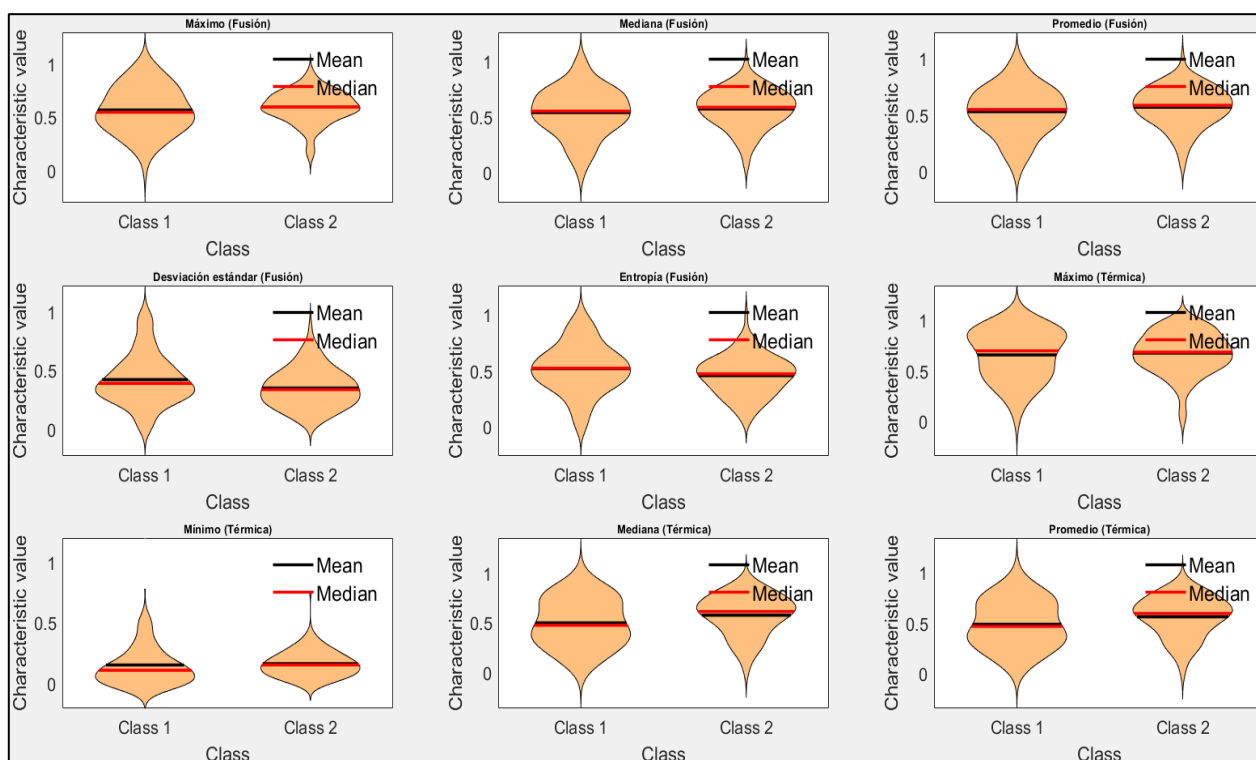
Figura 38. Resultados diagramas de violín características significativas.



Fuente. Elaboración propia.

Por otro lado, en la Figura 39 se presentan los diagramas de violín de las características que no mostraron diferencias estadísticamente significativas entre las clases. Se observa que sus distribuciones tienden a solaparse considerablemente, lo cual explica su limitada capacidad discriminativa frente a las características significativas. Estos resultados complementan el análisis, ya que permiten descartar variables poco informativas en el proceso de clasificación.

Figura 39. Resultados diagramas de violín características NO significativas.



Fuente. Elaboración propia.

Los resultados de ANOVA mostraron que 13 de las 23 características presentan diferencias estadísticamente significativas, es decir, tienen potencial para discriminar entre pies sanos y pies diabéticos y concuerdan con los gráficos mostrados en la figura 38 y 39. El veredicto de ANOVA es registrado en la tabla 7.

Tabla 7. Resultados ANOVA para las 23 características.

Característica	ANOVA Resultado
Máximo (Fusión)	NO significativa
Mínimo (Fusión)	SI significativa
Mediana (Fusión)	NO significativa
Promedio (Fusión)	NO significativa
Desviación estándar (Fusión)	NO significativa
Curtosis (Fusión)	NO significativa
Asimetría (Fusión)	SI significativa
Contraste (Fusión)	SI significativa
Energía (Fusión)	SI significativa
Homogeneidad (Fusión)	SI significativa

Entropía (Fusión)	NO significativa
Máximo (Térmica)	NO significativa
Mínimo (Térmica)	NO significativa
Mediana (Térmica)	NO significativa
Promedio (Térmica)	NO significativa
Desviación estándar (Térmica)	SI significativa
Curtosis (Térmica)	SI significativa
Asimetría (Térmica)	SI significativa
Contraste (Térmica)	SI significativa
Energía (Térmica)	SI significativa
Homogeneidad (Térmica)	SI significativa
Diferencia promedio temperatura pies	SI significativa
Entropía (Térmica)	SI significativa

Fuente: Elaboración propia.

Además, se observó que algunos pacientes diabéticos presentaban una diferencia significativa entre los promedios de intensidad de ambos pies, lo cual puede sugerir alteraciones en la circulación o el funcionamiento normal del pie. Es importante aclarar que los valores presentados no corresponden a grados Celsius reales, sino a intensidades relativas obtenidas del procesamiento de las imágenes térmicas. En la tabla 8, los valores resaltados identifican aquellos casos en los que la diferencia promedio de intensidad entre ambos pies supera el máximo habitual observado en la mayoría de los sujetos sanos. Si bien en algunos casos aislados de la población sana también se evidenciaron valores elevados, estos fueron significativamente menos frecuentes en comparación con los pacientes diabéticos. La relevancia radica en que, aun sin representar valores absolutos de temperatura, las diferencias relativas fueron consistentemente mayores en pacientes diabéticos frente a los sujetos sanos, lo cual sugiere un patrón clínicamente significativo.

Tabla 8. Característica diferencia promedio de temperatura entre pies para fragmento de pacientes seleccionados.

Pa ciente	1	2	3	4	5	6	7	...	16	17	18	19	20	21	22	23
Sa no	4. 61	3. 39	7.4 8	11. 14	5.2 3	7.1 3	5. 40	...	7.5 1	5.2 4	11. 09	12. 40	8. 92	5. 92	5. 04	7.1 6
Diabé tico	3. 11	1. 17	12. 38	5.9 5	18. 26	10. 78	6. 88	...	43. 14	32. 40	11. 65	4.3 7	2. 38	3. 53	6. 98	19. 85

Fuente: Elaboración propia.

No obstante, debe considerarse que las imágenes térmicas empleadas reportan intensidades relativas con una escala interna propia para cada paciente, dado que no se contó con información radiométrica calibrada. Esto implica que los valores obtenidos reflejan contrastes dentro de cada imagen más que magnitudes comparables de forma directa entre individuos. Por ello, los resultados aquí expuestos deben interpretarse como indicativos de tendencias y diferencias relativas dentro de los grupos, más que como medidas absolutas de temperatura homogéneas entre pacientes.

A partir de los análisis realizados, que incluyeron pruebas estadísticas como ANOVA y la inspección visual mediante diagramas de violín, se seleccionaron dos conjuntos de características para su posterior evaluación. El primero está compuesto por las 23 características listadas en la Tabla 7, mientras que el segundo incluye 13 características identificadas como estadísticamente significativas por ANOVA, ambos con las características con normalización aplicada de 0 a 1. Esta selección tiene como objetivo comparar el desempeño de ambos conjuntos durante la etapa de clasificación y validar la relevancia de las características seleccionadas.

3.3. CLASIFICACIÓN ENTRE PIES SANOS Y PIES DIABÉTICOS

Con los conjuntos de características ya discriminados, se procedió con la fase de clasificación, cuyo objetivo fue establecer modelos capaces de diferenciar correctamente entre pacientes con y sin riesgo de desarrollar pie diabético.

Dado que estudios previos, como el de Muhammad Adam Abdul Rahim y el trabajo de Arteaga N., Hernández G., Ortega J. y Ruiz J., reportaron resultados prometedores utilizando máquinas de vectores de soporte (SVM) con precisiones de hasta 89,39 % para el primer trabajo y métricas destacadas del segundo como una exactitud del 89,21%, precisión del 96,15%, sensibilidad del 81,93% y F1-score de 88,34% se decidió emplear inicialmente este clasificador como referencia para el presente estudio.

No obstante, con el fin de ampliar el análisis y explorar si otros métodos podrían superar el rendimiento de SVM en este contexto, se implementó paralelamente una serie de clasificadores adicionales. Por un lado, se evaluaron modelos estadísticos tradicionales como Naive Bayes, LDA (Análisis Discriminante Lineal), QDA (Análisis Discriminante Cuadrático), K-Nearest Neighbors (KNN), así como la implementación de clasificadores basados en redes neuronales, específicamente el Perceptrón Multicapa, la red Patternet y la Red de Base Radial (RBF), cuyo potencial para modelar relaciones no lineales y patrones complejos podría ofrecer ventajas en la clasificación de datos biomédicos.

Adicionalmente, se aplicaron técnicas de regularización durante el entrenamiento, tales como la regularización por caída (dropout) y la parada anticipada (early stopping), con el objetivo de mejorar la capacidad de generalización de los modelos y mitigar el sobreajuste.

Del total de 415 imágenes disponibles en la base de datos, se descartaron aquellas capturadas bajo condiciones de estimulación térmica por frío, las cuales correspondían exclusivamente a pacientes con diagnóstico de diabetes. Esta decisión se fundamentó en dos razones principales. En primer lugar, se busca desarrollar un sistema de clasificación que opere de forma no invasiva y fácilmente replicable en entornos clínicos reales; por tanto, incluir imágenes obtenidas tras una prueba de estrés térmico implicaría condicionar el uso del sistema a la aplicación previa de dicho estímulo, lo cual no es deseable desde el punto de vista práctico ni ético, ya que en seres humanos la exposición prolongada o repetida al frío, aun sin causar molestias inmediatas, puede producir lesiones nerviosas permanentes [85], además de requerir consentimiento explícito del paciente. En

segundo lugar, estas imágenes corresponden a un enfoque de termografía activa, ya que la estimulación por frío induce un estado fisiológico transitorio distinto al de reposo térmico. Incluir este tipo de imágenes podría introducir sesgos en el entrenamiento de los clasificadores, al mezclar condiciones de adquisición heterogéneas y no representativas del enfoque pasivo que se adopta en este trabajo. Además, esta exclusión permite que los resultados obtenidos sean comparables con los reportados en estudios previos que también emplearon únicamente imágenes adquiridas en condiciones pasivas.

Una vez excluidas estas imágenes, la base de datos se redujo en gran medida y, con el objetivo de mantener un equilibrio equitativo entre clases que minimizara el sesgo por desbalance de datos, se seleccionaron 200 imágenes para la fase de entrenamiento (100 de pacientes sanos y 100 de pacientes con diagnóstico de diabetes). Aunque este número puede considerarse limitado, resulta acorde con estudios previos en el área, incluso mayor que trabajos como el de Ferrin [14], quien trabajó con tan solo 17 pacientes, de los cuales, solo 7 eran diabéticos. En este contexto, más que la cantidad absoluta de muestras, lo relevante es la preservación de la proporcionalidad entre categorías y la homogeneidad en las condiciones de captura, lo cual permite entrenar y evaluar clasificadores de manera consistente. además, dentro del grupo de pacientes diabéticos, se consideraron casos clasificados en las categorías R0, R1 y R2, en función del riesgo de desarrollar complicaciones asociadas al pie diabético.

Aunque en la literatura es común emplear una partición 80/20 [86], en este trabajo se adoptó un 70% de entrenamiento y un 30% de prueba, con el fin de disponer de un conjunto de evaluación más robusto y minimizar el riesgo de sobreentrenamiento, dada la cantidad limitada de muestras disponibles. La separación de los conjuntos se realizó de manera fija, es decir, los mismos pacientes se mantuvieron en los respectivos subconjuntos (entrenamiento y prueba) a lo largo de todas las ejecuciones. Esta decisión tuvo como objetivo asegurar condiciones homogéneas para la comparación entre clasificadores, garantizando que cualquier diferencia en el rendimiento se debiera al comportamiento del modelo y no a la variabilidad en los datos de entrada.

3.3.1. Máquinas de Vectores de Soporte

Para la fase inicial de clasificación, se optó por emplear Máquinas de Vectores de Soporte (SVM) debido a su eficacia comprobada en trabajos previos relacionados con la detección de pie diabético, utilizando como insumo los dos conjuntos de características extraídas previamente.

Con el fin de establecer una arquitectura adecuada para el clasificador, se comenzó evaluando el impacto de distintas funciones núcleo, específicamente aquellas comúnmente referenciadas en la literatura: el núcleo lineal (ecuación 31), el núcleo polinomial de segundo grado (ecuación 32) y el núcleo de base radial (RBF, ecuación 33).

Además del tipo de núcleo, el entrenamiento de una SVM requiere resolver un problema de optimización cuadrática, lo que puede volverse computacionalmente costoso a medida que aumenta el tamaño del conjunto de datos. Por esta razón, se selecciona el algoritmo Sequential Minimal Optimization (SMO), procedimiento que divide el problema global de optimización en una serie de subproblemas más simples, resolviendo en cada iteración únicamente dos variables, además de que es ampliamente utilizado por su simplicidad computacional [63].

Dado que las Máquinas de Vectores de Soporte no generan directamente probabilidades, fue necesario aplicar un método de calibración posterior para obtener estimaciones probabilísticas interpretables y asignar una opción de rechazo, un parámetro fundamental en el ámbito médico y

que los antecedentes analizados no mencionan. Para ello, se empleó la transformación logística propuesta por Platt, la cual permite convertir las salidas del modelo (originalmente márgenes de decisión o score) en probabilidades continuas con valores entre 0 y 1 para la clase positiva (en este caso, “pie diabético”). A partir de estas probabilidades calibradas, se implementó una estrategia basada en umbrales para la toma de decisiones: se definió una zona de rechazo con el objetivo de evitar decisiones clasificatorias en situaciones de incertidumbre. Específicamente, solo se clasificaron como “pie diabético” aquellas muestras cuya probabilidad fue mayor o igual al 80 %, y como “sano” aquellas con una probabilidad menor o igual al 20 %. Las muestras con probabilidades intermedias (entre 20 % y 80 %) fueron etiquetadas como “rechazadas”, lo cual permite aumentar la confiabilidad del sistema, especialmente en aplicaciones clínicas donde es preferible abstenerse de clasificar a riesgo de cometer errores. En la literatura, Charoenphakdee et al. [87] y Mozannar y Sontag [88] destacan la importancia de incorporar opciones de rechazo en sistemas de clasificación como un mecanismo para reducir riesgos en escenarios de incertidumbre. Sin embargo, no se identificaron estudios con enfoque médico que establezcan porcentajes de umbral recomendados para dicha opción. En este trabajo, los valores adoptados fueron definidos de manera experimental, considerando el contexto clínico de la detección del pie diabético y priorizando la seguridad del paciente sobre la exhaustividad clasificatoria.

En todos los casos, se mantuvo constante la proporción de 70 % de las muestras para entrenamiento y 30 % para validación, garantizando condiciones homogéneas para la etapa de comparación de clasificadores. Los resultados para la SVM son:

Conjunto	Kernel	Exactitud	Precisión	Recall	Especificidad	F1-score	Nro. rechazos
23 características	Lineal	75%	72.73%	80%	70%	76.19%	32
	Cuadrático	85%	75%	85.71%	84.62%	80%	40
	RBF	63.33%	58.7%	90%	36.67%	71.05%	43
13 características	Lineal	75%	70.27%	86.67%	63.33%	77.61%	31
	cuadrático	75%	70.27%	86.67%	63.33%	77.61	51
	RBF	65%	61.54%	80%	50%	69,6%	38

Tabla 9. Resultados SVM para diferentes kernel.

De acuerdo con la tabla 9, se evaluaron clasificadores SVM con kernels lineal, cuadrático y RBF, usando conjuntos de 23 y 13 características. El kernel cuadrático con 23 características mostró el mejor desempeño global, con una exactitud del 85 % y F1-score del 80 %, aunque con un número elevado de rechazos (40), mientras que el RBF, si bien alcanzó el mayor recall (90 %), tuvo la peor especificidad y mayor número de errores. Al reducir a 13 características, los kernels lineal y cuadrático empataron en exactitud (75 %), pero el modelo lineal mostró menos rechazos, sugiriendo mayor estabilidad con menor complejidad. En este contexto, la cantidad de rechazos es un indicador relevante para evaluar el equilibrio del clasificador. Sin embargo, dado que estos resultados se basan en una sola partición del conjunto de datos, se implementó una validación cruzada para obtener un análisis más robusto del comportamiento general del modelo, para ello, se probaron 200 particiones aleatorias de los datos para sus respectivas categorías de sano y diabético.

Promedio de resultados						
Conjunto	Kernel	Exactitud	Precisión	Recall	Especificidad	F1-score
23 características	Lineal	82,2%	81,9%	83,3%	81,2%	82,3%
	Polinomial	80,2%	78,6%	83,8%	76,7%	80,9%
	RBF	64,7%	60,9%	92,4%	37,1%	72,0%

13 características	Lineal	82,1%	80,6%	85,2%	79,1%	82,6%
	Polinomial	76,0%	75,4%	77,9%	74,0%	76,4%
	RBF	72,5%	69,3%	84,3%	60,7%	75,2%

Tabla 10. Promedio de métricas de evaluación para SVM en clasificación binaria.

Máximos alcanzados						
Conjunto	Kernel	Exactitud	precisión	Recall	Especificidad	F1-score
23 características	Lineal	95,0%	96,6%	93,3%	96,7%	94,9%
	Polinomial	91,7%	87,9%	96,7%	86,7%	92,1%
	RBF	81,7%	91,3%	70,0%	93,3%	79,2%
13 características	Lineal	91,7%	87,9%	96,7%	86,7%	92,1%
	Polinomial	86,7%	92,3%	80,0%	93,3%	85,7%
	RBF	86,7%	80,6%	96,7%	76,7%	87,9%

Tabla 11. Mejor caso de clasificación con SVM según kernel y número de características.

Al evaluar el desempeño de los clasificadores SVM sobre el conjunto de datos con 23 características en la tabla 10 y 11, con particiones aleatorias, el kernel lineal fue el de mejor rendimiento, tanto en promedio (exactitud: 82,2 %, precisión: 81,9 %, sensibilidad: 83,3 %, especificidad: 81,2 %, F1-score: 82,3 %) como en su mejor caso individual, alcanzando una exactitud de 95,0 %, precisión de 96,6 %, sensibilidad de 93,3 % y un F1-score de 94,9 %. Estos resultados superan los reportados en trabajos previos de seguimiento, donde el mejor desempeño alcanzado fue una exactitud de 89,2 %, precisión de 96,2 %, sensibilidad de 81,9 % y F1-score de 88,3 %.

En ese sentido, también se exploró la reducción del conjunto a 13 características, pero los resultados obtenidos no representaron una mejora. Aunque el kernel lineal mantuvo un desempeño aceptable, tanto el valor promedio como el mejor caso presentaron disminuciones claras en precisión, sensibilidad y F1-score respecto al conjunto completo. Por ejemplo, el mejor F1-score obtenido con 13 características fue 92,1 %, frente a 94,9 % con 23 características. Esto sugiere que la reducción de características no fue favorable en términos de rendimiento diagnóstico para el caso de las SVM, por lo que el conjunto completo resulta más adecuado para maximizar la confiabilidad del modelo.

El mejor resultado individual obtenido presenta tres errores de clasificación, como se muestra en la Tabla 12. De estos, uno corresponde a un falso positivo, es decir, un paciente sano que fue clasificado erróneamente como portador de pie diabético. Más preocupante aún, se identificaron dos falsos negativos, lo que implica que dos pacientes diabéticos fueron diagnosticados incorrectamente como sanos. Aunque lo ideal sería no tener ningún falso negativo, dado que estos pueden tener mayores consecuencias graves, este resultado representa una mejora significativa frente al estudio de Adahm, quien reportó un falso positivo y seis falsos negativos en su modelo.

Tipo	Valor
VP (Verdadero Positivo)	28
FN (Falso Negativo)	2
FP (Falso Positivo)	1
VN (Verdadero Negativo)	29

Tabla 12. Resultados matrices de confusión SVM para el mejor de los resultados.

Sin embargo, para el mejor resultado obtenido se observó que, al aplicar la opción de rechazo basada en umbrales de confianza ($\geq 0,8$ para clasificar como diabético y $\leq 0,2$ como sano), 28 muestras fueron rechazadas, mientras que 14 fueron clasificadas como diabéticas y 18 como sanas. Este hallazgo indica que, incluso en el escenario de mejor desempeño del modelo, existen casos cuya probabilidad de pertenencia no es lo suficientemente clara, y que podrían beneficiarse de una revisión adicional antes de emitir una decisión automática. No obstante, no fue posible establecer una comparación directa con los antecedentes revisados, ya que estos no reportan si aplicaron o consideraron mecanismos de rechazo en sus enfoques.

Dado el contexto clínico de la aplicación, se priorizó el análisis de métricas como la precisión, la sensibilidad y el F1-score, ya que ofrecen una evaluación más relevante que la exactitud general. La sensibilidad (ecuación 39) indica la capacidad del modelo para detectar correctamente los casos de pie diabético, lo cual es crucial para evitar falsos negativos que podrían tener consecuencias graves. La precisión (ecuación 38) refleja cuántos de los casos identificados como positivos realmente lo son, ayudando a reducir falsos positivos. El F1-score (ecuación 40), al combinar ambas métricas, permite evaluar el equilibrio entre detectar casos positivos y evitar errores, lo cual es especialmente importante en contextos médicos donde no todos los errores tienen el mismo peso. No obstante, se muestra métricas como exactitud y especificidad para efectos comparativos con trabajos previos.

3.3.2. Clasificadores estadísticos

Para los clasificadores estadísticos se implementó un sistema de clasificación considerando el costo asociado a cada tipo de error y acierto, además de las probabilidades a priori de las clases. Para ello, se definió una matriz de costos que penaliza con mayor severidad los errores en los que un paciente diabético es clasificado como sano, dado el riesgo que implica esta equivocación.

$$Cost = \begin{bmatrix} 0 & 2 \\ 4 & 0 \end{bmatrix} \quad (46)$$

La matriz de costos definida en este trabajo responde a un criterio proporcional entre tipos de error. En concordancia con lo planteado en la literatura sobre cost-sensitive learning [89], las clasificaciones correctas se asignaron con costo cero, al no generar consecuencias clínicas ni prácticas. En cuanto a los errores, se estableció un costo de 2 para el falso positivo (clasificar a un paciente sano como diabético) y de 4 para el falso negativo (clasificar a un paciente diabético como sano). Esta relación refleja que el segundo escenario implica un riesgo clínico considerablemente mayor, ya que podría retrasar el diagnóstico y tratamiento oportuno, mientras que el primero solo conlleva exámenes adicionales de bajo impacto. En otras palabras, el error más crítico fue ponderado con un valor aproximadamente el doble del menos severo, siguiendo el principio de que en escenarios binarios los falsos negativos deben ser penalizados de manera más estricta que los falsos positivos [89]. Así mismo, se establecieron probabilidades a priori para las clases “sano” y “pie diabético”, de modo que el clasificador tuviera en cuenta la distribución real de los datos que en este caso corresponde a 50% para sanos y 50% para no sanos por el balanceo de clases.

$$Priori = [0.5 \quad 0.5] \quad (47)$$

Estos elementos se incorporaron directamente en los clasificadores estadísticos, de manera que no solo se basaran en la probabilidad posterior máxima, sino en la minimización del riesgo esperado. Es decir, cada clasificador, al recibir un nuevo dato, calcula el riesgo esperado asociado a asignar dicho dato a cada clase, usando la matriz de costos y las probabilidades a priori. La clase con menor riesgo era seleccionada como predicción.

No obstante, para aumentar la confiabilidad del sistema, se estableció nuevamente un umbral de decisión de 0.8. Esto significa que un clasificador solo emite una decisión si la probabilidad posterior corregida (o el puntaje asociado) supera dicho umbral para alguna de las clases. En caso contrario, el resultado es “rechazado”, lo que indica que el sistema no tiene suficiente certeza para clasificar ese caso. Este enfoque permite una toma de decisiones más cautelosa en contextos sensibles como el diagnóstico médico, priorizando la precisión por encima de la cobertura completa.

Luego de aplicar esta lógica, se recolectaron las decisiones emitidas por cada clasificador para el conjunto de datos con 23 características y 13 características. Estas decisiones incluyen los casos clasificados como “sano”, “pie diabético” y “rechazado” cuando el umbral de confianza no fue alcanzado. Las Tablas 13 y 14 presentan el resumen de estas decisiones.

	Naive Bayes	LDA	QDA	KNN
Pie diabético	37	26	31	28
Rechazado	4	18	1	22
Sano	19	16	28	10

Tabla 13. Resultados clasificadores estadísticos para 23 características con opción de rechazo.

	Naive Bayes	LDA	QDA	KNN
Pie diabético	34	25	34	26
Rechazado	7	20	6	23
Sano	19	15	20	11

Tabla 14. Resultados clasificadores estadísticos para 13 características con opción de rechazo.

En general, se observa que el clasificador QDA fue el que menos rechazos generó en ambos escenarios (1 y 6 instancias rechazadas respectivamente), lo que sugiere una mayor seguridad del modelo en sus predicciones. Por el contrario, LDA y KNN presentaron un número considerable de rechazos, particularmente KNN con 22 y 23 casos, lo que indica mayor incertidumbre en su clasificación. Sin embargo, es necesario analizar estas decisiones en conjunto con los resultados de las métricas implementadas para estos clasificadores.

Clasificador	Exactitud	Precisión	Recall	Especificidad	F1-score
Naive Bayes	70.0%	80.0%	53.3%	86.6%	64.0%
LDA	73.3%	85.0%	56.6%	90.0%	68.0%
QDA	85.0%	86.2%	83.3%	86.6%	84.7%
K-NN	73.3%	88.8%	53.3%	93.3%	66.6%

Tabla 15. Métricas clasificadores estadísticos para 23 características.

Clasificador	Exactitud	Precisión	Recall	Especificidad	F1-score
Naive Bayes	73.3%	85.0%	56.6%	90.0%	68.0%
LDA	76.6%	90.0%	60.0%	93.3%	72.0%
QDA	83.3%	91.6%	73.3%	93.3%	81.5%
K-NN	73.3%	88.8%	53.3%	93.3%	66.6%

Tabla 16. Métricas clasificadores estadísticos para 13 características.

Al analizar los resultados en función de las métricas clásicas de evaluación (Tablas 15 y 16), se puede evidenciar, primero, que la reducción de características favorece al aumento en el desempeño de los clasificadores (resaltador verde) y como segundo dato relevante, que el clasificador QDA

logró una alta precisión y sensibilidad en ambos conjuntos de características, así como un F1 Score cercano al de los antecedentes y superior al resto de los clasificadores.

Los resultados también se analizaron a través de matrices de confusión (Tablas 17 y 18), permitiendo identificar el comportamiento específico de cada clasificador.

CLASIFICADOR	BAYES	LDA	QDA	KNN
Verdadero positivo	14	17	25	14
Falso negativo	16	13	5	16
Verdadero negativo	26	27	26	28
Falso positivo	4	3	4	2

Tabla 17. Resultados matrices de confusión 23 características.

CLASIFICADOR	BAYES	LDA	QDA	KNN
Verdadero positivo	15	16	20	17
Falso negativo	15	14	10	13
Verdadero negativo	27	29	27	28
Falso positivo	3	1	3	2

Tabla 18. Resultados matrices de confusión 13 características.

Para el caso de las 23 características, el clasificador QDA alcanzó el mayor número de verdaderos positivos (25) y presentó un desempeño algo desfavorable al registrar más falsos negativos (5) que falsos positivos (4), ya que los falsos negativos representan un mayor riesgo clínico. Sin embargo, al compararlo con el antecedente reportado por Adahm, cuyos resultados incluyeron 1 falso positivo y 6 falsos negativos, se observa que la diferencia en desempeño es relativamente pequeña, lo cual permite considerar que el modelo propuesto ofrece un comportamiento competitivo con el conjunto de 23 características frente a trabajos previos similares.

3.3.3. Clasificadores basados en redes neuronales

Para el desarrollo de los clasificadores basados en redes neuronales se estableció una configuración común con el fin de asegurar condiciones de entrenamiento comparables entre las distintas arquitecturas. Del total de los datos, se utilizó un 70% para el conjunto de entrenamiento y validación, el cual se subdividió internamente en un 80% para entrenamiento propiamente dicho y un 20% para validación, mientras que el 30% restante se destinó a prueba.

Los clasificadores fueron entrenados utilizando un número máximo de 1000 épocas, parámetro que nunca fue alcanzado ya que por otro lado se dejó establecido un error mínimo deseado de $1e-5$, lo cual define el criterio de parada del entrenamiento. Además, se estableció una tasa de aprendizaje de 0.001 y una constante de momento de 0.9, parámetros que influyen directamente en la dinámica de ajuste de los pesos durante el entrenamiento, especialmente en redes como el Perceptrón Multicapa y la red Patternet. La elección de estas dos se fundamentó en la naturaleza y el tamaño de la base de datos empleada, la cual está compuesta por vectores de características estadísticas y morfológicas derivados de imágenes, y no por las imágenes crudas en sí mismas. Bajo este escenario, redes neuronales más complejas, suelen requerir un mayor volumen de datos para evitar sobreajuste, mientras que las redes clásicas permiten un entrenamiento más estable y reproducible con bases pequeñas o moderadas, según Haykin y Bishop en sus obras [90], [91]. Asimismo, Dreiseitl y Ohno-Machado [92] muestran que estas arquitecturas han sido ampliamente utilizadas en la literatura biomédica como clasificadores de referencia en escenarios exploratorios, lo que

facilita la comparabilidad con estudios previos y reduce la necesidad de un ajuste excesivo de hiperparámetros. Adicionalmente, MATLAB ofrece implementaciones optimizadas y ampliamente validadas de estas redes, lo cual garantiza reproducibilidad y consistencia en los experimentos realizados. Por lo tanto, el criterio principal para su selección fue equilibrar simplicidad, solidez metodológica y adecuación al tamaño de la muestra disponible.

Para garantizar la reproducibilidad de los resultados, se fijó una semilla de aleatoriedad. Este paso es fundamental en el entrenamiento de redes neuronales en MATLAB, ya que los pesos iniciales de la red son asignados aleatoriamente, y esta inicialización puede tener un impacto significativo en el desempeño final del clasificador. Dependiendo de los valores iniciales de los pesos y sesgos, la red puede converger a soluciones distintas, algunas con mejores resultados que otras.

Es así como se realizó una exploración previa utilizando más de 200 valores de semilla distintos en MATLAB y cambiando la función de entrenamiento. Se evaluaron los resultados obtenidos para cada una de estas semillas y se observó que la semilla 126 y 199 producía una convergencia estable y resultados consistentemente buenos en cuanto a métricas de rendimiento como la precisión, la sensibilidad y el F1 Score. En la tabla 19 se muestran las dos mejores funciones de entrenamiento y las semillas con los resultados más altos para las métricas evaluadas.

Modelo	Mejor Semilla	Precisión	Sensibilidad	Función
Perceptron	126 (trainbfg)	80.0%	70%	trainbfg
Patternnet	199 (trainlm)	83.3%	80%	trainlm
RBF	199 (trainlm)	78.3%	76.7%	trainlm
Parada	199 (trainlm)	75.0%	73.7%	trainlm

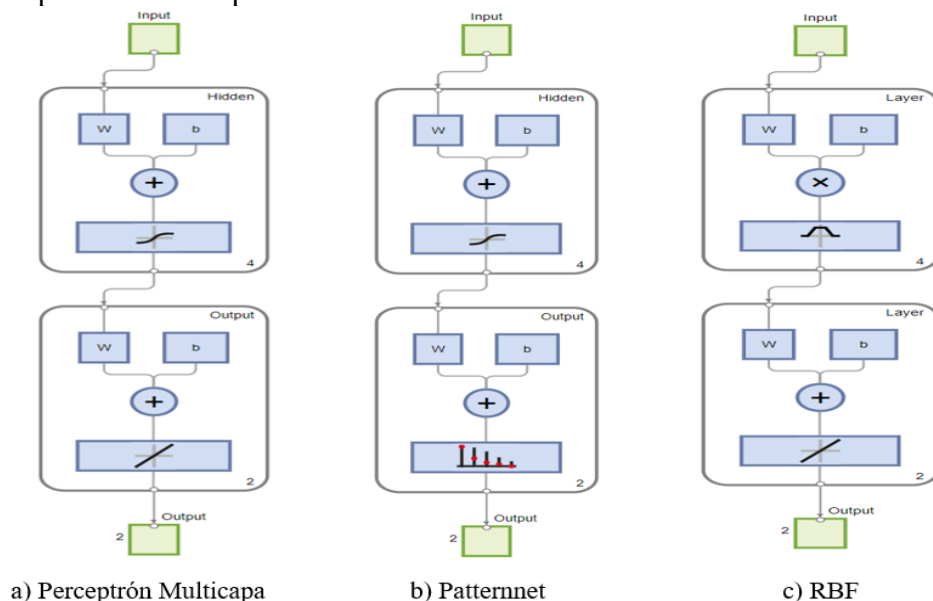
Tabla 19. Evaluación semilla y función para clasificadores con redes neuronales.

De acuerdo con la teoría de redes neuronales, una sola capa oculta es suficiente para aproximar cualquier función continua, y como máximo, dos pueden resolver problemas más complejos. Siguiendo este principio, se evaluaron arquitecturas con una capa oculta y diferentes cantidades de neuronas, dando el caso en que para 1, 2 y 3 neuronas en la capa oculta, se obtuvieron precisiones que oscilaron entre el 75% y el 85% y un valor de sensibilidad entre 70% y 82%.

No obstante, al utilizar un valor de 4 neuronas en la capa oculta, se observaron mejoras significativas y consistentes en el desempeño de la red, alcanzando precisiones estables entre el 80% y el 88.5% para ambas métricas, lo cual motivó la elección de esta configuración para los experimentos finales y se utilizó la función de entrenamiento trainlm (Levenberg–Marquardt) en base a los resultados vistos en la Tabla 18, donde se observaron los mejores porcentajes.

El resultado fue la selección de tres arquitecturas neuronales, representadas en la Figura 40. En (a) se muestra el Perceptrón Multicapa (MLP), una red de tipo feedforward con función de activación sigmoide en la capa oculta y salida lineal. En (b) se presenta la red Patternnet, variante del MLP diseñada para problemas de clasificación de patrones, cuya principal diferencia radica en la inclusión de la función softmax en la capa de salida, lo que permite obtener probabilidades normalizadas por clase. Finalmente, en (c) se ilustra la Red de Funciones de Base Radial (RBF), que emplea funciones radiales en la capa oculta y salida lineal, configurada con un valor de spread igual a 1 y un paso de entrenamiento inicial de 1, lo que define su sensibilidad y la velocidad de ajuste de parámetros.

Figura 40. Arquitectura final para clasificadores con redes neuronales.



Fuente: Generado por MATLAB.

Adicionalmente, se incluyeron dos variantes de la red Perceptrón Multicapa con el objetivo de mitigar el sobreajuste. La primera corresponde a la regularización por caída, el cual ajusta los pesos de la red penalizando valores grandes. La segunda corresponde a la regularización por parada temprana, en la que el entrenamiento se detiene de manera automática cuando el error de validación deja de disminuir. La arquitectura de la red permanece inalterada, diferenciándose únicamente en la estrategia de entrenamiento utilizada

Una vez definida la arquitectura de cada una de las redes neuronales utilizadas como clasificadores incluyendo parámetros como el número de neuronas ocultas, función de activación, tasa de aprendizaje y algoritmo de entrenamiento. Se procedió a evaluar su desempeño utilizando los mismos conjuntos de datos previamente empleados en los clasificadores estadísticos. Se mantuvo la división original, sin realizar mezclas adicionales, con el objetivo de asegurar una comparación directa y justa entre enfoques. Las pruebas se realizaron tanto con el conjunto completo de 23 características como con la versión reducida de 13 características, previamente seleccionadas. En ambos casos, se calcularon las métricas priorizadas de evaluación: precisión, sensibilidad y F1 Score, y métricas como la exactitud y especificidad para efectos comparativos con los trabajos previos. Los resultados se resumen en las Tablas 20 y 21.

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score
Perceptrón Multicapa	70.00 %	75.00 %	60.00 %	80.00 %	72.73 %
Patternnet	83.30 %	83.30 %	83.30 %	83.30 %	83.33 %
RBF	70.00 %	87.50 %	46.70 %	93.30 %	75.68 %
Regularización Caída	60.00 %	60.70 %	56.70 %	63.30 %	61.29 %
Parada Temprana	75.00 %	82.60 %	63.30 %	86.70 %	77.61 %

Tabla 20. Resultado métricas evaluativas con clasificadores para 23 características con redes neuronales.

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score
Perceptrón Multicapa	70.00 %	87.50 %	46.70 %	93.30 %	75.68 %
Patternnet	75.00 %	94.10 %	53.30 %	96.70 %	79.45 %
RBF	75.00 %	85.70 %	60.00 %	90.00 %	78.26 %
Regularización Caída	81.70 %	85.20 %	76.70 %	86.70 %	82.54 %
Parada Temprana	76.70 %	86.40 %	63.30 %	90.00 %	79.41 %

Tabla 21. Resultado métricas evaluativas con clasificadores para 13 características con redes neuronales.

Se observa que la red Patternnet obtuvo un rendimiento superior, en el conjunto de 23 características, mantuvo un equilibrio notable, mientras que otros clasificadores, como la RBF, si bien presentaron valores altos en precisión y especificidad, mostraron una sensibilidad considerablemente más baja.

3.3.4. Proceso de validación cruzada para clasificadores estadísticos y redes neuronales

Una vez obtenidos los resultados de los clasificadores estadísticos y de redes neuronales utilizando una partición fija de entrenamiento y prueba (sin mezcla de datos), se procedió a realizar una validación cruzada con el fin de evaluar la robustez y generalización de cada clasificador como se realizó para las SVM.

Para ello, se replanteó la distribución de los datos, mezclando aleatoriamente las muestras y dividiendo nuevamente los conjuntos de entrenamiento y prueba en cada intento. Esta técnica permite observar cómo se comportan los clasificadores ante variaciones en los datos, imitando condiciones más cercanas a escenarios reales donde la distribución no es necesariamente constante.

Durante el proceso se realizaron múltiples iteraciones, cada una con una nueva partición aleatoria de los datos, se conservaron las proporciones entre clases para evitar sesgos por desbalanceo, en cada iteración, se entrenaron nuevamente los clasificadores y se evaluaron sobre el nuevo conjunto de prueba. En el caso de los clasificadores estadísticos, se realizaron 20 iteraciones para cada conjunto de características (23 y 13 respectivamente). Los resultados promedio de estas ejecuciones para datos de prueba se presentan en las Tablas 22 y 23, con un rango de valores entre el 78% y 90% en las métricas evaluadas y colocando al clasificador QDA como uno de los de mejor respuesta en precisión, sensibilidad y F1-score (marcador verde), permitiendo observar el comportamiento de cada clasificador bajo condiciones de variabilidad controlada.

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score
BAYES	76.00 %	74.14 %	71.00 %	71.00 %	74.42 %
LDA	79.33 %	76.45 %	73.67 %	73.67 %	78.05 %
QDA	81.17 %	84.16 %	85.33 %	85.33 %	81.94 %
KNN	77.00 %	75.88 %	73.67 %	73.67 %	76.00 %

Tabla 22. Promedio de 20 intentos por validación cruzada para clasificadores estadísticos con 23 características.

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score
BAYES	75.83 %	80.89 %	84.00 %	84.00 %	77.72 %
LDA	80.33 %	80.16 %	82.67 %	92.67 %	82.62 %
QDA	80.50 %	82.49 %	83.33 %	83.33 %	81.06 %
KNN	76.17 %	80.02 %	82.33 %	82.33 %	77.65 %

Tabla 23. Promedio de 20 intentos por validación cruzada para clasificadores estadísticos con 13 características

Además, el valor máximo reportado entre las iteraciones se obtuvo a través del clasificador QDA, con los siguientes resultados reportados en la tabla 24:

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score	Nro. rechazos
QDA	90.0%	92.9%	86.7%	93.3%	89.7%	7

Tabla 24. Mejor desempeño entre clasificadores estadísticos.

En el caso de las redes neuronales, el proceso fue más exhaustivo: se realizaron más de 100 intentos utilizando diferentes semillas aleatorias, con el fin de analizar el impacto de la inicialización sobre el desempeño. De estos, se seleccionó la semilla que ofrecía los resultados más consistentes y con mejores métricas promedio. Se evidenció que las redes neuronales conservaron un comportamiento más estable en un rango de valores entre el 80% y el 88.5% para métricas como exactitud, precisión, sensibilidad y F1 Score, lo cual sugiere una alta capacidad de adaptación del modelo a distintas configuraciones de datos.

Los resultados promedios de los mejores modelos seleccionados tras validación cruzada se presentan en las Tablas 25 y 26. Se destaca que la red RBF con 13 características y el uso de parada temprana con 23 características son las configuraciones que alcanzan un mejor balance entre las métricas.

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score
Perceptrón Multicapa	74.17 %	71.87 %	81.33 %	67.00 %	75.86 %
Patternnet	75.00 %	74.91 %	77.67 %	72.33 %	75.67 %
RBF	75.00 %	76.63 %	72.67 %	77.33 %	74.35 %
Regularización Caída	71.50 %	71.25 %	72.67 %	70.33 %	71.56 %
Parada Temprana	80.17 %	82.23 %	77.67 %	82.67 %	79.40 %

Tabla 25. Promedio de 20 intentos por validación cruzada para redes neuronales con 23 características

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score
Perceptrón Multicapa	78.50 %	79.73 %	79.09 %	77.85 %	78.79 %
Patternnet	79.67 %	81.33 %	78.14 %	81.30 %	79.45 %
RBF	83.83 %	84.77 %	83.42 %	84.28 %	83.85 %
Regularización Caída	75.00 %	76.54 %	72.82 %	77.23 %	74.45 %
Parada Temprana	82.83 %	84.66 %	81.38 %	84.28 %	82.53 %

Tabla 26. Promedio de 20 intentos por validación cruzada para redes neuronales con 13 características.

Por último, se realiza un nuevo conteo de los resultados considerando la opción de rechazo y se guarda registro en la tabla 27. A través de este análisis, es posible evidenciar que, en algunos casos, existe una combinación desfavorable entre las métricas de desempeño y la cantidad de rechazos. Es decir, aunque se logran valores aceptables en precisión u otras métricas, estos resultados se alcanzan a costa de un número elevado de muestras rechazadas, lo cual limita la eficacia práctica de los modelos en un entorno real.

Red	Nro_Sano	Nro_Diabetic	Nro_Rechazado
Patternnet	21	16	23
RBF	21	10	29
Perceptron	28	9	23
Caída	25	23	12
Parada	21	15	24

Tabla 27. Conteo resultados redes neuronales.

El valor máximo reportado entre las iteraciones se obtuvo a través del clasificador RBF. En la tabla 28 se muestran los resultados:

Clasificador	Exactitud	Precisión	Sensibilidad	Especificidad	F1 Score	Nro. rechazos
RBF	90.0%	96.2%	81.9%	93.3%	89.3%	29

Tabla 28. Mejor desempeño entre clasificadores con redes neuronales.

3.3.5. Análisis de resultados

Se analizaron los resultados obtenidos tras evaluar distintos clasificadores: SVM, estadísticos y basados en redes neuronales, con el fin de validar la idoneidad de las características seleccionadas previamente a partir de estudios antecedentes. En dichos trabajos, estas características ya habían sido sometidas a procesos de selección y reducción, y los resultados actuales permiten corroborar que efectivamente son representativas y adecuadas para la clasificación del pie diabético.

Durante las pruebas, se evidenció que, al aplicar una reducción de características, las métricas de rendimiento de los clasificadores SVM disminuyeron levemente; sin embargo, este mismo ajuste produjo mejoras notables en el desempeño del resto de los clasificadores, especialmente en los estadísticos y los basados en redes neuronales. Esto sugiere que, si bien las SVM eran más sensibles a la pérdida de información por reducción, otros modelos lograron beneficiarse.

En cuanto a los resultados obtenidos, el mejor desempeño global se logró con el clasificador SVM, alcanzando una exactitud de 95,0 %, una precisión de 96,6 %, una sensibilidad de 93,3 % y un F1-score de 94,9 %. Entre los clasificadores estadísticos, se destaca el QDA que obtuvo una exactitud de 90 %, con una precisión de 92,9 %, una sensibilidad de 86,7 %, una especificidad de 93,3 % y un F1-score de 89,7 %. Por su parte, el clasificador basado en RBF también alcanzó una exactitud de 90 %, con precisión de 96,2 %, sensibilidad de 83,3 %, especificidad de 96,7 % y un F1-score de 89,3 %.

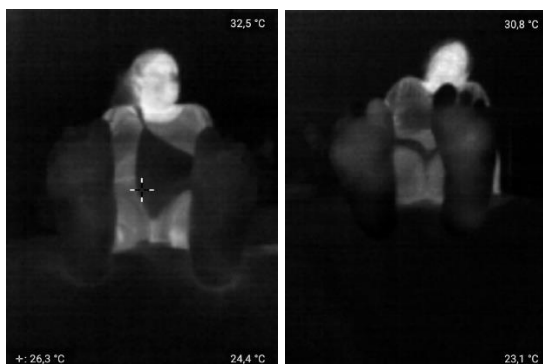
Estos resultados superan claramente el desempeño de los antecedentes revisados, los cuales reportaban una exactitud de 89,2 %, precisión de 96,2 %, sensibilidad de 81,9 % y un F1-score de 88,3 %, confirmando que la implementación de técnicas de fusión de imágenes como etapa previa a la extracción de características contribuyó significativamente a este resultado. La combinación de información del espectro visible e infrarrojo permitió generar representaciones más completas y discriminativas, lo que a su vez facilitó un aprendizaje más robusto por parte de los clasificadores. Esto valida que el uso de fusión multimodal no solo es viable, sino recomendable en sistemas de apoyo al diagnóstico clínico como el abordado en este estudio.

Al aplicar una opción de rechazo durante la clasificación, es decir, identificar como "rechazadas" aquellas muestras cuya probabilidad de pertenencia a una clase no superaba el umbral del 80 %, se obtuvo un comportamiento dispar entre los clasificadores evaluados. En este escenario, el clasificador SVM presentó 28 casos rechazados, el RBF rechazó 29, mientras que QDA solo rechazó 7.

Aunque SVM fue el modelo que alcanzó los mejores resultados en métricas, su alto número de rechazos indica que, a pesar de ser muy preciso cuando decide, lo hace con menor frecuencia bajo criterios estrictos de confianza. En contraste, el clasificador QDA, mantuvo un equilibrio adecuado entre rendimiento y confiabilidad al presentar solo 7 rechazos.

Para comprender mejor los casos en los que la clasificación resultó difícil, se realizó un análisis puntual de algunas de las muestras rechazadas. En particular, al revisar las imágenes correspondientes a la segmentación manual, se evidenció que ciertas imágenes térmicas no fueron capturadas con el cuidado o control necesarios. En algunos pacientes clasificados como R0 (sin riesgo visible), las imágenes térmicas mostraban una distribución de temperatura muy similar a la del entorno, dificultando la diferenciación entre el pie y el fondo. Esto sugiere posibles problemas en la toma de las imágenes, como una mala calibración de la cámara o condiciones ambientales inadecuadas. Algunos de estos casos se ilustran a continuación en la figura 41 para resaltar la importancia de un protocolo riguroso en la adquisición de datos térmicos:

Figura 41. Casos con bajo contraste térmico pie-fondo.



Fuente: Elaboración propia.

Además, se requiere que el pie esté completamente contenido dentro de una región mínima de 480×384 píxeles para garantizar un recorte adecuado; de lo contrario, partes del pie podrían quedar fuera del análisis después de la primera segmentación, realizada mediante el análisis de histogramas.

Por lo tanto, si bien SVM sigue siendo el más eficaz en términos de métricas puras, el QDA se considera el clasificador más idóneo en escenarios clínicos reales, ya que combina buen desempeño con un nivel de decisión más confiado.

Adicionalmente, se evaluó el desempeño de los clasificadores al ser aplicados sobre las imágenes segmentadas automáticamente mediante el algoritmo propuesto, manteniendo los mismos parámetros de entrenamiento y validación para los mejores resultados. En este escenario y para las mismas condiciones donde los clasificadores obtuvieron sus máximos resultados, la SVM obtuvo una precisión del 84,38 %, una sensibilidad del 90,00 % y un F1-score de 87,10 %, con un total de 46 rechazos. El clasificador QDA, por su parte, alcanzó una precisión de 82,14 %, una sensibilidad de 83,33 % y un F1-score de 80,65 %, con 11 rechazos. Estos resultados fueron inferiores a los obtenidos previamente con imágenes segmentadas manualmente, lo que evidencia que, si bien la segmentación automática permitió mantener un rendimiento aceptable, aún presenta ciertas limitaciones en cuanto a la conservación de información relevante para la clasificación. Esto pone de manifiesto no solo la necesidad de seguir perfeccionando los métodos automáticos de segmentación, sino también la importancia crítica de preservar la mayor cantidad posible de información significativa de los pies, ya que cualquier pérdida puede impactar negativamente en la capacidad diagnóstica del sistema.

3.3.6. Resultados adicionales

En las etapas iniciales de este trabajo, no se abordó inicialmente la discriminación entre dos clases, se desarrolló una clasificación de tres clases: pacientes sanos, pacientes R1 (aquellos con neuropatía sin isquemia), y pacientes R2 (casos isquémicos). Aunque esta estructura triple no fue el objetivo final de la tesis, los experimentos realizados ofrecieron información valiosa para el desarrollo posterior del sistema de clasificación binaria.

Los mejores resultados en esta etapa se obtuvieron con el clasificador Naive Bayes. El entrenamiento fue aplicando los mismos principios de selección y reducción de características, validación cruzada y control de aleatoriedad utilizados posteriormente en la clasificación binaria. Las métricas obtenidas se resumen en la tabla 29, derivadas directamente de las matrices de confusión:

Naive Bayes			
Métrica	Clase 1	Clase 2	Clase 3
	(Sanos)	(R1)	(R2)
Sensibilidad (Recall)	91.7 %	71.4 %	57.9 %
Precisión	73.3 %	66.7 %	73.3 %
F1-score	81.5 %	68.9 %	64.7 %

Tabla 29. Resultados clasificación tres clases.

Estos resultados reflejan que la clasificación de pacientes sanos fue especialmente robusta. Sin embargo, la discriminación entre las clases R1 y R2 fue más compleja, probablemente debido a la superposición de características fisiológicas entre ambas condiciones, además de que al hacer esta separación, se contaba con muy pocos datos de entrenamiento y prueba, ya que cada clase quedó establecida con un total de 50 sujetos al buscar un balance. Esta limitación, combinada con la similitud entre ambas afecciones, sugiere que una mayor profundidad en el análisis de estas clases podría ser necesaria para mejorar su diferenciación en futuras investigaciones.

Dado que la clasificación entre tipos de afección diabética no era el objetivo principal de esta tesis, se optó por unificar las clases R0, R1 y R2 en una única categoría de "pacientes con pie diabético". Esta decisión facilitó el entrenamiento de clasificadores más robustos y adaptados al propósito final: detectar la presencia de riesgo o afectación diabética en los pies, independientemente del tipo clínico subyacente, resultados que fueron ilustrados a lo largo del trabajo.

4. CONCLUSIONES

El presente trabajo propuso el desarrollo de un sistema automático para la detección del pie diabético a partir de imágenes en el espectro visible e infrarrojo, sin requerir información antropométrica ni exámenes clínicos, con el fin de simplificar y automatizar el proceso de diagnóstico preliminar.

El método de fusión basado en DWT se destacó como el más eficaz, al obtener el mayor puntaje global. Su ventaja radica en su capacidad para realizar una descomposición multiresolución, separando las imágenes en bandas de frecuencia que capturan tanto los detalles finos (bordes, texturas) como las variaciones suaves.

Por su parte, la DCT también mostró resultados cercanos en algunas métricas, gracias a su eficiencia en la compactación de energía, lo cual es útil para representar patrones dominantes con pocos coeficientes. Sin embargo, su limitación principal es que opera globalmente sobre la imagen, lo que puede llevar a una pérdida de detalles locales sutiles que son clínicamente importantes en este caso.

Por otro lado, la TPL trabaja en múltiples escalas, como DWT, pero su enfoque está más orientado a la detección de bordes y contrastes locales, lo que, si bien puede ser útil en ciertos contextos visuales, podría no capturar adecuadamente la gradación térmica suave que caracteriza las anomalías en el pie diabético.

En síntesis, aunque DCT y TPL ofrecieron rendimientos aceptables y podrían ser optimizadas con parámetros específicos, la DWT presentó una mayor capacidad de representación tanto de estructura como de contenido térmico, lo que la convierte en la opción más adecuada para la etapa de fusión en este sistema.

Durante el proceso de segmentación se identificaron desafíos relevantes relacionados con la similitud cromática entre el color de piel y ciertos objetos del entorno, especialmente aquellos con tonos café, naranja o amarillos. Estos colores presentaron comportamientos similares en todos los espacios de color evaluados, dificultando el diseño de filtros lógicos consistentes. Por ejemplo, al analizar la diferencia entre las capas G (verde) y B (azul) en el espacio visible, se observó que, para colores como el naranja y el café, esta variación oscilaba entre 10 a 20 unidades, un rango insuficientemente discriminativo para segmentar de forma confiable sin incluir elementos no deseados del fondo.

A pesar de estas dificultades, la etapa de operaciones morfológicas (en particular, la combinación de erosión y dilatación) jugó un papel crucial en la eliminación de residuos y zonas ambiguas, permitiendo una mejora considerable en los resultados finales con un promedio global de similitud alcanzó un 90.01%, lo que demuestra un alto grado de concordancia y robustece la fiabilidad del proceso, aunque se reconoce que aún existen casos extremos donde la segmentación puede verse comprometida por interferencias cromáticas del entorno.

Para la clasificación entre pies sanos y pies con afecciones diabéticas, se implementaron diversos enfoques que incluyeron tanto SVM, clasificadores estadísticos, como redes neuronales. Todos ellos fueron entrenados con características extraídas a partir de imágenes previamente segmentadas y fusionadas, considerando tanto segmentaciones manuales como automáticas.

Si bien, las SVM superaron los resultados de los trabajos previos, presentaron un número significativamente mayor de rechazos. Este comportamiento, aunque útil en sistemas donde se permite o valora la indecisión frente a casos ambiguos, puede limitar su aplicabilidad en entornos clínicos o automáticos donde se requiere una respuesta clara e inmediata. Por esta razón, el desempeño más destacado fue obtenido por el clasificador Quadratic Discriminant Analysis (QDA), que alcanzó una precisión de 92,9 %, una sensibilidad de 86,7 % y un F1-score de 89,7 % acompañado de solo 7 rechazos o casos de indecisión. Además de sus buenos resultados, QDA demostró una notable consistencia frente a diferentes métodos de segmentación, lo cual refuerza su capacidad de generalización ante datos heterogéneos.

La etapa de clasificación fue objeto de un proceso iterativo de optimización, donde se exploró una amplia variedad de configuraciones. Por ejemplo, en el caso de las redes neuronales se probaron distintos algoritmos de entrenamiento, tasas de aprendizaje, funciones de error, tamaños de red, número de épocas, e incluso diferentes semillas aleatorias de inicialización de pesos para estudiar su impacto sobre la estabilidad y el rendimiento.

Un aspecto fundamental en la mejora de los resultados fue la selección de características. A través de técnicas estadísticas y el análisis visual de diagramas de bandas, se logró una reducción significativa del conjunto inicial de características, conservando únicamente aquellas con mayor capacidad discriminante. Esta reducción no solo mejoró la precisión de los clasificadores, sino que también incrementó la eficiencia computacional, lo cual es deseable en escenarios de implementación real.

La evaluación con imágenes segmentadas automáticamente evidenció un desempeño inferior al obtenido con segmentación manual, lo que resalta la necesidad de seguir perfeccionando los métodos automáticos de segmentación. Conservar la información relevante durante este proceso resulta fundamental, ya que su pérdida puede afectar significativamente la capacidad diagnóstica del sistema.

Finalmente, es importante señalar que estas mismas características seleccionadas también se utilizaron durante una fase exploratoria inicial del estudio, en la que se abordó una clasificación más específica en tres clases: sanos, R1 (neuropáticos no isquémicos) y R2 (isquémicos). A pesar de la complejidad añadida del problema multiclase, los clasificadores alcanzaron desempeños aceptables, con sensibilidades especialmente altas en la detección de pacientes sanos. Si bien la diferenciación entre R1 y R2 resultó más desafiante, los resultados obtenidos en esta etapa preliminar ofrecen una base prometedora y sugieren que, con una selección adecuada de características, es posible sostener métricas razonables.

5. TRABAJOS FUTUROS

Como parte del desarrollo futuro de esta línea de investigación, se proponen diversas mejoras y ampliaciones al sistema presentado, que permitirían incrementar su precisión, escalabilidad y aplicabilidad en entornos clínicos reales:

- Diseño e implementación de una interfaz gráfica para el usuario final, que permita cargar imágenes, visualizar resultados de segmentación y clasificación, y emitir reportes de forma intuitiva. Esto facilitaría el uso del sistema por parte de personal médico no especializado en programación o visión artificial.
- Desarrollo de un modelo de segmentación basado exclusivamente en redes neuronales profundas, entrenado desde cero y enfocado específicamente en la segmentación precisa de las plantas de los pies. Esta estrategia podría superar las limitaciones observadas con YOLO y los filtros en espacios de color, especialmente frente a fondos complejos o variaciones en las condiciones de captura.
- Ampliación de la base de datos utilizada, incorporando un mayor número de imágenes provenientes de distintas fuentes, sujetos y condiciones clínicas. Esto permitiría mejorar la capacidad de generalización de los clasificadores y redes neuronales, reducir el sesgo del sistema y aumentar su confiabilidad.
- Construcción o identificación de una base de datos que contenga imágenes térmicas en formato radiométrico junto con sus correspondientes imágenes RGB del mismo caso, con el fin de evaluar la detección del pie diabético mediante técnicas de fusión en un entorno más cercano al clínico real.
- Exploración de técnicas de fusión de imágenes, ya sea profundizando aún más las abordadas en este estudio o trabajar con métodos nuevos. Evaluar estas técnicas podría conducir a una representación más rica y discriminativa para el proceso de clasificación.
- Implementación de nuevos clasificadores, así como la construcción de comités de clasificadores, explorando diferentes configuraciones y técnicas de combinación (votación, ponderación, entre otros) que optimicen la precisión y reduzcan la tasa de rechazo, especialmente en casos de difícil discriminación.

6. REFERENCIAS

- [1] Vista de la detección de úlceras del pie diabético impulsada por aprendizaje automático con YOLOv5 (s.f.). [Online]. Disponible en: <https://healthinformaticsjournal.com/index.php/IJMI/article/view/816/773> (Consultado el 15 de Oct de 2024)
- [2] F. E. Godoy, *Métodos clásicos de clasificación: comparación y aplicación*, trabajo especial de licenciatura en Matemática, Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Argentina, jul. 2021.
- [3] A. Richaud, *Guía completa de K-Nearest Neighbors (Con ejemplos de código en Python)*.
- [4] International Diabetes Federation (s.f.). IDF Diabetes Atlas: Global Diabetes Data & Statistics. [Online]. Disponible en: <https://diabetesatlas.org/> (Consultado el 12 de sep de 2024)
- [5] Cuenta de Alto Costo (2023, 11, 14). *Día mundial de la diabetes 2023*. [Online]. Disponible en: <https://cuentadealtocosto.org/general/dia-mundial-de-la-diabetes-2023/> (Consultado el 12 de Sep de 2024)
- [6] Organización Panamericana de la Salud (s.f.). *Diabetes - OPS/OMS*. [Online]. Disponible en: <https://www.paho.org/es/temas/diabetes> (Consultado el 12 de sep de 2024)
- [7] Ministerio de Salud Pública y Bienestar Social (s.f.). *Cada 20 segundos, una persona con diabetes es amputada en el mundo*. [Online]. Disponible en: <https://www.mspbs.gov.py/portal-16631/cada-20-segundos-una-persona-con-diabetes-es-amputada-en-el-mundo.html> (Consultado el 9 de enero de 2025)
- [8] J. D. May y M. W. J. Morris (2017, 12, 01). *Mobile phone generated vibrations used to detect diabetic peripheral neuropathy*. *Foot and Ankle Surgery*, vol. 23, no. 4, pp. 281–284. [Online]. Disponible en: <https://doi.org/10.1016/j.fas.2016.08.004>
- [9] Gobierno de Canarias, Consejería de Sanidad. (2017). *Guía de actuación: Pie diabético en Canarias*. Dirección General de Programas Asistenciales. ISBN: 978-84-16878-08-6.
- [10] L. Muñoz, A. Uribe y D. Martínez (s.f.). *Conocimientos del pie diabético en el personal médico en formación de la Universidad de Antioquia*. [Online]. Disponible en: <http://www.scielo.org.co/pdf/iat/v35n1/0121-0793-iat-35-01-40.pdf> (Consultado el 4 de febrero de 2024)
- [11] Avila Cabrera, D. (2023). *Aplicación clínica y protocolo de medición de temperatura en la planta del pie en pacientes diabéticos con scanner IR desarrollado en el laboratorio LAREMUS* (Tesis de maestría). Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Zacatenco.
- [12] FLIR SYSTEMS, ‘Manual del Curso de Termografía Infrarroja Nivel 1’.
- [13] Instituto de Astrofísica de Canarias (s.f.). *IACTEC Tecnología Médica | Instituto de Astrofísica de Canarias – IAC*. [Online]. Disponible en: <https://www.iac.es/es/proyectos/iactec-tecnologia-medica> (Consultado el 5 de febrero de 2024)
- [14] P. Carlos and D. Ferrin Bolaños, “Sistema de Visión Artificial Multiespectral para Análisis de la Huella Plantar,” 2014.
- [15] M. Bharara (2007, 05, 01). *Liquid crystal thermography in neuropathic assessment of the diabetic foot*. Tesis doctoral, Bournemouth University.
- [16] M. A. Abdul Rahim (2018). *Detection of diabetic foot with infrared thermography*. Tesis de maestría, Nanyang Technological University, Singapur. [Online]. Disponible en: <https://doi.org/10.32657/10356/74240> (Consultado el 27 de mayo de 2025)

- [17] Vilcahuamán, L. (2013). *Early diagnosis of diabetic foot using thermal images* (Tesis doctoral, Université d'Orléans, Laboratoire PRISME). Université d'Orléans. [14]
- Hernández Contreras, D. A. (2019). *Determinación temprana de riesgo de ulceración en el pie diabético mediante el análisis de imágenes térmicas* (Tesis doctoral). Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Tonantzintla, Puebla, México.
- [18] M. C. D. A. Hernández Contreras (2019). *Determinación temprana de riesgo de ulceración en el pie diabético mediante el análisis de imágenes térmicas*. Tesis doctoral, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), México.
- [19] N. Arteaga-Marrero, A. Hernández-Guedes, J. Ortega-Rodríguez, and J. Ruiz-Alzola, "State-of-the-Art Features for Early-Stage Detection of Diabetic Foot Ulcers Based on Thermograms," *Biomedicines*, vol. 11, no. 12, Dec. 2023, doi: 10.3390/biomedicines11123209.
- [20] IEEE DataPort (s.f.). *Base de datos de termogramas plantares para el estudio de las complicaciones del pie diabético*. [Online]. Disponible en: <https://ieee-dataport.org/open-access/plantar-thermogram-database-study-diabetic-foot-complications> (Consultado el 7 de febrero de 2025)
- [21] D. Bouallal et al. (2022, 06, 01). *STANDUP database of plantar foot thermal and RGB images for early ulcer detection*. *Open Research Europe*, vol. 2, p. 77. [Online]. Disponible en: <https://doi.org/10.12688/openreseurope.14706.1>
- [22] Fernández Manchado, C. (s.f.). *Anatomía del pie y ciclos de la marcha humana* [Anexo I]. Escuela Universitaria de Ingeniería Técnica Industrial de Zaragoza.
- [23] B. Paula, G. Martín y C. Moreno (s.f.). *26 Aprendiendo sobre patología del pie: Manual didáctico y autoevaluación*. Colección manuales UEX-126.
- [24] M. Duraes, G. Captier, and F. Bonnel, "Vascularización arterial y venosa del pie," *EMC - Podología*, vol. 24, no. 1, pp. 1–9, Feb. 2022, doi: 10.1016/S1762-827X(21)45963-8.
- [25] M. Duraes, G. Captier y F. Bonnel (2022, 02). *Vascularización arterial y venosa del pie*. *EMC - Podología*, vol. 24, no. 1, pp. 1–9. [Online]. Disponible en: [https://doi.org/10.1016/S1762-827X\(21\)45963-8](https://doi.org/10.1016/S1762-827X(21)45963-8)
- [26] *Imagen glucosa* (s.f.). [Online]. Disponible en: <https://www.niddk.nih.gov/health-information/informacion-de-la-salud/diabetes/informacion-general/que-es> (Consultado el 5 de abril de 2025)
- [27] C. Rodríguez, C. Nieto, P. Sedano, D. Correa y E. Garcés (2017). *CLASO CLÍNICO Colgajo plantar medial sensitivo para cubrimiento de defectos en talón. A propósito de un caso*. *Sensitive plantar medial flap for covering defects in heel. About a case. RCCP*, vol. 23.
- [28] *Pie diabético | Lima | Endocrinología - INM* (s.f.). [Online]. Disponible en: <https://www.endocrinologia-inm.com/pie-diabetico> (Consultado el 5 de ene de 2025)
- [29] *Neuropatía e isquemia* (s.f.). [Online]. Disponible en: <https://scielo.isciii.es/pdf/angiologia/v75n4/0003-3170-angiologia-75-04-242.pdf> (Consultado el 5 de abril de 2025)
- [30] *Comparación de la respuesta de sensibilidad con el monofilamento de Semmes Weinstein con el método de la velocidad de conducción sensitiva para detectar neuropatía diabética subclínica en un grupo de pacientes del Hospital de la Policía de enero del 2000 a mayo del 2000* (s.f.). [Online]. Disponible en: <https://cybertesis.unmsm.edu.pe/backend/api/core/bitstreams/e02c7b85-46aa-407e-9083-ea712c82ab4e/content> (Consultado el 5 de abril de 2025)
- [31] *Prueba de monofilamentos: MedlinePlus enciclopedia médica ilustración* (s.f.). [Online]. Disponible en: https://medlineplus.gov/spanish/ency/esp_imagepages/19960.htm (Consultado el 5 de abril de 2025)

- [32] E. Talaya-Navarro et al. (2022). *Prevención de amputaciones relacionadas con el pie diabético*. *Journal of Negative and No Positive Results*, vol. 7, no. 2, pp. 235–265. [Online]. Disponible en: <https://doi.org/10.19230/JONNPR.4450>
- [33] Dros J, Wewerinke A, Bindels PJ, van Weert HC. Accuracy of monofilament testing to diagnose peripheral neuropathy: a systematic review. *Ann Fam Med*. 2009 Nov-Dec;7(6):555–8. doi:10.1370/afm.1016.
- [34] Bril V, Kojic J, Ngo M, Clark K. Comparison of a neurothesiometer and vibratron in measuring vibration perception thresholds and relationship to nerve conduction studies. *Diabetes Care*. 1997 Sep;20(9):1360–2. doi:10.2337/diacare.20.9.1360.
- [35] A. V. Orozco, A. María y T. Ocampo (s.f.). *Visión artificial, una alternativa a la sensorica convencional en la implementación de los sistemas de control*. *Artificial vision, an alternative to the conventional sensor in the implementation of control systems*
- [36] H. Loaiza, “Material de Clase Visión Artificial,” Universidad del Valle, Cali, 2024.
- [37] *Vista de PREPROCESAMIENTO DE DATOS ESTRUCTURADOS* (s.f.). [Online]. Disponible en: <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/4123/5790> (Consultado el 1 de agos de 2024)
- [38] N. Palomino y U. Román (2009). *Técnicas de Segmentación en Procesamiento Digital de Imágenes*.
- [39] J. Armando, R. Frutos, I. M. Andrés y D. Oliva (2023). *Estimation of Distribution Algorithm for Image Segmentation by Multilevel Thresholding*. *Research in Computing Science*.
- [40] Y. Binbin (2019). *An improved infrared image processing method based on adaptive threshold denoising*. *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, pp. 1–12. [Online]. Disponible en: <https://doi.org/10.1186/S13640-018-0401-8/FIGURES/5>
- [41] *HSV color space and RGB color transformation* (s.f.). [Online]. Disponible en: https://www.researchgate.net/figure/HSV-color-space-and-RGB-color-transformation_fig4_312678134 (Consultado el 15 de marzo de 2025)
- [42] F. G. Ortiz Zamora (2002). *Fundamentos del Color*. Universidad de Alicante. [Archivo PDF]. Disponible en: <https://rua.ua.es/dspace/bitstream/>
- [43] J. A. Marcial Basilio, G. A. Torres, G. Sánchez Pérez, K. Toscano Medina y H. M. Pérez Meana (2012). *Novedosa técnica para la detección de imágenes pornográficas empleando modelos de color HSV y YCbCr*. *Novel method for pornographic image detection using HSV and YCbCr color models*.
- [44] W. A. Ceballos, L. E. Salazar y A. I. Oviedo. *Clasificador de Páginas Web Pornográficas Basado en el Contenido de las Imágenes*.
- [45] Reinke A, Eisenmann M, Tizabi MD, Sudre CH, Rädtsch T, Antonelli M et al. Common Limitations of Image Processing Metrics: A Picture Story. *arXiv [Internet]*. 2021 Apr 12. Disponible en: <http://arxiv.org/abs/2104.05642>
- [46] X. Jin et al. (2017). *A survey of infrared and visual image fusion methods*. *Infrared Physics & Technology*, vol. 85, pp. 478–501, Sep. [Online]. Disponible en: <https://doi.org/10.1016/J.INFRARED.2017.07.010>
- [47] A. Orellana y R. Reinier (2020). *Análisis y selección de técnicas para la fusión de imágenes PET/CT basado en software*. Habana, pp. 0–14.
- [48] M. F. Lillo-Saavedra y C. Gonzalo (2008). *Multidirection-Multiresolution Fusion Images Methology (MDMR) applied to turbidity lake estimation*. *Información Tecnológica*, vol. 19, no. 5, pp. 137–146. [Online]. Disponible en: <https://doi.org/10.1612/inf.tecnol.3978it.07>
- [49] A. Marín (2013). *Implementación y evaluación de algoritmos de fusión de imagen en el contexto de la imagen médica*.
- [50] L. Mendoza, *Procesamiento de Datos Discretos en 1D y 2D, Usando Fourier, Coseno y Wavelet*, vol. 1. 2020.

- [51] W. Wang y F. Chang (2011). *A multi-focus image fusion method based on Laplacian pyramid*. *J Comput (Taipei)*, vol. 6, no. 12, pp. 2559–2566. [Online]. Disponible en: <https://doi.org/10.4304/jcp.6.12.2559-2566>
- [52] P. J. Burt y R. J. Kolczynski, “Enhanced image capture through fusion,” *4th International Conference on Computer Vision*, Berlin, Germany, pp. 173–182, 1993, doi: 10.1109/ICCV.1993.378222.
- [53] F. Calderon, A. Garnica-Carrillo, and J. J. Flores, “Fusión de Imágenes Multi Foco basado en la Combinación Lineal de Imágenes utilizando Imágenes Incrementales,” *RIAI - Revista Iberoamericana de Automatica e Informatica Industrial*, vol. 13, no. 4, pp. 450–461, Oct. 2016, doi: 10.1016/j.riai.2016.07.002.
- [54] A. Orellana, R. Rodríguez, D. Yanez, and P. Valdés-Sosa, “Fusion of PET / CT neuroimaging using a Wavelet-based and the Haar discrete transform scheme,” *RIAI - Revista Iberoamericana de Automatica e Informatica Industrial*, vol. 18, no. 1, pp. 75–81, 2021, doi: 10.4995/RIAI.2020.12977.
- [55] J. Medina, L. Joyanes, C. Pinilla, and M. Madrid, “Aplicativo Web para la Fusión de Imágenes Satelitales”, doi: 10.4304/risti.11.17-30.
- [56] H. Loaiza, “Material de clases Reconocimiento de patrones”. Universidad del Valle, Cali, 2024.
- [57] J. Andrea, M. García, y I. A. Uribe, “Técnicas para detección de outliers multivariantes,” *Revista en telecomunicaciones e informática*, vol. 3, no. 5, pp. 11–25, 2013.
- [58] “Clasificadores, extracción y selección.” Accedido: 16 mayo. 2024. [En línea]. Disponible en: https://oa.upm.es/50271/1/TFG_CRISTINA_BLANCO_GARRIDO.pdf
- [59] C. A. Cattaneo *et al.*, “Métodos de umbralización de imágenes digitales basados en entropía de Shannon y otros,” 2011. [En línea]. Disponible en: <http://www.amcaonline.org.ar>
- [60] *Elementos básicos de estadística descriptiva para el análisis de datos*, Universidad Católica Luis Amigó, 2016.
- [61] A. Bereciartua Pérez, *Desarrollo de algoritmos de procesamiento de imagen avanzado para interpretación de imágenes médicas. Aplicación a segmentación de hígado sobre imágenes de Resonancia Magnética multisequencia*, tesis doctoral, Universidad del País Vasco (UPV/EHU), 2016.
- [62] E. D. Monroy Jordan, J. E. Pérez Neira y E. Gómez Vásquez, *Máquinas de soporte vectorial (SVM)*, tesis de grado, Programa de Ingeniería Electrónica, Universidad Tecnológica de Bolívar, Cartagena de Indias, 2005.
- [63] J. C. Platt, “Probabilistic outputs for support vector machines and comparisons,” en *Advances in Large Margin Classifiers*, A. J. Smola, P. L. Bartlett, B. Schölkopf, y D. Schuurmans, Eds. Cambridge, MA: MIT Press, 2000, pp. 61–74.
- [64] A. Carlos, A. Jaimes, *Optimización del clasificador ‘Naive Bayes’ usando árbol de decisión C4.5*, tesis de maestría, Facultad de Ciencias Matemáticas, Unidad de Posgrado, Universidad Nacional Mayor de San Marcos, Perú
- [65] C. Alberto Ruiz, M. S. Basualdo y D. Jorge Matich, *Cátedra: Informática Aplicada a la Ingeniería de Procesos - Orientación I Redes Neuronales: Conceptos Básicos y Aplicaciones*.
- [66] “Qué son las redes neuronales y sus funciones - ATRIA Innovation,” [En línea]. Disponible en: <https://atriainnovation.com/blog/que-son-las-redes-neuronales-y-sus-funciones/>. [Accedido: 16-abr-2024].
- [67] J. Francisco, C. García, A. Por, I. Enrique y R. Carballo, “Fundamentos para la implementación de red neuronal perceptrón multicapa mediante software.”

- [68] L. R. Barba Guamán, *Uso de técnicas deep learning para reconocimiento de objetos en áreas rurales*, tesis doctoral, Escuela Técnica Superior de Ingeniería de Sistemas Informáticos, Universidad Politécnica de Madrid, 2021.
- [69] D. Saúl y R. Raneros, *Estudio de la arquitectura YOLO para la detección de objetos mediante deep learning*, Escuela Técnica Superior de Ingeniería de Telecomunicación, Universidad de Valladolid.
- [70] J. Redmon, S. Divvala, R. Girshick y A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” [En línea]. Disponible en: <http://pjreddie.com/yolo/>. [Accedido: 17-abr-2025].
- [71] S. Afaq y S. Rao, “Significance of epochs on training a neural network,” [En línea]. Disponible en: www.ijstr.org. [Accedido: 17-abr-2025].
- [72] S. J. Pan y Q. Yang, “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009, doi: 10.1109/TKDE.2009.191.
- [73] A. B. Muñoz Ruiz, “Los sistemas automatizados de reconocimiento de emociones en el trabajo en el reglamento europeo de inteligencia artificial,” *Labores Sociales*, 2024, doi: 10.20318/labos.2024.9033.
- [74] A. B. Muñoz Ruiz, “Los sistemas automatizados de reconocimiento de emociones en el trabajo en el reglamento europeo de inteligencia artificial,” *Labores Sociales*, 2024, doi: 10.20318/labos.2024.9033.
- [75] “Roboflow: Computer vision tools for developers and enterprises,” [En línea]. Disponible en: <https://roboflow.com/>. [Accedido: 12-nov-2024].
- [76] J. Redmon, S. Divvala, R. Girshick y A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. Disponible en: <http://pjreddie.com/yolo/>.
- [77] “COCO - Common Objects in Context,” [En línea]. Disponible en: <https://cocodataset.org/#home>. [Accedido: 12-nov-2024].
- [78] Benjamin Planche, and, and Eliot Andres, *Hands-On Computer Vision with TensorFlow 2*.
- [79] D. Rodríguez Esparragón, A. Medina y F. M. Ruiz, “Introducción a la fusión de imágenes y sus métricas de calidad,” Departamento de Señales y Comunicaciones, Universidad de Las Palmas de Gran Canaria, España.
- [80] E. Fidalgo, E. Alegre, L. Fernández-Robles, and V. González-Castro, “Early fusion of multi-level saliency descriptors for image classification,” *RIAI - Revista Iberoamericana de Automatica e Informatica Industrial*, vol. 16, no. 3, pp. 358–368, 2019, doi: 10.4995/RIAI.2019.10640.
- [81] M. A. Rojas Carrasco, *Fusión de imágenes visibles y térmicas aplicando técnicas metaheurísticas para sistemas de reconocimiento de rostros*, informe final para optar al título de Ingeniero Electrónico, Pontificia Universidad Católica de Valparaíso, Valparaíso.
- [82] B. Yang, S. Li, and F. Sun, “Image fusion using nonsubsampling contourlet transform,” in *Proceedings of the 4th International Conference on Image and Graphics, ICIG 2007*, 2007, pp. 719–724. doi: 10.1109/ICIG.2007.124.
- [83] H. S. De la Cruz Mondragón y D. F. Soler Franco, *Fusión de imágenes utilizando la transformada wavelet para la evaluación y análisis multitemporal del cauce del río Meta en la zona de Puerto López, Colombia*, proyecto de grado para optar al título de Ingeniero Catastral y Geodesta, Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas.
- [84] Leal de Araújo A, Diana da Silva Negreiros F, Sampaio Florêncio R, Karanini Paz de Oliveira S, Vilarouca da Silva AR, Magalhães Moreira TM. How to cite this article. Rev Latino-Am Enfermagem. 2022;30:e3525. doi:10.1590/1518-8345.5663.3525.

- [85] Rutkove SB. AAEM Minimonograph #14: effects of temperature on neuromuscular electrophysiology. *Muscle and Nerve*. 2001;24(7):867–882. doi: 10.1002/mus.1084.
- [86] Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press; 2016. Disponible en: <http://www.deeplearningbook.org>
- [87] Charoenphakdee N, Lee J, Sugiyama M. Classification with rejection based on cost-sensitive classification. In: *Proceedings of the 38th International Conference on Machine Learning*. PMLR; 2021.
- [88] Mozannar H, Sontag D. *Consistent estimators for learning to defer to an expert*. In: *International Conference on Machine Learning*. PMLR; 2020.
- [89] Pes B, Lai G. Cost-sensitive learning strategies for high-dimensional and imbalanced data: a comparative study. *PeerJ Comput Sci*. 2021;7:e832. doi:10.7717/peerj-cs.832.
- [90] Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River (NJ): Prentice Hall; 1999.
- [91] Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer; 2006.
- [92] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform*. 2002 Oct;35(5-6):352–9. doi:10.1016/S1532-0464(03)00034-0.