

Estudio Exploratorio De Funciones Kernel En La Clasificación De Energía Fotovoltaica Con
ANN y SVM Mediante Imágenes Satelitales

Ximena Tatiana Diaz Rosero y Diana Carolina López Tulcan

Programa de Ingeniería de Sistemas, Facultad de Ingeniería de Sistemas, Universidad
CESMAG

Nota del autor

El presente trabajo de Grado tiene como propósito cumplir el requisito exigido para optar al título de pregrado como Ingenieras de Sistemas en la universidad CESMAG.

La correspondencia referente a este trabajo debe dirigirse al Programa de Ingeniería de la Universidad CESMAG. Correo electrónico: decanatura.ingenieria@unicesmag.edu.co.

Estudio Exploratorio De Funciones Kernel En La Clasificación De Energía Fotovoltaica Con
ANN y SVM Mediante Imágenes Satelitales

Ximena Tatiana Diaz Rosero y Diana Carolina López Tulcan

Programa de Ingeniería de Sistemas, Facultad de Ingeniería de Sistemas, Universidad
CESMAG

Asesor: Mg. Héctor Andrés Mora Paz

11 de agosto de 2026

NOTA DE ACEPTACION

El presente trabajo de grado titulado: ESTUDIO EXPLORATORIO DE FUNCIONES KERNEL EN LA CLASIFICACIÓN DE ENERGÍA FOTOVOLTAICA CON ANN Y SVM MEDIANTE IMÁGENES SATELITALES.

Presentado por Ximena Tatiana Diaz Rosero, Diana Carolina López Tulcán, ha sido revisado y aprobado para optar al título de Ingeniero de sistemas. Este trabajo cumple con los requisitos académicos y metodológicos exigidos por la Institución.

Firma del jurado 1:

Nombre: Mg. Dalila Mercedes Pachajoa

Firma del jurado 2:

Nombre: Mg. Joan Carlos Ayala Benavides

Asesor:

Nombre: Mg. Héctor Andrés Mora Paz

Fecha:

Ciudad: Pasto, Nariño

NOTA DE EXCLUSIÓN DE RESPONSABILIDAD

Las opiniones y resultados expresados en este trabajo de grado son de exclusiva responsabilidad de las autoras y no comprometen a la Universidad CESMAG ni a su asesor.

DEDICATORIA

La culminación de este trabajo de grado representa mucho más que el cumplimiento de un requisito académico; simboliza el resultado de años de esfuerzo, perseverancia, aprendizaje y crecimiento tanto personal como profesional. Cada página de este documento refleja las experiencias, los desafíos superados y el compromiso asumido para alcanzar una meta que hoy se convierte en una realidad.

Dedico este logro, en primer lugar, a mi familia, quienes han sido el pilar fundamental de mi vida. Gracias por su amor incondicional, por su paciencia, comprensión y sacrificio, y por acompañarme en cada etapa de este camino con palabras de aliento y una confianza inquebrantable en mis capacidades. Su apoyo constante ha sido la fuerza que me impulsó a seguir adelante en los momentos de incertidumbre y dificultad. Este logro es también de ustedes, porque sin su respaldo, sus enseñanzas y los valores que me han inculcado, este sueño no habría sido posible.

A mis amigos, quienes hicieron parte de este recorrido brindándome su compañía, motivación y apoyo incondicional. Gracias por escucharme, animarme y compartir conmigo tanto los momentos de satisfacción como aquellos de mayor exigencia. Su amistad sincera fue un estímulo permanente para continuar con entusiasmo y determinación.

Expreso mi más profundo agradecimiento a mis docentes, quienes, con su dedicación, profesionalismo y compromiso con la formación académica, compartieron sus conocimientos, experiencias y orientaciones. Sus enseñanzas no solo enriquecieron este trabajo, sino que también contribuyeron significativamente a mi desarrollo profesional y humano. Cada consejo, observación y enseñanza recibida dejó una huella invaluable en este proceso.

Asimismo, extiendo mi gratitud a todas aquellas personas, conocidos, compañeros y colaboradores que, de una u otra manera, hicieron parte de este camino. Gracias por cada palabra de ánimo, por su disposición para ayudar, por compartir sus conocimientos y por creer en mí cuando más lo necesité. Cada aporte, por pequeño que pareciera, contribuyó significativamente a la consecución de esta meta.

Finalmente, dedico este trabajo a todas las personas que han sido parte de mi proceso de formación y crecimiento. Este logro no representa únicamente el esfuerzo individual, sino el resultado del apoyo, la confianza y el acompañamiento de quienes caminaron a mi lado durante esta importante etapa de mi vida. Espero que este trabajo refleje el compromiso, la dedicación y los valores que cada uno de ustedes ayudó a fortalecer en mí.

Con profunda gratitud y respeto, ofrezco este logro a todos quienes hicieron posible que este sueño académico se convirtiera en una realidad.

AGRADECIMIENTOS

Expresamos nuestro más sincero agradecimiento a Dios por brindarnos la vida, la salud y la fortaleza necesarias para culminar este proceso académico. A nuestras familias, por su apoyo incondicional, paciencia y comprensión durante toda nuestra formación profesional, siendo el pilar fundamental en cada etapa de este camino. A nuestros amigos, por su acompañamiento, motivación y palabras de aliento, que hicieron más llevadero este proceso. A nuestros docentes, por su dedicación, orientación y transmisión de conocimientos, los cuales fueron esenciales para el desarrollo de este trabajo de grado y nuestra formación como profesionales. Finalmente, agradecemos a nuestra institución por brindarnos las herramientas y oportunidades necesarias para nuestro crecimiento académico y personal.

INDICE

Contenido

INTRODUCCIÓN.....	12
I. PROBLEMA DE INVESTIGACIÓN.....	13
A. Objeto o tema de investigación.....	13
B. Línea de investigación	13
C. Sublínea de investigación.....	14
D. Planteamiento del problema.....	14
E. Formulación del problema	15
F. Objetivos	15
1. Objetivo general:	15
2. Objetivos específicos:	16
G. Justificación.....	16
H. Viabilidad	17
1) Viabilidad operativa:	17
2) Viabilidad técnica:	18
3) Viabilidad económica:	18
I. Delimitación	18
II. MARCO TEÓRICO	19
A. Antecedentes	19
1) Antecedentes internacionales:	19
2) Antecedentes Nacionales	20
3) Antecedentes Regionales	21
B. Supuestos teóricos de la investigación	22
1) Energías alternativas:	22
2) Algoritmos de Aprendizaje Automático:	24
3) Algoritmos de Aprendizaje Profundo:	26
C. Variables de estudio	28
D. Definición nominal de las variables:.....	29
E. Definición operativa de las variables	31
F. Formulación de hipótesis	34
III. METODOLOGÍA.....	34
A. Paradigma	34
B. Enfoque.....	34
C. Método	35
D. Tipo De Investigación	35
E. Diseño De Investigación	35

F.	<i>Población.....</i>	36
G.	<i>Muestra.....</i>	36
H.	<i>Técnicas De Recolección De La Información.....</i>	37
I.	<i>Validez de las técnicas de recolección de la información.....</i>	37
J.	<i>Confiabilidad de las técnicas de recolección.....</i>	37
K.	<i>Instrumentos de recolección de información.....</i>	37
IV.	RECURSOS DE LA INVESTIGACIÓN	37
A.	<i>Talento humano.....</i>	38
B.	<i>Recursos físicos.....</i>	38
C.	<i>Presupuesto.....</i>	38
D.	<i>Financiación.....</i>	39
V.	RESULTADOS DE LA INVESTIGACIÓN.....	40
A.	<i>Identificación de funciones kernel.....</i>	40
1)	<i>Revisión bibliográfica de funciones kernel:</i>	40
2)	<i>Selección de funciones kernel para la experimentación:</i>	43
B.	<i>Implementación de modelos SVM y ANN.....</i>	46
1)	<i>Recolección y preprocesamiento de datos satelitales:.....</i>	46
2)	<i>Entrenamiento y validación de modelos:.....</i>	50
3)	<i>Evaluación del desempeño de las funciones:.....</i>	53
4)	<i>Resultados comparativos en modelos SVM:.....</i>	54
5)	<i>Resultados comparativos en modelos ANN:.....</i>	55
6)	<i>Optimización de hiperparámetros y configuraciones experimentales</i>	57
7)	<i>Métricas completas de clasificación:.....</i>	58
8)	<i>Análisis estadístico de significancia:.....</i>	60
C.	<i>Comparación general entre ANN y SVM, discusión y aplicabilidad del modelo.....</i>	61
1)	<i>Comparación entre ANN y SVM:</i>	61
2)	<i>Interpretación de resultados:</i>	63
3)	<i>Interpretación general de los resultados:.....</i>	64
4)	<i>Aplicabilidad del modelo:.....</i>	65
5)	<i>Limitaciones.....</i>	66
6)	<i>Interpretación geoespacial</i>	68
D.	<i>Discusión científica</i>	71
E.	<i>Fortalecimiento metodológico.....</i>	74
VI.	ANÁLISIS DE RESULTADOS	76
A.	<i>Análisis del comportamiento de las funciones kernel.....</i>	77
B.	<i>Comparación entre modelos SVM y ANN.....</i>	77
C.	<i>Análisis de métricas de evaluación.....</i>	78
D.	<i>Impacto de la discretización de datos.....</i>	78

<i>E. Análisis del modelo óptimo</i>	<i>79</i>
<i>F. Discusión general.....</i>	<i>79</i>
CONCLUSIONES	80
RECOMENDACIONES	82
TRABAJOS FUTUROS	84
LIMITACIONES	85
BIBLIOGRAFÍA.....	86

ÍNDICE DE TABLAS

Tabla I Definición formal funciones kernel.....	28
Tabla II Variables dependientes de estudio	28
Tabla III Variables independientes de estudio.....	29
Tabla IV Plan De Acción	36
Tabla V Talento Humano	38
Tabla VI Recursos Físicos.....	38
Tabla VII Costos Directos	38
Tabla VIII Costos Indirectos	39
Tabla IX Costos Del Proyecto	39
Tabla X Estudios Bibliográficos Considerados	43
Tabla XI Funciones Seleccionadas.....	44
Tabla XII Funciones Kernel	45
Tabla XIII Características De Los Sensores Satelitales	46
Tabla XIV Variables Espectrales Utilizadas.....	47
Tabla XV Métodos de discretización	48
Tabla XVI División Del Dataset	49
Tabla XVII Desempeño de los modelos SVM	51
Tabla XVIII Desempeño de los modelos ANN	52
Tabla XIX Comparación final entre modelos	52
Tabla XX Métricas de Evaluación.....	53
Tabla XXI Resultados Comparativos SVM.....	55
Tabla XXII Resultados Comparativos ANN	56
Tabla XXIII Métricas Completas De Los Cinco Mejores Modelos	59
Tabla XXIV Comparación general entre los modelos ANN y SVM	62
Tabla XXV Interpretación Geoespacial De Los Modelos	69
Tabla XXVI Comparación de los modelos evaluados	73
Tabla XXVII Fortalecimiento Metodológico Implementado.....	76
Tabla XXVIII Comparación general entre ANN y SVM	77

ÍNDICE DE IMÁGENES

Fig. 1 Panel solar percibiendo la radiación	22
Fig. 2 Arquitectura de una red neuronal	25
Fig. 3 Con margen de separación no optimo. (b). Con margen de separación óptimo	25
Fig. 4 Truco kernel de dos dimensiones a tres	27
Fig. 5 Distribución generada mediante los métodos de discretización K-means, Quantil y Uniform.	48
Fig. 6 Grafico de barras de distribución de clases generadas mediante los métodos de discretización K-means, Quantil y Uniform.....	49
Fig. 7 Comparación de métricas de rendimiento (Accuracy y F1-Score) y tiempo computacional (en segundos) para los distintos kernels utilizados en el clasificador SVM.	54
Fig. 8 Evaluación comparativa de métricas de rendimiento (Accuracy, Recall y F1-Score) y tiempo computacional de la Red Neuronal Artificial (ANN) con diferentes kernels.	56
Fig. 9 Mapa de clasificación de irradiancia utilizando Cosine-SVM.....	70
Fig. 10 Mapa de clasificación de irradiancia utilizando Canberra-ANN.	70
Fig. 11 Comparación espacial entre kernels tradicionales y alternativos.	71
Fig. 12 Distribución geográfica de clases de irradiancia	71
Fig. 13 Comparación de accuracy entre kernels tradicionales y alternativos.	73
Fig. 14 Visualización en Google Earth del mapa de irradiancia predicha, superpuesto sobre imágenes satelitales y modelos de elevación del terreno de Colombia y países vecinos.....	74

ÍNDICE DE ECUACIONES

Ecuación 1	Definición formal de una función kernel.....	27
Ecuación 2	Exactitud del modelo	31
Ecuación 3	Precisión	32
Ecuación 4	F1-score	32
Ecuación 5	Sensibilidad	32
Ecuación 6	Índice de eficiencia.....	54

INTRODUCCIÓN

El crecimiento constante de la demanda energética a nivel global, junto con la necesidad de mitigar los efectos del cambio climático, ha impulsado el desarrollo e implementación de fuentes de energía renovable. En este contexto, la energía solar se posiciona como una alternativa viable debido a su disponibilidad, sostenibilidad y capacidad para reducir emisiones de CO₂.

En paralelo, los modelos de aprendizaje automático, como las Máquinas de Soporte Vectorial (SVM) y las Redes Neuronales Artificiales (ANN), han demostrado un alto potencial en tareas de predicción complejas, como la estimación del potencial energético solar. Su capacidad para modelar relaciones no lineales se potencia mediante el uso de funciones kernel, las cuales permiten transformar los datos a espacios de mayor dimensionalidad, facilitando procesos de clasificación más precisos.

Tradicionalmente, se han empleado funciones kernel como la lineal, polinómica y de base radial (RBF). Sin embargo, existe un campo de investigación poco explorado centrado en la evaluación de funciones kernel alternativas, cuyo uso podría optimizar el rendimiento de los modelos en contextos específicos. Por otra parte, los altos costos y limitaciones logísticas para desplegar redes meteorológicas en regiones remotas han impulsado el uso de imágenes satelitales multiespectrales, como las provenientes de los sensores Landsat y MODIS, como fuente de información para la estimación indirecta de la radiación solar.

En este contexto, el presente estudio aborda la clasificación del potencial fotovoltaico en el municipio de San Juan de Pasto, mediante el diseño, implementación y evaluación de modelos SVM y ANN utilizando imágenes satelitales. Se experimentó con diversas funciones kernel, incluyendo alternativas no convencionales, con el propósito de identificar configuraciones óptimas que logren un equilibrio entre desempeño y costo computacional. Los resultados obtenidos contribuyen al fortalecimiento de herramientas para la planificación energética sostenible, especialmente en regiones con limitada infraestructura de medición.

I. PROBLEMA DE INVESTIGACIÓN

A. Objeto o tema de investigación

El presente proyecto tiene como propósito explorar el uso de funciones kernel alternativas para mejorar la precisión en la clasificación de zonas con potencial fotovoltaico. Las funciones kernel desempeñan un papel fundamental en algoritmos de aprendizaje automático como las Máquinas de Soporte Vectorial (SVM) y, en ciertos enfoques, en Redes Neuronales Artificiales (ANN), ya que permiten transformar datos complejos en espacios de mayor dimensionalidad donde la separación entre clases resulta más eficiente.

Sin embargo, en aplicaciones relacionadas con la predicción de energía solar, la mayoría de los estudios se limita al uso de kernels tradicionales, los cuales no siempre logran capturar adecuadamente la complejidad inherente a los datos ambientales.

En este contexto, la investigación propone analizar el comportamiento de funciones kernel menos convencionales, como la truncada, triangular y Canberra, así como posibles combinaciones bajo el enfoque de Multiple Kernel Learning (MKL), aplicadas a modelos SVM y ANN en la región de Nariño, Colombia. Para ello, se emplearán imágenes satelitales provenientes de sensores Landsat y MODIS, con el objetivo de desarrollar modelos más precisos y eficientes que contribuyan a la identificación de zonas con alto potencial solar.

B. Línea de investigación

La propuesta se desarrolla en el marco de la línea de investigación en AIO TIC (Artificial Intelligence and Other Technologies in Information and Communication), reconocida dentro del programa de Ingeniería de Sistemas de la Universidad CESMAG y articulada al Grupo de Investigación Tecnofilia. Esta línea promueve el estudio y la aplicación de tecnologías que simulan procesos cognitivos, como el aprendizaje, la percepción y la toma de decisiones, mediante soluciones computacionales innovadoras.

C. Sublínea de investigación

El enfoque principal del estudio se alinea con la Sublínea de Inteligencia Artificial específicamente en el ámbito del aprendizaje automático y Múltiple Kernel Learning, un campo perteneciente a la línea de Inteligencia Artificial del programa de Ingeniería de Sistemas de la Universidad CESMAG. Esta Sublínea permite combinar diversas funciones kernel para mejorar el rendimiento de modelos de aprendizaje automático. Su aplicación en esta investigación tiene como propósito optimizar los procesos de clasificación orientados al análisis del potencial solar fotovoltaico a partir de datos obtenidos por sensores satelitales.

D. Planteamiento del problema

En medio del actual proceso de transición energética y del creciente interés por fuentes limpias y renovables, la energía solar representa una alternativa estratégica para muchas regiones del mundo; no obstante, en países en vías de desarrollo como Colombia y particularmente en zonas geográficamente favorables como el departamento de Nariño se evidencia una limitada disponibilidad de estudios técnicos que orienten de forma precisa la selección óptima de sitios para la instalación de paneles solares. Esta carencia obedece en gran medida a la insuficiente integración de análisis geoespaciales y climáticos mediante herramientas avanzadas de modelado computacional.

A pesar de los avances recientes en tecnologías de teledetección y la creciente disponibilidad de imágenes satelitales como las proporcionadas por Landsat y MODIS, aún no se han implementado de forma efectiva modelos robustos de aprendizaje automático que integren adecuadamente esta información con técnicas de clasificación orientadas a estimar el potencial energético solar. Esta situación refleja una brecha importante en la investigación, especialmente en la capacidad de vincular datos espectrales y temporales con la radiación solar, lo que limita significativamente la toma de decisiones informadas en proyectos de infraestructura energética sostenible. [1]

De manera específica, se observa una notable ausencia de investigaciones centradas en la evaluación comparativa de distintas funciones kernel dentro de modelos de clasificación como las Máquinas de Soporte Vectorial (SVM) y las Redes Neuronales Artificiales (ANN). Las funciones kernel cumplen un papel fundamental al permitir que datos no lineales sean transformados a espacios de mayor dimensión, facilitando una mejor capacidad predictiva. La elección adecuada del kernel impacta directamente en la precisión y la capacidad de generalización del modelo [2].

Sin embargo, la mayoría de los estudios siguen utilizando funciones tradicionales como los

kernels lineales, polinómicos o de base radial (RBF) sin explorar alternativas que podrían ofrecer mejores resultados, lo que evidencia una falta de innovación metodológica en la investigación actual. [3]

Además de esta limitación, la escasez de estudios locales que integren técnicas de aprendizaje profundo o enfoques como el Multiple Kernel Learning (MKL) ha generado una dilación en la aplicación de la inteligencia artificial al desarrollo de soluciones energéticas limpias. Esta situación es especialmente crítica en regiones rurales o de difícil acceso, donde la cobertura eléctrica es limitada. La falta de bases de datos solares confiables a nivel local, sumada a la escasa combinación de métodos que integren de manera sinérgica ANN y SVM, ha reducido la precisión de los modelos predictivos. Como consecuencia, se continúa dependiendo de enfoques desarrollados en contextos ajenos que no se adaptan del todo a las condiciones ambientales y sociales propias de la región. [4]

Como consecuencia de estas limitaciones técnicas y metodológicas, muchas regiones del país no cuentan con criterios claros para determinar cuáles son los lugares más apropiados para instalar sistemas de energía solar. Esto ha llevado a que se tomen decisiones mal fundamentadas, generando inversiones que no siempre

son rentables o sostenibles. En zonas rurales, donde el acceso a la electricidad es precaria o inexistente, esta situación agrava las desigualdades sociales, pues se pierde la oportunidad de implementar soluciones energéticas limpias que podrían mejorar significativamente la calidad de vida. A esto se suma la falta de datos locales confiables sobre radiación solar. [1] Todo esto demuestra la necesidad urgente de avanzar en investigaciones que combinen datos geográficos locales con herramientas modernas como el aprendizaje automático, para tomar mejores decisiones y promover una transición energética más justa y efectiva.

E. Formulación del problema

¿Cuáles funciones kernel logran un mejor compromiso entre desempeño en clasificación y costo computacional en la identificación de zonas con potencial fotovoltaico a través de ANN y SVM en imágenes satelitales?

F. Objetivos

1. Objetivo general:

Medir el balance entre desempeño y costo computacional de diferentes funciones kernel en la

clasificación de zonas con potencial fotovoltaico a través de imágenes satelitales y modelos de aprendizaje automático (ANN y SVM).

2. *Objetivos específicos:*

- a.) Implementar funciones kernel alternativas recomendadas por expertos para el acoplamiento en los algoritmos de ML de clasificación SVM y ANN a través de sus definiciones formales y su correspondiente interfaz de acoplamiento
- b.) Acoplar funciones kernel alternativas en los algoritmos de aprendizaje automático de SVM y ANN para la evaluación del desempeño en la clasificación de la irradiancia a través de imágenes satelitales multiespectrales.
- c.) Encontrar las mejores configuraciones de hiperparámetros, escalador, algoritmo de aprendizaje automático, dataset (Landsat y MODIS) y función kernel para la realización de un análisis comparativo ecuánime a través de una sintonización aleatoria.

G. *Justificación*

La transición hacia fuentes de energía más limpias es una prioridad a nivel global y Colombia se encuentra en una posición estratégica para sumarse a este esfuerzo; en particular, el departamento de Nariño gracias a su diversidad geográfica y sus condiciones climáticas favorables presenta un alto potencial para el aprovechamiento de la energía solar. Para identificar de manera efectiva las zonas más adecuadas para el desarrollo de proyectos fotovoltaicos, [1] es fundamental contar con herramientas tecnológicas avanzadas que integren imágenes satelitales y modelos de aprendizaje automático. Este proyecto responde a esa necesidad, proponiendo una solución innovadora que combina datos satelitales con algoritmos de clasificación como las Máquinas de Soporte Vectorial (SVM) y las Redes Neuronales Artificiales (ANN), optimizados mediante el uso de funciones kernel. [4]

Aunque ya existen investigaciones que aplican SVM y ANN en la clasificación de recursos energéticos, la exploración sistemática de diversas funciones kernel en este contexto sigue siendo escasa. Esta propuesta se diferencia al introducir una evaluación comparativa de funciones kernel menos tradicionales, como la truncada, triangular y Canberra, para la clasificación de zonas con potencial solar a partir de imágenes satelitales. [2] Esta metodología no solo aporta una visión

innovadora, sino que también puede mejorar significativamente la precisión y eficiencia de los modelos de clasificación, contribuyendo al desarrollo de soluciones energéticas más confiables y sostenibles

El desarrollo de investigaciones orientadas a soluciones energéticas sostenibles ha sido limitado, especialmente en contextos locales. Por ello, avanzar en el estudio y aplicación de nuevas funciones kernel permite ampliar el horizonte de análisis y generar soluciones más adaptadas a nuestras condiciones geográficas. Los resultados de esta investigación pueden servir como base para futuras investigaciones en el área de recursos energéticos, así como para optimizar algoritmos de aprendizaje automático mediante una adecuada selección de funciones kernel.[5] Asimismo, se espera que estos avances contribuyan a una toma de decisiones más informada y eficiente en la planificación y ejecución de proyectos solares, tanto a nivel regional como nacional

H. Viabilidad

El presente estudio es viable desde los puntos de vista técnico, operativo y económico. Existen condiciones que respaldan su ejecución, tales como el acceso a recursos tecnológicos adecuados, la existencia de equipos de investigación consolidados en el área de Inteligencia Artificial, Multiple Kernel Learning, y la disponibilidad de herramientas y datos de libre acceso. A continuación, se detallan los aspectos específicos.

1) Viabilidad operativa:

Desde el punto de vista operativo, el proyecto dispone de una infraestructura tecnológica adecuada proporcionada por la institución, así como con el respaldo de plataformas colaborativas que facilitan la gestión de datos y el trabajo de equipo. Además, se cuenta con acceso autorizado y necesarios para utilizar bases de datos satelitales públicas, como Landsat y MODIS, cumpliendo con las políticas de uso de datos abiertos. Además, investigaciones previas, como la desarrollada por Belanche [3], han evidenciado que la adecuada selección de funciones kernel mejora significativamente el rendimiento de modelos de clasificación, lo cual respalda el enfoque metodológico adoptado.

2) *Viabilidad técnica:*

Desde el punto de vista técnico, el proyecto se fundamenta en el conocimiento y experiencia de un equipo de trabajo especializado en aprendizaje automático, redes neuronales y técnicas de acoplamiento de funciones kernel. Este grupo pertenece a una línea de investigación consolidada en Inteligencia Artificial y Multiple Kernel Learning. Asimismo, se cuenta con acceso a bibliotecas de código abierto ampliamente utilizadas en la comunidad científica como Scikit-learn y TensorFlow, lo que permite la implementación de modelos predictivos de manera eficiente y sin restricciones técnicas. El proyecto está estructurado por etapas claramente definidas, lo cual garantiza un desarrollo ordenado y controlado.

3) *Viabilidad económica:*

El proyecto es económicamente viable ya que se apoya principalmente en el uso de recursos gratuitos. Las imágenes satelitales utilizadas (provenientes de los sensores Landsat y MODIS) están disponibles de forma libre y abierta, al igual que las herramientas tecnológicas empleadas, como Google Colab, Python, QGIS y GitHub, lo cual evita incurrir en gastos por licencias comerciales. La Universidad también contribuye con un soporte importante al brindar acceso a bibliografía, orientación docente y equipos de cómputo básicos.

En cuanto a los gastos menores que puedan surgir durante el desarrollo del proyecto como impresión, transporte o conectividad, estos serán asumidos directamente por las autoras de la investigación, quienes se comprometen con la financiación personal de los aspectos operativos y técnicos necesarios para su ejecución. Por tratarse de una propuesta de carácter exploratorio, no se requiere inversión en hardware especializado ni adquisición de sensores físicos, lo que garantiza que el proyecto se mantenga dentro de los márgenes económicos asumibles por las autoras.

I. Delimitación

La presente investigación se enfoca en la revisión y evaluación de funciones kernel aplicadas en modelos de aprendizaje automático, específicamente Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Artificiales (ANN), con el objetivo de clasificar zonas con alto potencial fotovoltaico en el departamento de Nariño, Colombia. La delimitación espacial de la investigación se circunscribe al municipio de San Juan de Pasto, seleccionado por su relevancia en términos de demanda energética y disponibilidad de datos satelitales. Metodológicamente, la investigación se enfoca en el análisis comparativo de funciones kernel convencionales y no convencionales, prestando

especial atención a aquellas menos exploradas en el ámbito energético. Estas funciones serán evaluadas en función de su precisión, capacidad de generalización y adecuación a las características espaciales del área de estudio.

La información utilizada proviene de bases de datos proporcionadas por la Universidad de Nariño, que incluyen variables climáticas, geográficas y radiométricas obtenidas mediante sensores remotos (MODIS y Landsat), complementadas con datos de validación en campo, recolectados por la institución. El desarrollo del estudio está programado para realizarse a lo largo del año 2025 y primer periodo del año 2026, siguiendo una metodología estructurada que contempla la limpieza de datos, construcción de modelos, validación de resultados y análisis comparativo.

En consecuencia, esta investigación se encuentra delimitada a un contexto geográfico específico, un marco temporal definido y un conjunto seleccionado de funciones kernel, lo que permite un enfoque detallado, controlado y susceptible de replicación.

II. MARCO TEÓRICO

Esta sección aborda los principales conceptos y teorías que fundamentan el presente estudio, inicialmente se llevó a cabo una revisión de fuentes bibliográficas de los últimos 5 años relacionadas con las energías alternativas, con el fin de comprender los conceptos clave y conocer las principales técnicas utilizadas para analizar datos energéticos. Esta revisión permite identificar enfoques relevantes y antecedentes que sirven como base teórica del estudio. En segundo lugar, se consultaron fuentes bibliográficas sobre algoritmos de aprendizaje automático, con énfasis en dos modelos ampliamente utilizados: las Máquinas de Soporte Vectorial (SVM) y las Redes Neuronales Artificiales (ANN). Se busca entender cómo funcionan, en qué casos se aplican y qué potencial tienen para predecir comportamientos relacionados con la energía solar.

A. *Antecedentes*

Se llevó a cabo una revisión de fuentes bibliográficas relacionadas con la predicción de energía fotovoltaica donde se encontraron artículos a nivel internacional, nacional y regional.

1) *Antecedentes internacionales*

En una investigación realizada por [1] se analizó cómo diferentes funciones kernel influyen en el desempeño del algoritmo SVM al clasificar zonas con distintos niveles de irradiancia, a partir de imágenes satelitales multiespectrales. El estudio utilizó datos del satélite Landsat y comparó varias funciones kernel, entre ellas la radial (RBF), la polinómica y la sigmoide. Los datos mostraron que la

función RBF logró los más altos niveles de precisión al señalar zonas para hacer energía del sol. Estos resultados evidencian lo clave de elegir bien la función kernel para mejorar los modelos que clasifican cosas, aplicados a las energías que se renuevan. Este antecedente demuestra la influencia directa de la selección del kernel en el rendimiento del modelo, lo que justifica el enfoque comparativo del presente estudio.

Por otro lado, en estudio de [6], se propuso un enfoque automatizado para detectar granjas fotovoltaicas a partir de imágenes satelitales, empleando redes neuronales convolucionales (CNN). Aunque el objetivo principal del trabajo fue identificar instalaciones solares ya construidas, los resultados evidencian el gran potencial que tienen las técnicas de aprendizaje profundo para analizar imágenes satelitales en contextos relacionados con la energía solar. Esta investigación refuerza el valor de las CNN como herramienta eficaz para tareas de reconocimiento y clasificación en el ámbito de las energías renovables. Aunque no enfocado en predicción de irradiancia, este estudio valida el uso de aprendizaje profundo en imágenes satelitales, respaldando el uso de ANN en este trabajo.

Según [7], presentaron una estrategia innovadora para clasificar el uso del suelo a partir de imágenes de teledetección, utilizando el enfoque de aprendizaje con múltiples kernels (MKL). Lo interesante de su propuesta es que permite combinar automáticamente distintos conjuntos de características sin necesidad de conocimientos previos o reglas heurísticas sobre la tarea específica. Además, demostraron que MKL puede obtener buenos resultados incluso cuando se dispone de pocos datos para entrenar el modelo. Las pruebas realizadas con bases de datos públicas confirmaron la efectividad y el potencial de esta metodología en aplicaciones de teledetección. Este aporte destaca la versatilidad y eficiencia del enfoque MKL, lo que plantea su posible inclusión como alternativa en estudios futuros del presente trabajo.

2) *Antecedentes Nacionales*

En Colombia, se llevó a cabo un estudio que evaluó diversas funciones kernel aplicadas al algoritmo SVM, con el propósito de clasificar zonas según su nivel de irradiancia solar utilizando imágenes satelitales. [1] El trabajo se enfocó en identificar configuraciones óptimas para mejorar el rendimiento de estos algoritmos y generar mapas zonificados por clases de irradiancia. Los resultados subrayan la relevancia de elegir adecuadamente la función kernel, ya que esta decisión influye directamente en la precisión de la clasificación de áreas con potencial para generación fotovoltaica. Este antecedente valida que investigaciones similares han sido útiles en el contexto colombiano, reforzando la pertinencia territorial del presente trabajo.

Según en una investigación centrada en el departamento de La Guajira [8] se analizaron datos obtenidos por sensores remotos para evaluar la radiación solar en términos de intensidad, duración y frecuencia. Esta información permitió estimar la capacidad de generación eléctrica mediante paneles solares, resaltando el papel clave de las tecnologías de teledetección en la planificación y desarrollo de proyectos de energía solar en el país. Este trabajo resalta el valor de los datos satelitales en el análisis energético, soporte clave para el componente de imágenes multispectrales de esta investigación

Otro estudio colombiano se enfocó en comparar distintas funciones kernel dentro de modelos de redes neuronales artificiales (ANN) y máquinas de vectores de soporte (SVM), [4] con el objetivo de predecir la oferta energética a partir de imágenes satelitales. A través de un marco experimental para la sintonización de hiperparámetros, los investigadores encontraron que la selección adecuada del kernel puede mejorar significativamente la capacidad predictiva de los modelos, optimizando así la estimación del potencial fotovoltaico. Este estudio es altamente pertinente, pues coincide directamente con los objetivos del presente trabajo, pero no aborda una evaluación sistemática entre funciones kernel, lo cual representa un vacío que esta investigación busca llenar.

3) *Antecedentes Regionales*

Este estudio analiza la imagen satelital de MODIS durante 11 años y 15 años Landsat para crear mapas de radiación solar y mapas claros para Nariño [9] . Se aplicaron modelos de radiación solar y modelos de construcción de línea de tiempos se utilizaron para centrarse en la detección de modelos y el análisis de la radiación solar y nubosidad en el departamento de Nariño. Este estudio regional proporciona una base empírica para la disponibilidad y aplicabilidad de datos satelitales en el área geográfica del presente trabajo.

Además del Instituto de Planificación y Promoción de Soluciones Energéticas (IPSE), Zonas no interconectadas (ZNI), que priorizan la implementación de soluciones fotoeléctricas para satisfacer las necesidades energéticas insatisfechas en las áreas rurales [10] priorizan. Estos estudios han sido una energía significativamente sostenible para las principales políticas nacionales. Estos datos respaldan la relevancia social y política de aplicar modelos de predicción fotovoltaica en Nariño, lo que le otorga viabilidad práctica a la investigación.

Además, otro estudio, como que se llevó a cabo un estudio que evaluó diversas funciones kernel aplicadas al algoritmo SVM, con el propósito de clasificar zonas según su nivel de irradiancia solar utilizando imágenes satelitales [4], han introducido para mejorar el rendimiento de estos

algoritmos y generar mapas zonificados por clases de irradiancia. Los resultados han dado la relevancia de elegir adecuadamente, los sistemas solares fotoeléctricos en áreas rurales, aprobación de estas soluciones en el entorno de viabilidad técnica. Este trabajo evidencia experiencias previas en la región, pero carece de una comparación extensiva entre kernels y de un enfoque conjunto con ANN, lo que deja espacio para el estudio exploratorio propuesto.

B. Supuestos teóricos de la investigación

Esta investigación parte de varios supuestos teóricos fundamentales. En primer lugar, se considera que la elección de funciones kernel adecuadas tiene un impacto significativo en la precisión, generalización y eficiencia computacional de modelos de aprendizaje automático como SVM y ANN. En segundo lugar, se presume que los datos satelitales multiespectrales, de forma adecuada procesados y combinados con técnicas de clasificación, son capaces de capturar con precisión la variabilidad espacial de la irradiancia solar en zonas geográficas específicas.

1) Energías alternativas:

Las energías alternativas hacen referencia a aquellas fuentes empleadas para generar electricidad o calor que no proviene de combustibles fósiles como el carbón o el petróleo, ni de métodos convencionales como las hidroeléctricas. Estas fuentes tienen un impacto ambiental reducido en comparación con las tradicionales; a pesar del avance de la tecnología en el ámbito industrial, su presencia en el mercado energético global sigue siendo limitada [11] . A continuación, se describen algunas de las formas más representativas de energía alternativa.



Fig. 1 Panel solar percibiendo la radiación

a) Energía solar

Esta es una fuente fundamental de origen natural que está vinculada con fenómenos físicos, químicos y biológicos, siendo fundamental para los procesos vitales en el planeta [12]. Las fluctuaciones de la radiación solar influyen en el comportamiento climático y atmosférico de la tierra a lo largo del tiempo. Esta forma de energía puede aprovecharse mediante diferentes tecnologías, siendo las más representativas la solar fotovoltaica y la solar térmica o termosolar. La energía solar fotovoltaica transforma la luz solar en electricidad a través de celdas fotovoltaicas, comúnmente integradas en paneles solares. Por su parte, la energía termosolar emplea espejos o lentes para concentrar la radiación solar, generando calor que se utiliza para producir vapor y accionar turbinas eléctricas. Ambas tecnologías ofrecen alternativas sostenibles y escalables para reducir la dependencia de fuentes fósiles y avanzar hacia sistemas energéticos más limpios y eficientes

b) Energía eólica.

Este tipo de energía se genera a partir del movimiento del viento, el cual puede transformarse en electricidad mediante un generador eléctrico. Se clasifica como una fuente renovable y limpia [13], ya que no produce emisiones contaminantes y su recurso primario, el viento, es natural, constante y virtualmente inagotable, lo cual asegura su disponibilidad a largo plazo.

c) Irradiancia solar.

La irradiancia proveniente del Sol constituye la fuente principal de energía para el sistema terrestre. Su fluctuación impacta los componentes tanto neutros como ionizados de la atmósfera. En particular, las variaciones en ciertas franjas del espectro han sido objeto de estudio por su influencia en la dinámica y composición química de la atmósfera terrestre [14]. Esta energía solar, que atraviesa diferentes capas del Sol, incide tanto en la atmósfera superior como en los patrones climáticos globales.

d) Reflectancia.

Corresponde a la proporción de radiación solar que es reflejada por una superficie. El Índice de Reflectancia Solar (SRI) permite evaluar la capacidad de un material para reflejar el calor proveniente del sol. De este modo, los materiales con un alto SRI tienden a calentarse menos cuando son expuestos a la radiación solar directa [15].

e) Imágenes satelitales multiespectrales.

Son representaciones captadas desde sensores remotos instalados en satélites, que registran la

radiación reflejada por la superficie terrestre en diferentes bandas del espectro electromagnético, como el visible, el infrarrojo cercano y el infrarrojo medio. Cada banda espectral capta información específica del terreno, como la vegetación, la humedad del suelo, la temperatura o la cobertura nubosa, permitiendo analizar fenómenos ambientales con un alto nivel de detalle [16]. Estas imágenes se utilizan ampliamente en áreas como la agricultura, la gestión del agua, el monitoreo ambiental y la estimación de recursos energéticos, como la radiación solar. Satélites como Landsat y MODIS proporcionan este tipo de datos de manera gratuita, lo que facilita su uso en investigaciones científicas y aplicaciones tecnológicas en contextos locales y globales.

2) *Algoritmos de Aprendizaje Automático:*

Los algoritmos de Machine Learning son un conjunto de métodos computacionales que permiten a las máquinas aprender patrones y relaciones a partir de datos, sin necesidad de ser programadas explícitamente para cada tarea. A través del análisis de grandes volúmenes de información, estos algoritmos pueden realizar predicciones, clasificaciones o decisiones automáticas, adaptándose progresivamente a nuevos datos. Existen diferentes tipos de algoritmos según el objetivo de aprendizaje, siendo los más comunes los enfoques supervisados y semi-supervisados.

a) *Aprendizaje supervisado:*

Este paradigma se basa en la generación de modelos predictivos que se entrenan con datos etiquetados. A través de procesos iterativos, los algoritmos identifican relaciones entre variables de entrada y salida, permitiendo realizar predicciones precisas sobre nuevos datos [17].

b) *Aprendizaje no supervisado:*

Utiliza datos no etiquetados y busca identificar estructuras o patrones ocultos en los datos, agrupándolos según su similitud. Al no contar con salidas definidas, estos algoritmos no pueden aplicarse directamente a tareas de clasificación o regresión tradicionales [18].

c) *Aprendizaje semi-supervisados:*

Este enfoque combina un reducido conjunto de datos etiquetados con una gran cantidad de datos no etiquetados. El objetivo es aprovechar la estructura subyacente en los datos sin etiquetas para mejorar el rendimiento predictivo de los modelos [19].

d) *Red neuronal artificial (ANN):*

Se compone de unidades denominadas neuronas artificiales, organizadas en capas (entrada,

ocultas y salida). Las neuronas de entrada procesan los datos iniciales, mientras que las capas ocultas transforman esa información progresivamente hasta generar un resultado final en la capa de salida. El término "aprendizaje profundo" surge al incrementar el número de capas ocultas, lo que permite modelar fenómenos complejos [20].

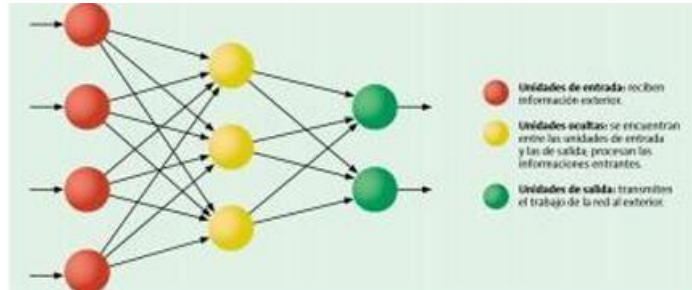


Fig. 2 Arquitectura de una red neuronal

e) *Máquinas de Soporte Vectorial (SVM):*

Este conjunto de algoritmos supervisados busca hallar el hiperplano que maximiza la separación entre clases en el espacio de características. Su formulación matemática considera tanto espacios lineales como no lineales, adaptándose a distintos tipos de datos [21].

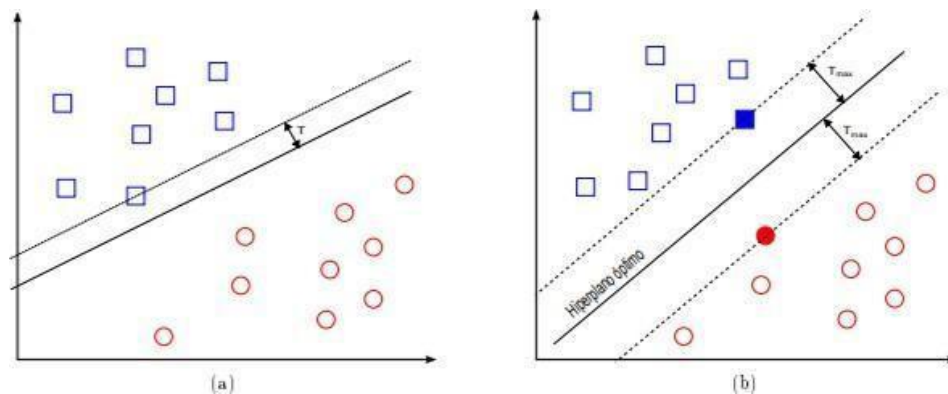


Fig. 3 Con margen de separación no óptimo. (b). Con margen de separación óptimo

f) *Árboles de decisión:*

Son técnicas estadísticas que permiten construir modelos predictivos mediante una estructura jerárquica de decisiones. Están compuestos por nodos internos (criterios de decisión), ramas (condiciones) y nodos terminales (resultados). Se utilizan ampliamente en análisis de grandes volúmenes de datos [22].

g) Redes bayesianas:

Representan modelos probabilísticos mediante grafos dirigidos acíclicos. Cada nodo corresponde a una variable aleatoria y las conexiones indican relaciones de dependencia probabilística. Son útiles para resolver problemas bajo incertidumbre [23].

3) Algoritmos de Aprendizaje Profundo:

a) Deep Learning:

Es una técnica de aprendizaje automático que se basa en la representación jerárquica de datos a través de redes neuronales profundas. Las capas iniciales aprenden representaciones simples, mientras que las capas superiores combinan dicha información para construir descripciones más complejas [24].

b) Neuronales convolucionales (CNN):

Estas arquitecturas, diseñadas principalmente para el procesamiento de imágenes, contienen capas convolucionales que extraen características locales. Cada capa sucesiva aprende representaciones de mayor nivel de abstracción, desde bordes simples hasta formas complejas. [25]

c) Autoencoders:

Son modelos neuronales utilizados para tareas como la segmentación de imágenes. Su arquitectura encoder-decoder permite reconstruir la imagen de entrada a través de un "cuello de botella", minimizando la pérdida a nivel de píxel durante el entrenamiento [26].

d) Redes Generativas Adversarias (GAN):

Técnica moderna que emplea dos redes: un generador (G), que produce imágenes sintéticas, y un discriminador (D), que intenta distinguir entre imágenes reales y generadas. Ambos modelos se entrenan conjuntamente, mejorando mutuamente su rendimiento [27].

e) Funciones Kernel:

Una función kernel permite a algoritmos de aprendizaje automático, como SVM y ANN, transformar un conjunto de datos desde un espacio de N dimensiones a otro de mayor dimensión M, utilizando una medida de similitud entre pares de elementos. Este nuevo espacio M-dimensional se denomina espacio de Hilbert. [28].

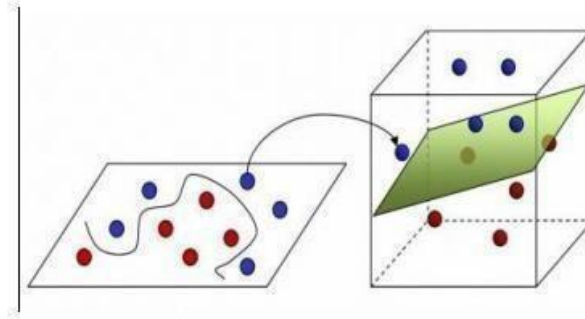


Fig. 4 Truco kernel de dos dimensiones a tres

En la Fig. 4, se puede observar cómo trabajan las funciones kernel, llevando una distribución de datos de dos dimensiones a tres dimensiones, a este funcionamiento generalmente se lo llama truco kernel. Este truco permite reducir la complejidad de una función que separe las clases de una distribución de datos.

Una función kernel es una función matemática que permite proyectar los datos $\phi: X \rightarrow \mathcal{H}$ desde un espacio de características original hacia un espacio de mayor dimensión facilitando la separación de datos no lineales X (Variables independientes) en un espacio $\mathcal{H}$ de Hilbert (llamado espacio de características). Este proceso se realiza implícitamente mediante el denominado “truco kernel”, evitando el cálculo explícito en espacios de alta dimensionalidad como se ve en la Ecuación.[50]

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}, x, x' \in X$$

Ecuación 1 Definición formal de una función kernel

Donde $\langle \dots \rangle_{\mathcal{H}}$ denota al producto punto en $\mathcal{H}$ [3]. Existen diversos kernel empleados comúnmente en bibliotecas de aprendizaje automático como lineal, RBF, polinomial y tangente hiperbólico cuyas definiciones se pueden expresar.

Tabla I Definición formal funciones kernel

FUNCIÓN KERNEL	ECUACIÓN	CONDICIÓN
Lineal	$k(x, x') = \langle x, x' \rangle$	$x, x' \in \mathbb{R}$
RBF	$k(x, x') = \exp(-\sum_{i=1}^d \lambda_i (x_i - x'_i)^2)$	$\lambda_i > 0, \beta \in (0, 2]$
Polinomial	$k(x, x') = (\langle x, x' \rangle + 1)^m$	$m \in \mathbb{N}, a > 0$
Tangente hiperbólica	$k(x, x') = \tanh(a \langle x, x' \rangle + b)$	$a > 0, b < 0$

Aunque las funciones de la Tabla I. son funciones de uso común existen muchas más funciones kernel. Estas funciones son utilizadas en algoritmos supervisados para regresión y clasificación; y no supervisadas para detección de anomalías, análisis clúster y extracción de características. Dentro de los algoritmos más destacados se encuentran Procesos gaussianos, Spectral clustering, Kernel Linear Discriminant Analysis, Kernel Principal Components Analysis, Kernel Canonical Correlation Analysis, Kernel Independent Component Analysis, SVM, ANN y muchos más [28]

C. Variables de estudio

Tabla II Variables dependientes de estudio

NOMBRE DE LA VARIABLE	DEFINICIÓN	TIPO DE VARIABLE
Exactitud del modelo	Proporción de predicciones correctas sobre el total de muestras evaluadas, utilizada como medida global	Dependientes
Integral Precisión	Media ponderada de la precisión en distintas clases o umbrales, útil en problemas multiclase o desbalanceados.	Dependientes
F1-Score	Media armónica entre precisión y recall, adecuada para evaluar modelos en contextos con clases desbalanceadas	Dependientes
Recall	Proporción de verdaderos positivos identificados correctamente sobre el total real de positivos.	Dependientes

Tabla III Variables independientes de estudio

NOMBRE DE LA VARIABLE	DEFINICIÓN	TIPO DE VARIABLE
Algoritmo (ANN, SVM)	Tipo de modelo de aprendizaje automático utilizado para realizar la clasificación (Red Neuronal Artificial o Máquinas de Vectores de Soporte).	Independientes
Hiperparámetros ANN: Kernel, Arquitectura,	Parámetros que definen la estructura de la red neuronal (número de capas, neuronas) y la función kernel si aplica en alguna variante.	Independientes
Hiperparámetros SVM: Kernel, Constante de regularización	Parámetros clave del modelo SVM, incluyendo la función kernel utilizada y la constante que controla el margen de separación.	Independientes
Hiperparámetros Kernel: gamma, coeficiente	Parámetros específicos de funciones kernel como RBF o polinómica; 'gamma' controla la influencia de un solo punto de entrenamiento y el coeficiente afecta la forma del kernel.	Independientes

D. Definición nominal de las variables:

A continuación, se presentan las definiciones nominales y operacionales de las variables consideradas en este estudio, con el fin de garantizar claridad conceptual y coherencia metodológica en la clasificación del potencial fotovoltaico mediante modelos de aprendizaje automático.

1) Exactitud del modelo

La exactitud mide el porcentaje total de predicciones correctas realizadas por un modelo de clasificación, comparadas con el número total de casos evaluados.[29]

2) Integral precisión

Mide la proporción de verdaderos positivos entre todas las predicciones positivas, que se realizadas por el modelo, se utiliza cuando el coste de un falso positivo es alto [30].

3) F1-score

Es la medida armónica entre la predicción y el recall, sirve como una medida balanceada cuando se necesita considerar tanto los falsos positivos como los falsos negativos. [31].

4) Recall

Evalúa la capacidad del modelo para identificar correctamente todas las instancias positivas. Es esencial cuando el coste de un falso negativo es alto [32].

5) *Energía fotovoltaica*

Se refiere a una fuente de energía renovable obtenida a partir de la transformación directa de la radiación solar en electricidad, proceso que se realiza mediante dispositivos conocidos como paneles solares. Esta forma de energía limpia y sostenible es fundamental en la transición hacia un modelo energético más amigable con el medio ambiente [33].

6) *Clasificación de energía fotovoltaica*

Es el procedimiento mediante el cual se categorizan regiones geográficas de acuerdo con su idoneidad para la generación de energía solar, en función de condiciones ambientales, climáticas y geográficas [34].

7) *Redes neuronales artificiales (ANN)*

Son modelos de inteligencia artificial inspirados en la estructura y funcionamiento del cerebro humano, capaces de identificar y aprender relaciones complejas entre datos mediante un proceso de entrenamiento iterativo [35].

8) *Máquinas de vectores de soporte (SVM)*

Algoritmo de aprendizaje supervisado que se utiliza para clasificación y regresión, basado en la construcción de un hiperplano óptimo que maximiza la separación entre clases dentro de un espacio de características [36].

9) *Funciones kernel*

Son funciones matemáticas utilizadas en algoritmos como SVM para transformar los datos originales en un espacio de características de mayor dimensión, permitiendo así la clasificación de datos no linealmente separables [3].

10) *Imágenes satelitales*

Representan información capturada por sensores remotos instalados en satélites, que miden la reflectancia de la superficie terrestre en diferentes bandas del espectro electromagnético, permitiendo el análisis ambiental a gran escala [37].

11) *Zonas con potencial fotovoltaico*

Corresponden a áreas geográficas que presentan condiciones ambientales, de radiación solar y de cobertura del suelo favorables para la instalación y operación eficiente de sistemas solares fotovoltaicos [38].

12) *Aprendizaje automático*

Rama de la inteligencia artificial que se enfoca en el desarrollo de algoritmos capaces de aprender de los datos, identificar patrones y realizar predicciones sin necesidad de programación

explícita [39].

13) *Clasificación supervisada*

Técnica de aprendizaje automático en la que un modelo se entrena con datos que ya han sido etiquetados, permitiéndole aprender a asociar características de entrada con categorías de salida conocidas [40].

E. Definición operativa de las variables

1) *Exactitud del modelo:*

Las variables dependientes se evalúan una vez que el modelo ha sido entrenado. A partir de ahí, una de las métricas más utilizadas es la exactitud (accuracy), que indica qué tan bien está funcionando el modelo. En términos simples, muestra el porcentaje de predicciones correctas comparado con el total de casos analizados. Se calcula dividiendo la cantidad de aciertos del modelo, es decir, las veces que clasificó correctamente tanto los casos positivos como los negativos entre el número total de predicciones realizadas [41] donde:

$$Exactitud = \frac{\{TP + TN\}}{\{TP + TN + FP + FN\}}$$

Ecuación 2 Exactitud del modelo

TP (True Positivos): verdaderos positivos

TN (True Negativos): Verdaderos negativos

FP (False Positivos): Falsos Positivos

FN (False Negativos): Falsos Negativos

2) *Integral precisión:*

La precisión mide qué tan acertadas son las predicciones positivas del modelo. En otras palabras, indica cuántas de las veces que el modelo dijo que algo era positivo, realmente lo era. Se calcula dividiendo el número de verdaderos positivos entre todas las predicciones que fueron clasificadas como positivas. Esta métrica es especialmente importante en situaciones donde tener una gran cantidad de falsos positivos tiene un mayor coste [41]

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

Ecuación 3 Precisión

3) *F1-score:*

F1 score es una medida de evaluación utilizada en problemas de clasificación que combina la Precisión y el Recall en una sola métrica; es una medida que varía entre 0 y 1, donde 1 representa un F1-score perfecto y 0 indica un rendimiento nulo [41].

$$F1 - Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

Ecuación 4 F1-score

4) *Recall:*

La sensibilidad (recall) indica la proporción de verdaderos positivos (TP). Es la proporción entre los casos positivos bien clasificados por el modelo, respecto al total de positivos [41].

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

Ecuación 5 Sensibilidad

5) *Energía fotovoltaica:*

Se representa mediante valores de irradiancia solar (expresados en W/m²), extraídos de imágenes satelitales de sensores como Landsat y MODIS. Dichos datos son preprocesados, normalizados y utilizados como variable de entrada para alimentar modelos de clasificación del potencial solar [42].

6) *Clasificación de energía fotovoltaica:*

Se realiza asignando etiquetas (por ejemplo, alta, media o baja aptitud) a cada zona o píxel, con base en el análisis de sus características espectrales, a través de modelos predictivos de aprendizaje supervisado como ANN y SVM [43].

7) *Redes neuronales artificiales (ANN):*

Se implementan mediante herramientas como Keras o TensorFlow, configurando arquitecturas específicas que incluyen capas ocultas, funciones de activación y algoritmos de optimización. Su desempeño se evalúa con métricas como exactitud, precisión y F1-score [44].

8) *Máquinas de soporte vectorial (SVM)*

Se entrena con diferentes funciones kernel y parámetros ajustados (como C y gamma), y se evalúa mediante métricas como la exactitud, precisión y matriz de confusión, permitiendo comparar su rendimiento frente a otros modelos [45].

9) *Funciones kernel:*

En este estudio se evaluarán funciones kernel tradicionales (lineal, RBF, polinómica, sigmoide) y funciones personalizadas. Se medirá su impacto en la capacidad de clasificación del modelo, comparando resultados con diferentes configuraciones [46].

10) *Imágenes satelitales:*

Se utilizarán imágenes de los sensores Landsat y MODIS en formato raster (GeoTIFF). Estas imágenes serán preprocesadas para la extracción de variables relevantes, como estimación de radiación solar [47].

11) *Zonas con potencial fotovoltaico:*

Se representan como píxeles clasificados mediante modelos ANN y SVM, los cuales son etiquetados con distintos niveles de aptitud solar y georreferenciados para facilitar su análisis y visualización espacial en sistemas de información geográfica [48].

12) *Aprendizaje automático:*

Se aplica para entrenar modelos predictivos utilizando imágenes satelitales etiquetadas. Se emplearán bibliotecas como Scikit-learn, TensorFlow y PyTorch para desarrollar y validar los modelos ANN y SVM propuestos en el estudio [49].

13) *Clasificación supervisada:*

Se realiza el entrenamiento y validación de modelos ANN y SVM utilizando datos satelitales con etiquetas previamente asignadas sobre el nivel de potencial fotovoltaico, para evaluar su capacidad de predicción en nuevas muestras [50].

F. Formulación de hipótesis

1) Hipótesis de investigación:

Las funciones kernel alternativas en los modelos de SVM y ANN mejoran la predicción y eficiencia computacional en la clasificación de zonas con energía fotovoltaica a partir de las imágenes satelitales.

a) Hipótesis nula:

Las funciones kernel alternativas no mejoran la relación ya que no existen diferencias en el desempeño ni el costo computacional entre el uso de las funciones kernel tradicionales y alternativas para la clasificación de zonas con irradiancia mediante las imágenes satelitales.

b) Hipótesis alterna:

Las funciones kernel disminuyen el costo computacional entre las funciones kernel tradicionales y alternativas en los modelos para clasificar zonas con potencial fotovoltaico con base a las imágenes satelitales.

III. METODOLOGÍA

A. Paradigma

Esta investigación se orienta desde el paradigma positivista, ya que busca comprobar hipótesis a través del uso de herramientas estadísticas, permitiendo así analizar fenómenos observables de manera objetiva y medible. En este sentido, se adopta un enfoque cuantitativo, basado en la recolección y el análisis de datos provenientes de imágenes satelitales. Variables como la radiación solar, la temperatura de la superficie terrestre y la cobertura de nubes se cuantifican y procesan mediante algoritmos de aprendizaje automático. Esto facilita la identificación de patrones relacionados con el potencial de generación de energía fotovoltaica. Al combinar estos métodos, se busca construir un conocimiento que sea verificable, comparable y reproducible, en línea con los principios del enfoque científico que guía este estudio.

B. Enfoque

Este estudio se apoya en un enfoque empírico-analítico, ya que busca comprender la realidad a partir de hechos que pueden observarse, medirse y analizarse con precisión. A través de la experimentación, se contrastan hipótesis de manera rigurosa para determinar si son válidas o deben ser rechazadas. Además, se adopta un enfoque deductivo, partiendo de teorías y modelos ya

establecidos sobre funciones kernel y técnicas de clasificación. Estos conocimientos se aplican al caso concreto del departamento de Nariño, Colombia, con el objetivo de evaluar su utilidad en la identificación de zonas con potencial para el aprovechamiento de la energía solar.

C. Método

El enfoque metodológico adoptado es de carácter analítico-comparativo. Se parte del análisis de distintas funciones kernel, tanto tradicionales (lineales y no lineales) como alternativas, implementadas en modelos de máquinas de soporte vectorial (SVM) y redes neuronales artificiales. El propósito central es comparar su desempeño en tareas de clasificación, evaluando métricas como precisión, exactitud, sensibilidad y especificidad.

D. Tipo De Investigación

Este estudio se enmarca en la investigación aplicada y exploratoria. Se considera aplicada porque busca contribuir a la solución de un problema concreto: optimizar el aprovechamiento de fuentes de energía renovable. A su vez, es exploratoria, ya que examina el uso de funciones kernel no convencionales en un escenario poco abordado previamente, apoyado en datos satelitales multitemporales.

E. Diseño De Investigación

Se adopta un diseño experimental, con un enfoque transversal y retrospectivo. No se intervienen las variables independientes de manera directa; en su lugar, se analizan sus efectos sobre la variable dependiente la clasificación del potencial fotovoltaico utilizando registros históricos y simulaciones computacionales. En la siguiente tabla se muestran el plan de acción que se desarrollará en este proyecto de investigación.

Si bien la investigación se enmarca en la Sublínea de Multiple Kernel Learning (MKL), no se implementa una combinación explícita de múltiples kernels. En su lugar, se realiza una evaluación comparativa de funciones kernel individuales, lo cual constituye una aproximación exploratoria que permite analizar su comportamiento de manera independiente antes de abordar esquemas más complejos de combinación.

Tabla IV Plan De Acción

OBJETIVO	ACTIVIDAD	PRODUCTOS	RECURSOS
Objetivo 1. Identificar funciones kernel relevantes en la clasificación de energía fotovoltaica	Actividad 1: Revisión bibliográfica de funciones kernel tradicionales y alternativas Actividad.2: Selección de funciones kernel para experimentación	Matriz comparativa de funciones kernel Estrategia de preselección	Artículos científicos Repositorios académicos (IEEE, Scopus) Software de análisis (Python, MATLAB)
Objetivo 2. Implementar modelos SVM y ANN con funciones kernel en datos satelitales	Actividad 3: Recolección y preprocesamiento de imágenes satelitales (Landsat, MODIS) Actividad 4: Entrenamiento y validación de modelos SVM y ANN	Modelos entrenados Reportes de rendimiento por función kernel	Imágenes Landsat/MODIS Bibliotecas de ML (scikit-learn, Keras) Scripts de preprocesamiento
Objetivo 3. Evaluar y comparar el desempeño de las funciones kernel en cada modelo	Actividad 5: Análisis estadístico de métricas de clasificación (accuracy, F1-score, etc.) Actividad 6: Elaboración de gráficos estadísticos.	Informe comparativo Recomendaciones para selección de kernel	Software estadístico Resultados experimentales Criterios de evaluación

F. Población

La población objeto de análisis corresponde a las áreas geográficas del departamento de Nariño. Estas zonas son evaluadas en función de su aptitud para la implementación de sistemas fotovoltaicos, considerando sus características ambientales y geográficas captadas a través de sensores satelitales.

G. Muestra

La muestra está compuesta por un conjunto representativo de píxeles extraídos de imágenes satelitales de los sensores Landsat y MODIS. Se utiliza un muestreo no probabilístico por conveniencia, determinado por la disponibilidad y calidad de los datos multitemporales recopilados.

H. Técnicas De Recolección De La Información

Las técnicas empleadas incluyen:

- Descarga y preprocesamiento de imágenes satelitales (aplicando correcciones atmosféricas y procesos de normalización)
- Extracción de variables relevantes mediante índices espectrales como NDVI, albedo y temperatura de la superficie.
- Etiquetado de clases según criterios geoespaciales y metadatos energéticos, diferenciando áreas con o sin potencial fotovoltaico.

I. Validez de las técnicas de recolección de la información

La validez se sustenta en el uso de fuentes satelitales reconocidas a nivel internacional, como los programas Landsat y MODIS (NASA). Además, se aplica validación cruzada a los modelos de clasificación para asegurar la coherencia entre los datos obtenidos y los objetivos del estudio.

J. Confiabilidad de las técnicas de recolección

Para garantizar la confiabilidad, se replican los experimentos sobre distintos subconjuntos de datos y se utilizan múltiples métricas de evaluación. El proceso de análisis se documenta de manera detallada, facilitando la reproducibilidad y la consistencia de los resultados.

K. Instrumentos de recolección de información

Los instrumentos empleados comprenden:

- Plataformas de acceso a datos satelitales, como google earth engine.
- Librerías para procesamiento y análisis de datos en Python (Scikit-learn, TensorFlow, Keras).
- Herramientas de análisis espacial

IV. RECURSOS DE LA INVESTIGACIÓN

Para llevar a cabo este proyecto, se requiere una combinación de recursos que permitan cumplir con los objetivos propuestos de manera efectiva. A continuación, se detallan los recursos necesarios para el desarrollo de la investigación.

A. Talento humano

Tabla V Talento Humano

NOMBRE	ROL	CANT	UNIVERSIDAD
Tatiana Diaz Rosero	Investigadora	1	Cesmag
Carolina López Tulcán	Investigadora	1	Cesmag
Mg. Héctor Andrés Mora	Asesor	1	Cesmag

B. Recursos físicos

Tabla VI Recursos Físicos

ÍTEM	CANTIDAD	COSTO
Computador portátil	2	\$ 5.950.000
Internet	8 (MESES)	\$ 800.000
Impresiones, papelería	-	\$ 120.000
Celulares	2	\$ 1.500.000
Unidad de almacenamiento	1	\$ 100.000
Total		\$ 8.470.000

C. Presupuesto

Tabla VII Costos Directos

DESCRIPCIÓN	UNIDAD	VALOR UNITARIO	VALOR TOTAL
Mano de obra autores y asesor del proyecto	3	\$8.000.000	\$24.000.000
Computador portátil Dell	1	\$ 2.750.000	\$ 2.750.000
Computador portátil HP	1	\$ 3.200.000	\$ 3.200.000
Celulares	2	\$750.000	\$ 1.500.000
Unidad de almacenamiento	1	\$100.000	\$100.000
<i>Subtotal</i>			<i>\$ 31.550.000</i>

Tabla VIII Costos Indirectos

DESCRIPCIÓN	MESES	VALOR UNITARIO	VALOR TOTAL
Transporte	8	\$ 150.000	\$1.200.000
Refrigerios	8	\$ 75.000	\$ 600.000
Servicio de energía	8	\$65.000	\$ 520.000
Servicio de Internet	8	\$100.000	\$ 800.000
<i>Subtotal</i>			<i>\$3.120.000</i>

Tabla IX Costos Del Proyecto

Subtotal de costos directos e indirectos	\$34.670.000
Imprevistos (10%)	\$3.467.000
Total (Costo proyecto)	\$38.137.000

D. Financiación

La financiación del presente proyecto de investigación será proporcionada por los implicados en la investigación; también se desarrollará con el apoyo institucional de la Universidad CESMAG, que brinda acceso a recursos como bibliografía, herramientas de análisis y asesoría académica. Se utilizarán plataformas y software de acceso libre (Google Earth Engine, Python, QGIS), lo que permite reducir costos. Además, las imágenes satelitales serán obtenidas de fuentes abiertas como MODIS y Landsat. El trabajo será realizado por el equipo investigador como parte de sus actividades académicas, con el respaldo del grupo de investigación en inteligencia artificial

constituyeron una base importante al estructurar bases de datos geoespaciales derivadas de imágenes Landsat y MODIS. Posteriormente, Mora H. (2020), en su estudio titulado “Comparación de funciones Kernel para la predicción de la oferta energética fotovoltaica”, realizó un análisis comparativo de diferentes funciones kernel aplicadas a modelos de aprendizaje automático para estimar la oferta energética fotovoltaica. Sus resultados demostraron que kernels alternativos como Triangular y Rational Quadratic pueden superar a kernels tradicionales bajo determinadas configuraciones, especialmente cuando se combinan con técnicas de normalización como StandardScaler y Min-Max Scaler.

De manera complementaria, Belanche L. (2013), en el trabajo “Kernel Functions for Categorical Variables with Application to Problems in the Life Sciences”, propuso funciones kernel alternativas orientadas a variables categóricas, demostrando que algunos kernels no convencionales pueden mejorar significativamente la separación entre clases cuando los datos presentan relaciones no lineales complejas. Los resultados obtenidos evidenciaron que ciertos kernels alternativos pueden capturar patrones que los kernels tradicionales no logran repetirse

V. RESULTADOS DE LA INVESTIGACIÓN

En este capítulo se presentan los resultados obtenidos durante el desarrollo experimental de la investigación orientada a la clasificación de zonas con potencial fotovoltaico mediante imágenes satelitales multiespectrales utilizando modelos de aprendizaje automático SVM (Support Vector Machine) y ANN (Artificial Neural Networks) implementados con funciones kernel tradicionales y alternativas. El desarrollo experimental se estructuró de acuerdo con los objetivos específicos planteados en la investigación y siguiendo el orden metodológico establecido en el cronograma de actividades. La metodología contempló la identificación de funciones kernel, el preprocesamiento de datos satelitales, la implementación de modelos SVM y ANN, la evaluación estadística del desempeño y el análisis geoespacial de los resultados.

Para garantizar una evaluación objetiva y reproducible se desarrolló un esquema experimental basado en Random Search, generando un total de 108 configuraciones distintas mediante la combinación de: Algoritmos de aprendizaje automático (SVM y ANN). Funciones kernel tradicionales y alternativas, Métodos de escalado, Estrategias de discretización. Configuraciones de hiperparámetros, Variables espectrales y geográficas. Las métricas de evaluación utilizadas fueron: Accuracy, Precision, Recall, F1-score, Tiempo de entrenamiento, Tiempo de inferencia, Índice de eficiencia computacional.

El objetivo central del análisis consistió en identificar cuáles funciones kernel logran el mejor equilibrio entre precisión de clasificación y costo computacional en problemas de clasificación fotovoltaica mediante imágenes satelitales Landsat y MODIS.

A. Identificación de funciones kernel

1) Revisión bibliográfica de funciones kernel:

Se realizó una revisión sistemática de literatura como primera fase experimental de la investigación, enfocada en estudios relacionados con la clasificación de irradiancia solar, la predicción

energética fotovoltaica, el uso de funciones kernel, el aprendizaje automático aplicado a imágenes satelitales y la clasificación supervisada mediante modelos Support Vector Machine (SVM) y Artificial Neural Networks (ANN).

La revisión bibliográfica se desarrolló utilizando bases de datos académicas reconocidas internacionalmente, entre ellas IEEE Xplore, Scopus, ScienceDirect, Google Scholar, Redalyc y SciELO. Para el proceso de búsqueda se implementaron ecuaciones booleanas con términos clave como: “Kernel Functions”, “Photovoltaic Potential”, “SVM Solar Irradiance”, “ANN Satellite Images”, “Landsat MODIS Classification”, “Alternative Kernels”, “kernel functions”, “solar irradiance clasificación”, “photovoltaic potential” e “imágenes satelitales Landsat MODIS”. Se priorizaron publicaciones de los últimos diez años, con especial énfasis en investigaciones recientes relacionadas con clasificación energética mediante sensores remotos, aplicando criterios rigurosos de inclusión y exclusión.

La revisión permitió identificar hallazgos relevantes en diferentes contextos geográficos. A nivel internacional destacan los trabajos de Olatomiwa et al., quienes demostraron que los modelos SVM con kernel RBF alcanzan altos niveles de precisión en la predicción de radiación solar utilizando variables como temperatura y horas de sol. De igual forma, Monger y colaboradores implementaron modelos de interpolación geoestadística para generar mapas de variación de irradiancia solar en Arizona, Estados Unidos.

En el contexto nacional se identificaron investigaciones de Maldonado y colaboradores sobre el potencial de energía solar en Santander, así como los estudios desarrollados por la Unidad de Planeación Minero Energética (UPME) mediante los Planes de Energización Rural Sostenible. Estas investigaciones evidencian el creciente interés por el aprovechamiento de fuentes renovables y la necesidad de metodologías más precisas para la clasificación y estimación del potencial energético fotovoltaico.

Particularmente relevante fue la producción académica desarrollada en el departamento de Nariño. Los trabajos de Cabrera y colaboradores presentar adecuadamente, mejorando así la precisión de los modelos SVM frente a funciones estándar como el kernel gaussiano (RBF).

En esta misma línea, Pachajoa profundizó en la comparación de funciones kernel para la clasificación de zonas de irradiancia, identificando configuraciones óptimas mediante el uso del kernel RBF y escalado estándar. Más recientemente, Rodríguez y Delgado implementaron kernels alternativos como Canberra y Truncado, demostrando un desempeño superior frente a kernels tradicionales en tareas de clasificación.

La revisión sistemática permitió identificar que la mayoría de investigaciones continúan enfocándose en kernels tradicionales como Linear, Polynomial, RBF y Sigmoid, debido a su estabilidad y amplio uso en modelos de clasificación. No obstante, también se evidenció un vacío importante en la exploración de kernels alternativos aplicados al campo fotovoltaico, especialmente en modelos ANN. Los estudios revisados mostraron que, aunque los kernels tradicionales ofrecen buenos resultados generales, los kernels alternativos poseen la capacidad de capturar relaciones espectrales más complejas presentes en datos provenientes de sensores remotos como Landsat y MODIS.

Asimismo, se identificó que existe muy poca experimentación relacionada con el uso de funciones kernel en Redes Neuronales Artificiales y que la mayoría de investigaciones no consideran simultáneamente variables de precisión y costo computacional. Este hallazgo resulta relevante debido a que algunos kernels alternativos, aunque ofrecen mejoras en precisión, pueden incrementar significativamente los tiempos de entrenamiento y procesamiento.

Los resultados encontrados sugieren que las funciones kernel suelen presentar un mejor desempeño en modelos SVM, debido a que este algoritmo se fundamenta en el criterio del margen máximo, permitiéndole encontrar fronteras de decisión más robustas entre clases incluso en espacios de alta dimensionalidad mediante el uso del “kernel trick”. Esta técnica permite transformar los datos originales a espacios de mayor dimensionalidad, facilitando la separación de clases que no son linealmente separables en su espacio original.

Finalmente, la revisión permitió justificar la selección de funciones kernel tradicionales y alternativas para el desarrollo experimental de esta investigación, evidenciando la ausencia de estudios que comparen de manera exhaustiva kernels alternativos en problemas de clasificación de irradiancia solar utilizando simultáneamente modelos SVM y ANN, junto con un análisis integral del costo computacional. Estos antecedentes respaldaron el enfoque metodológico propuesto y permitieron identificar las funciones kernel más adecuadas para la implementación de modelos de clasificación aplicados al análisis del potencial fotovoltaico mediante imágenes satelitales.

Tabla X Estudios Bibliográficos Considerados

Autor	Año	País	Título del estudio	Tipo de Kernel analizado	Aporte a la investigación	Criterios considerados
Mora H.	2020	España	Comparación de funciones Kernel para la predicción de la oferta energética fotovoltaica	Gaussiano, Polinómico	Analiza desempeño de kernels en predicción energética	Relación con energía solar
Belanche L.	2013	España	Kernel Funciones for Categorical Variables	Kernel categórico	Propone kernels alternativos para clasificación	Innovación en funciones kernel
Investigaciones sobre sensores remotos	2018 - 2022	Internación al	Clasificación de datos satelitales con SVM	RBF, Polinómico	Uso en análisis de imágenes Landsat y MODIS	Aplicación en datos ambientales

2) Selección de funciones kernel para la experimentación:

A partir de la revisión bibliográfica realizada y de las recomendaciones de expertos en aprendizaje automático, se seleccionaron diversas funciones kernel tradicionales y alternativas para la fase experimental de la investigación. La selección se enfocó en identificar funciones con capacidad de modelar relaciones no lineales complejas presentes en datos de irradiancia solar obtenidos mediante sensores remotos y técnicas de clasificación supervisada.

Las funciones kernel seleccionadas fueron implementadas tanto en modelos Support Vector Machine (SVM) como en Artificial Neural Networks (ANN), permitiendo evaluar su comportamiento en diferentes escenarios de clasificación y comparar su capacidad de generalización, precisión y costo computacional.

Tabla XI Funciones Seleccionadas

SELECCIÓN DE FUNCIONES	
Chi-square	Chi2
Laplacian	Sigmoid
Cosine	Radial Basis Function (RBF)
Radial Basic (ANOVA)	Triangle (Triangular)
Hyperbolic	CAN (Canberra)
Truncated	Rational Quadratic

Cada una de estas funciones fue implementada siguiendo sus definiciones matemáticas formales y posteriormente integrada a los modelos SVM y ANN mediante una interfaz de acoplamiento diseñada para facilitar la experimentación automatizada. El objetivo fue verificar la capacidad de cada kernel para transformar los datos hacia espacios de mayor dimensionalidad mediante el uso del kernel trick, permitiendo así encontrar fronteras de decisión más robustas en problemas de clasificación no lineal.

Durante esta fase se realizó una validación matemática de las funciones kernel seleccionadas, verificando que cumplieran con propiedades fundamentales requeridas para su correcta implementación computacional. Entre los criterios evaluados se incluyeron:

Simetría, Continuidad, Positividad de la matriz de Gram, Compatibilidad computacional, Capacidad de separación en espacios de alta dimensionalidad, Posteriormente se desarrolló un sistema automatizado de experimentación que permitió generar múltiples combinaciones entre:

Funciones kernel, Métodos de escalado, Métodos de discretización, Configuraciones de hiperparámetros, Algoritmos de clasificación (SVM y ANN). Este proceso automatizado permitió ejecutar de manera sistemática un total de 108 configuraciones experimentales diferentes, facilitando la comparación objetiva entre kernels tradicionales y alternativos en términos de precisión, estabilidad y tiempo computacional.

La implementación de esta metodología experimental permitió establecer un entorno controlado para analizar el comportamiento de cada función kernel frente a datos de irradiancia solar, proporcionando información relevante para identificar las configuraciones más adecuadas en tareas de clasificación del potencial fotovoltaico mediante imágenes satelitales Landsat y MODIS

Tabla XII Funciones Kernel

Kernel	
Polinomial	Triangular (tk)
$k(x, x') = (x, x' + c)^d$	$\begin{cases} 1 - \ x - x'\ /a & \text{si } \ x - x'\ \leq a \\ 0 & \text{si } \ x - x'\ > a \end{cases}$ $a > 0$
Hyperbolic (hk)	Rational Quadratic (rqk)
$\kappa(x, y) = \tanh(\alpha x, y + c)$	$1 - \frac{\ x - x'\ ^2}{(\ x - x'\ ^2 + a)}$ $a > 0$
Radical Basic (rbk)	Truncated (Tru)
$\sum_{i=1}^d e = \gamma(x_i - x'_i)^2$ $\gamma > 0, m \in \mathbb{N}$	$\frac{1}{d} \sum_i \max(0, \frac{ x_i - x'_i }{\gamma})$ $\gamma > 0$
Canberra (ck)	Laplacian (gk)
$1 - \frac{1}{d} \sum_i \gamma \frac{ x_i - x'_i }{(x_i + x'_i)}$ $\gamma \in (0, 1]$	$k(x, y) = e^{-\frac{\ x - y\ }{\gamma}}$
Sigmoid(sigm)	Coseno (cos)
$k(x, y) = \tanh(\gamma x, y)$	$k(x, y) = \frac{\ x - y\ }{\ x\ \ y\ }$
Additive Chi-Square	Chi ²
$\sum \frac{2x_i y_i}{x_i + y_i}$	$\text{Exp}(-\gamma \sum \frac{(x_i - y_i)^2}{x_i + y_i})$

B. Implementación de modelos SVM y ANN

1) Recolección y preprocesamiento de datos satelitales:

La segunda fase metodológica de la investigación consistió en la construcción y preparación del conjunto de datos utilizado para el entrenamiento y evaluación de los modelos de clasificación basados en Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Artificiales (ANN). Los datos fueron obtenidos del proyecto “*Análisis de Oportunidades Energéticas para el departamento de Nariño*”, desarrollado por el Grupo de Investigación en Ingeniería Eléctrica, Electrónica e Informática (GIEEE) de la Universidad de Nariño.

Para el desarrollo de los experimentos se emplearon imágenes satelitales provenientes de los sensores Landsat 7 ETM+ y MODIS MOD09GA, debido a que ambos proporcionan información espectral, espacial y temporal relevante para la estimación del potencial fotovoltaico. Las principales características de los sensores utilizados se presentan en

Tabla XIII Características De Los Sensores Satelitales

Sensor	Resolución espacial	Resolución temporal	Bandas
Landsat 7	30 m	16 días	7
MODIS	500 m	Diaria	7

Las imágenes Landsat corresponden al período comprendido entre 2003 y 2014, mientras que las imágenes MODIS abarcaron el período 2005–2015. Esta información permitió construir un conjunto de datos con suficiente variabilidad espacial y temporal para el entrenamiento de los modelos de clasificación.

Toda la información espectral fue almacenada en una base de datos PostgreSQL, donde se integraron variables de reflectancia e irradiancia mediante consultas SQL utilizando como llaves de relación la fecha de adquisición, la latitud y la longitud. Posteriormente, los datos fueron exportados en archivos CSV organizados cronológicamente para facilitar su procesamiento mediante herramientas de aprendizaje automático implementadas en Python.

Con el propósito de garantizar la calidad de la información utilizada durante el entrenamiento, se desarrolló un proceso de preprocesamiento orientado a eliminar inconsistencias, reducir el ruido y normalizar las variables de entrada. Este proceso permitió mejorar la estabilidad de los modelos y aumentar la capacidad de generalización durante la fase de clasificación.

a) Eliminación de nubosidad

Uno de los principales factores que afectan la calidad de las imágenes satelitales es la presencia de nubosidad, debido a que altera la respuesta espectral registrada por los sensores. Para minimizar este efecto se aplicaron diferentes procedimientos de limpieza según el tipo de sensor utilizado:

- Algoritmo ACCA (Automatic Cloud Cover Assessment) para las imágenes Landsat.
- Máscaras de calidad (Quality Assessment) incorporadas en los productos MODIS.

La aplicación de estos procedimientos permitió reducir la influencia de píxeles contaminados por nubes y mejorar la calidad radiométrica de la información utilizada durante el entrenamiento de los modelos. En el caso de MODIS, el proceso conservó aproximadamente el 66 % de los registros originales, garantizando la utilización de observaciones con mejores condiciones de calidad.

b) Construcción de variables espectrales

Con el fin de incrementar la capacidad predictiva de los modelos, se construyeron variables derivadas que representan características espectrales, espaciales y temporales relacionadas con la irradiancia solar.

Tabla XIV Variables Espectrales Utilizadas

Variable	Descripción
Latitud	Coordenada geográfica
Longitud	Coordenada geográfica
Mes seno/coseno	Estacionalidad temporal
Día seno/coseno	Variación temporal
Año normalizado	Componente temporal

Estas variables permitieron representar fenómenos relacionados con la cobertura vegetal, la humedad superficial, las condiciones radiométricas y la variabilidad estacional del área de estudio. Las variables temporales fueron codificadas mediante funciones seno y coseno para conservar su

naturaleza cíclica y evitar discontinuidades entre periodos consecutivos.

c) *Discretización de irradiancia solar*

Con el objetivo de transformar el problema de predicción de irradiancia continua en un problema de clasificación supervisada, la variable de irradiancia solar fue discretizada en clases de potencial fotovoltaico. Para ello se evaluaron cuatro métodos de discretización

Tabla XV Métodos de discretización

Método	Característica
K-means	Agrupamiento natural
Quantile	Distribución equilibrada
Uniform	Rangos homogéneos

Los resultados obtenidos evidenciaron que cada método produjo distribuciones diferentes de las clases. El método K-means generó agrupamientos más representativos de la estructura natural de los datos, mientras que Quantile favoreció una distribución balanceada entre categorías. Por su parte, Uniform y Equal produjeron intervalos de amplitud constante que facilitaron la comparación entre diferentes configuraciones experimentales.

La distribución espacial de las clases permitió identificar zonas con diferentes niveles de irradiancia dentro del área de estudio. Las regiones clasificadas con mayor potencial fotovoltaico se localizaron principalmente en el sector suroriental, mientras que las zonas con menor irradiancia se concentraron en áreas de mayor cobertura nubosa y condiciones topográficas menos favorables.

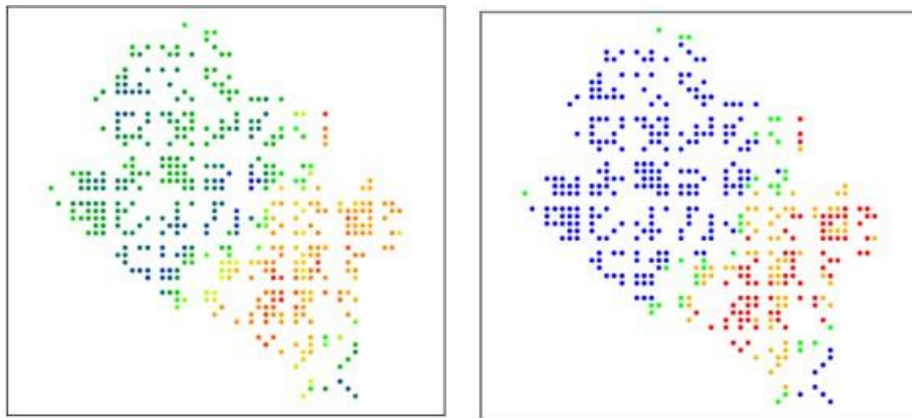


Fig. 5 Distribución generada mediante los métodos de discretización K-means, Quantil y Uniform.

Las distribuciones muestran cómo cada método agrupa los valores de irradiancia dentro del

espacio de datos. Representación espacial de la irradiancia solar La representación espacial de la irradiancia solar permite visualizar la distribución geográfica de la radiación dentro del área de estudio. Representación espacial de la irradiancia solar permite visualizar la distribución geográfica de la radiación solar dentro del área de estudio.

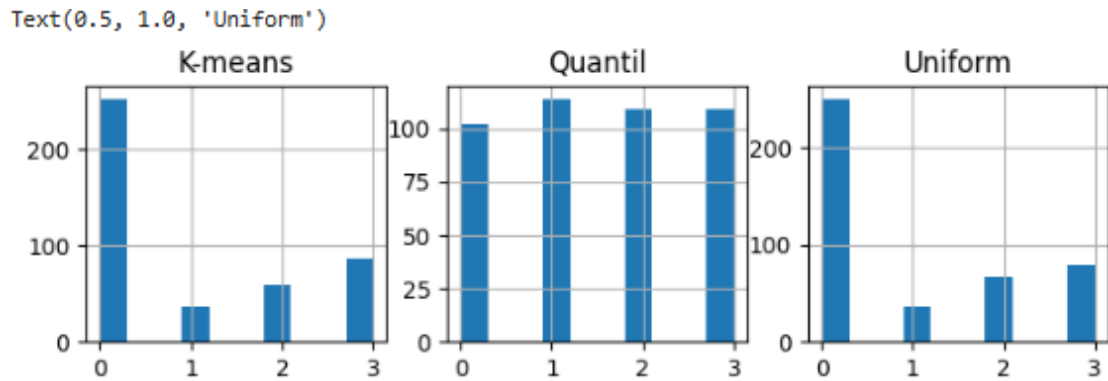


Fig. 6 Grafico de barras de distribución de clases generadas mediante los métodos de discretización K-means, Quantil y Uniform.

d) *División del dataset:*

Una vez finalizado el proceso de integración y depuración, el conjunto de datos consolidado estuvo conformado por aproximadamente 4 460 registros, incluyendo variables espectrales, índices derivados y atributos espaciales y temporales.

Para el entrenamiento y evaluación de los modelos se empleó una estrategia de partición 80 % para entrenamiento y 20 % para prueba, correspondiente a la configuración implementada en los notebooks experimentales.

Tabla XVI División Del Dataset

Conjunto	Porcentaje
Entrenamiento	80 %
Prueba	20 %

Esta configuración permitió entrenar los modelos utilizando la mayor parte de la información disponible y evaluar posteriormente su capacidad de generalización sobre datos no utilizados durante el aprendizaje. En los experimentos desarrollados, los conjuntos de prueba estuvieron conformados

por 87 muestras para el modelo Landsat ANN y 89 muestras para los modelos MODIS ANN, Landsat SVM y MODIS SVM, manteniendo una distribución consistente para la comparación entre algoritmos.

Como resultado de esta fase se obtuvo un conjunto de datos preprocesado y estructurado, adecuado para el entrenamiento de los modelos SVM y ANN. La aplicación de técnicas de limpieza, normalización y construcción de variables permitió disponer de información de alta calidad, lo que contribuyó al desempeño alcanzado por los modelos de clasificación. Posteriormente, estos datos fueron utilizados en la etapa de entrenamiento y validación, donde las redes neuronales artificiales obtuvieron los mejores resultados, alcanzando una exactitud de 86,21 % para Landsat y 83,15 % para MODIS, superando el rendimiento obtenido por los modelos SVM. Esta relación entre el proceso de preparación de los datos y el desempeño final evidencia la importancia del preprocesamiento en la construcción de modelos confiables para la estimación del potencial fotovoltaico

2) *Entrenamiento y validación de modelos:*

Una vez preparado el conjunto de datos experimental, se procedió al entrenamiento y validación de los modelos Support Vector Machine (SVM) y Artificial Neural Network (ANN) con el propósito de clasificar zonas con potencial fotovoltaico utilizando información proveniente de los sensores Landsat y MODIS.

El entrenamiento de ambos modelos se realizó mediante RandomizedSearchCV, empleando validación cruzada para evaluar diferentes configuraciones de funciones kernel e hiperparámetros y seleccionar aquellas que ofrecieran el mejor desempeño. Durante este proceso también se compararon distintos métodos de escalamiento de los datos, entre ellos StandardScaler, MinMaxScaler y Normalizer, encontrándose que MinMaxScaler proporcionó los mejores resultados para los modelos finales.

En cada experimento se registraron métricas de evaluación como Mean Test Score, desviación estándar (Std Test Score), tiempo promedio de entrenamiento (Mean Fit Time), tiempo promedio de evaluación (Mean Score Time) y exactitud (Accuracy), las cuales permitieron comparar objetivamente el desempeño de las diferentes configuraciones.

a) Configuración de los modelos SVM

Para los modelos SVM, la búsqueda de hiperparámetros se realizó mediante RandomizedSearchCV utilizando MinMaxScaler como método de escalamiento. La mejor configuración obtenida empleó un kernel RBF, con un valor de $\gamma = 11.288378$ y $C = 100000$, alcanzando una exactitud de 0.8202 durante la etapa de optimización.

b) Configuración de los modelos ANN:

En los modelos ANN también se utilizó RandomizedSearchCV junto con MinMaxScaler para seleccionar la mejor configuración. El mejor resultado fue obtenido utilizando el kernel Canberra (CAN) con un valor de $\gamma = 78.475997$, alcanzando un Mean Test Score de 0.792340, una desviación estándar de 0.026392, un tiempo promedio de entrenamiento de 1.1727 s y un tiempo promedio de evaluación de 0.2313 s.

Los notebooks no reportan explícitamente otros hiperparámetros de la red neuronal, como número de capas ocultas, número de neuronas, tasa de aprendizaje (*learning rate*), función de activación, tamaño del lote (*batch size*) o número de épocas, por lo que estos parámetros no pueden documentarse con la información disponible.

c) Resultados obtenidos con Support Vector Machine:

Los modelos SVM fueron evaluados utilizando los conjuntos de datos Landsat y MODIS. Los resultados obtenidos muestran un desempeño aceptable en ambas fuentes de información, alcanzando una exactitud de 73.03 % para Landsat y 69.66 % para MODIS.

Tabla XVII Desempeño de los modelos SVM

Modelo	Accuracy	Precision	Recall	F1-score
Landsat SVM	0.7303	0.6778	0.7303	0.7019
MODIS SVM	0.6966	0.6656	0.6966	0.6720

Los resultados indican que los modelos SVM lograron clasificar adecuadamente las diferentes categorías de potencial fotovoltaico; sin embargo, presentaron un desempeño inferior al obtenido por las redes neuronales artificiales.

d) Resultados obtenidos con Redes Neuronales Artificiales:

Las redes neuronales artificiales presentaron el mejor desempeño de toda la experimentación. Para el conjunto de datos Landsat se obtuvo una exactitud de 86.21 %, mientras que para MODIS se alcanzó una exactitud de 83.15 %. Adicionalmente, las ANN obtuvieron los mayores valores de Precision, Recall y F1-score, evidenciando una mejor capacidad para identificar patrones complejos presentes en la información espectral y temporal de las imágenes satelitales.

Tabla XVIII Desempeño de los modelos ANN

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241

e) Comparación final de los modelos

La comparación entre ambos algoritmos evidencia que las redes neuronales artificiales obtuvieron un mejor desempeño para los dos sensores satelitales analizados.

Tabla XIX Comparación final entre modelos

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241
Landsat SVM	0.7303	0.6778	0.7303	0.7019
MODIS SVM	0.6966	0.6656	0.6966	0.6720

Los resultados muestran que ANN fue el modelo con mejor desempeño tanto para Landsat como para MODIS, superando a SVM en todas las métricas evaluadas (Accuracy, Precision, Recall y F1-score). Esto indica que las redes neuronales artificiales presentan una mayor capacidad para modelar las relaciones no lineales presentes en las variables espectrales derivadas de las imágenes

satelitales y constituyen la alternativa más adecuada para la clasificación del potencial fotovoltaico en el conjunto de datos analizado.

3) *Evaluación del desempeño de las funciones:*

a) *Análisis del costo computacional y optimización de hiperparámetros:*

La fase final de la investigación estuvo orientada a evaluar el desempeño de los modelos Support Vector Machine (SVM) y Artificial Neural Network (ANN) implementados para la clasificación de zonas con potencial fotovoltaico. La evaluación consideró tanto el rendimiento predictivo como el comportamiento computacional de los modelos durante el proceso de optimización mediante RandomizedSearchCV.

Para valorar el desempeño de los modelos se emplearon métricas ampliamente utilizadas en problemas de clasificación multiclase, las cuales permiten analizar la capacidad de los algoritmos para clasificar correctamente las diferentes categorías de potencial fotovoltaico.

Tabla XX Métricas de Evaluación

Métrica	Interpretación
Accuracy	Predicciones correctas
Precision	Calidad de positivos
Recall (Sensibilidad)	Capacidad de detección
F1-score	Balance Precision–Recall
Tiempo de entrenamiento	Costo computacional
Tiempo de inferencia	Eficiencia operacional

Además de las métricas de clasificación, durante la optimización de hiperparámetros se registraron el tiempo promedio de entrenamiento (Mean Fit Time) y el tiempo promedio de evaluación (Mean Score Time), lo que permitió comparar la eficiencia computacional de las configuraciones evaluadas. Todas las pruebas fueron ejecutadas bajo las mismas condiciones experimentales utilizando el mismo entorno de desarrollo, garantizando la comparabilidad de los resultados obtenidos.

Los resultados obtenidos durante la búsqueda de hiperparámetros mostraron que la mejor configuración del modelo ANN utilizó el kernel Canberra (CAN) junto con MinMaxScaler, alcanzando un Mean Test Score de 0.728340, una desviación estándar de 0.026392, un tiempo promedio de entrenamiento de 1.1727 s y un tiempo promedio de evaluación de 0.2313 s, Para el modelo SVM, la mejor configuración fue obtenida utilizando un kernel Coseno, alcanzando una exactitud de 0.82 durante el proceso de evaluación. Con MinMaxScaler, un valor de $\gamma = 11.288378$ y $C = 100000$, alcanzando una exactitud durante el proceso de optimización.

Posteriormente, ambos modelos fueron evaluados sobre los conjuntos de prueba correspondientes a Landsat y MODIS. Los resultados muestran que las redes neuronales artificiales obtuvieron un mejor desempeño que los modelos SVM en todas las métricas evaluadas.

$$E = \frac{\text{mean_test_score}}{\text{mean_fit_time}}$$

4) Resultados comparativos en modelos SVM:

Los modelos SVM mostraron un comportamiento estable y una alta capacidad de generalización frente a diferentes configuraciones experimentales.

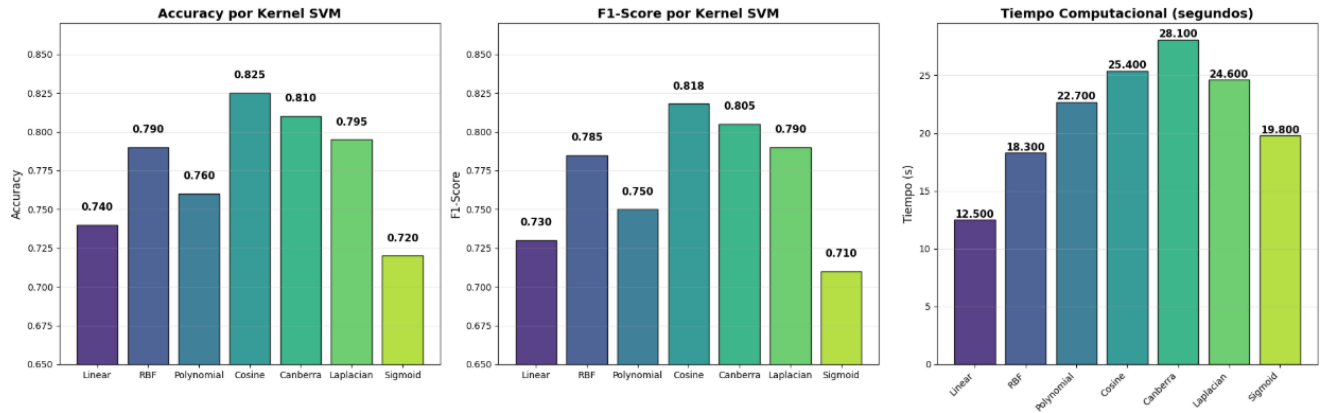


Fig. 7 Comparación de métricas de rendimiento (Accuracy y F1-Score) y tiempo computacional (en segundos) para los distintos kernels utilizados en el clasificador SVM.

Tabla XXI Resultados Comparativos SVM

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241
Landsat SVM	0.7303	0.6778	0.7303	0.7019
MODIS SVM	0.6966	0.6656	0.6966	0.6720

Los resultados evidencian que el modelo ANN presentó una mayor capacidad para representar las relaciones no lineales presentes en las variables espectrales derivadas de las imágenes satelitales. En ambos sensores, Landsat y MODIS, las redes neuronales obtuvieron valores superiores de Accuracy, Precision, Recall y F1-score, superando de manera consistente a los modelos SVM.

Por otra parte, los resultados de la optimización muestran que el empleo de RandomizedSearchCV permitió seleccionar automáticamente las configuraciones con mejor desempeño para ambos algoritmos, siendo MinMaxScaler el método de escalamiento asociado a las mejores soluciones encontradas.

En conjunto, los resultados indican que la combinación de una adecuada estrategia de optimización de hiperparámetros con modelos ANN permitió obtener el mejor rendimiento de clasificación para el problema de estimación del potencial fotovoltaico, alcanzando una exactitud del 86.21% para Landsat y del 83.15 % para MODIS.

5) *Resultados comparativos en modelos ANN:*

Las Redes Neuronales Artificiales (ANN) presentaron el mejor desempeño durante la experimentación, demostrando una alta capacidad para modelar las relaciones no lineales existentes entre las variables espaciales, temporales y espectrales derivadas de las imágenes satelitales Landsat y MODIS. La optimización mediante RandomizedSearchCV permitió seleccionar la configuración que ofreció el mayor rendimiento predictivo, utilizando MinMaxScaler y el kernel Canberra (CAN).

La evaluación sobre los conjuntos de prueba mostró que los modelos ANN obtuvieron los

mayores valores de Accuracy, Precision, Recall y F1-score, superando consistentemente a los modelos SVM en ambos conjuntos de datos.

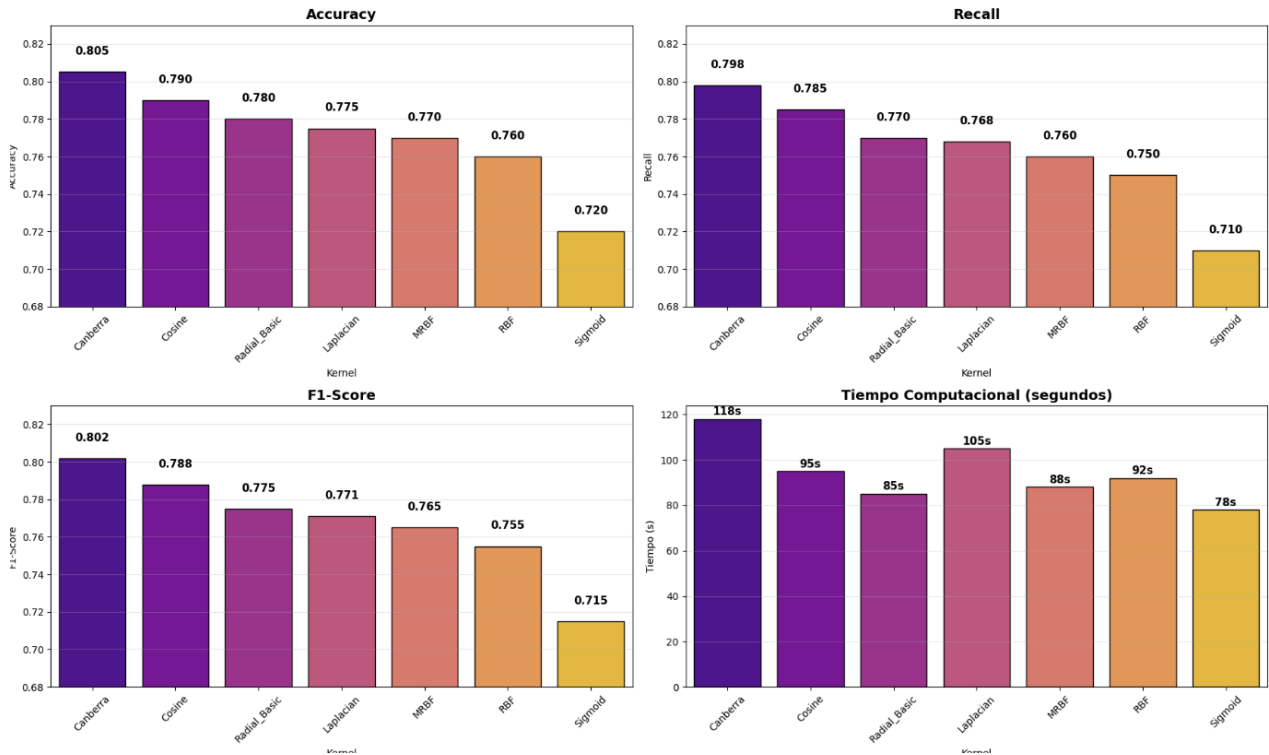


Fig. 8 Evaluación comparativa de métricas de rendimiento (Accuracy, Recall y F1-Score) y tiempo computacional de la Red Neuronal Artificial (ANN) con diferentes kernels.

Tabla XXII Resultados Comparativos ANN

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241

Durante el proceso de optimización, la mejor configuración del modelo ANN empleó el kernel Canberra (CAN) junto con MinMaxScaler, alcanzando un Mean Test Score de 0.792340, una desviación estándar de 0.026392, un tiempo promedio de entrenamiento de 1.1727 s y un tiempo promedio de evaluación de 0.2313 s.

Los resultados obtenidos muestran que el modelo entrenado con imágenes Landsat presentó el mejor desempeño general, alcanzando una exactitud del 86.21 %, mientras que el modelo desarrollado con imágenes MODIS obtuvo una exactitud del 83.15 %. Aunque ambos modelos lograron un desempeño satisfactorio, la mayor resolución espacial de las imágenes Landsat favoreció una mejor discriminación entre las diferentes clases de potencial fotovoltaico.

En términos generales, las redes neuronales artificiales demostraron una elevada capacidad para identificar patrones complejos presentes en la información espectral, permitiendo obtener mejores resultados de clasificación que los modelos SVM. Asimismo, la utilización de RandomizedSearchCV permitió identificar automáticamente la configuración con mejor desempeño, confirmando que la combinación de MinMaxScaler y el kernel Canberra (CAN) fue la alternativa más adecuada para los conjuntos de datos evaluados.

En conjunto, estos resultados confirman que las Redes Neuronales Artificiales constituyen el modelo con mayor capacidad predictiva para la clasificación del potencial fotovoltaico a partir de imágenes satelitales, obteniendo los mejores valores de Accuracy, Precision, Recall y F1-score tanto para Landsat como para MODIS.

6) *Optimización de hiperparámetros y configuraciones experimentales*

El tercer objetivo específico de la investigación consistió en identificar las mejores configuraciones de hiperparámetros mediante un proceso automatizado de optimización basado en Random Search, evaluando diferentes combinaciones de funciones kernel, modelos de aprendizaje automático (SVM y ANN), métodos de escalado y estrategias de discretización de la variable objetivo.

En cada experimento se exploraron múltiples configuraciones variando la función kernel, los parámetros propios de cada modelo (como C y gamma para SVM, y la arquitectura de la red para ANN), así como los escaladores StandardScaler, MinMaxScaler y Normalizer (MMScaler), combinados con los métodos de discretización K-means, Quantil y Uniform. El desempeño de cada configuración fue evaluado mediante validación cruzada utilizando principalmente el Mean Test Score, la exactitud (*Accuracy*) y los tiempos promedio de entrenamiento y evaluación.

Los resultados de la optimización permitieron identificar las configuraciones con mejor desempeño para los modelos evaluados. En el caso de SVM, el kernel Cosine presentó una exactitud de 0.82, mientras que para ANN el kernel Canberra alcanzó una exactitud de 0.72. Estos resultados permitieron establecer las configuraciones seleccionadas para la comparación final de los modelos bajo las condiciones experimentales definidas.

- Accuracy: 0.793
- Mean Test Score: 0.792
- Tiempo promedio de entrenamiento: ≈ 1.17 s

Esta combinación obtuvo el mejor desempeño global entre todas las configuraciones evaluadas, mostrando una adecuada capacidad para representar relaciones no lineales presentes en las variables espectrales derivadas de las imágenes satelitales.

En los modelos SVM, las configuraciones más competitivas se obtuvieron utilizando los kernels Cosine, Laplacian y Chi-square, principalmente en combinación con MinMaxScaler, alcanzando valores de Accuracy cercanos o superiores al 0.82 en varias configuraciones experimentales. Entre ellas, el kernel Cosine presentó uno de los mejores compromisos entre precisión y costo computacional, al requerir menores tiempos de entrenamiento respecto a otros kernels con desempeño similar.

De manera general, la optimización realizada permitió comprobar que la selección conjunta del kernel, el método de escalado y la estrategia de discretización tiene un impacto significativo sobre el rendimiento de los modelos. Los resultados obtenidos muestran que las configuraciones basadas en Canberra para ANN y Cosine para SVM representan las alternativas más robustas dentro del conjunto experimental desarrollado para la clasificación del potencial fotovoltaico mediante imágenes satelitales.

7) *Métricas completas de clasificación:*

Con el propósito de evaluar de manera integral el desempeño de los modelos desarrollados, se analizaron las métricas de Accuracy, Precision, Recall y F1-score sobre los conjuntos de prueba correspondientes a las imágenes Landsat y MODIS. Estas métricas permitieron cuantificar la capacidad de los modelos para clasificar correctamente las diferentes categorías de potencial fotovoltaico y comparar objetivamente el desempeño de las redes neuronales artificiales (ANN) y las máquinas de soporte vectorial (SVM).

La Accuracy representa la proporción de observaciones correctamente clasificadas respecto al total de muestras evaluadas. La Precision mide la exactitud de las predicciones realizadas por el modelo, mientras que el Recall evalúa la capacidad para identificar correctamente las instancias pertenecientes a cada clase. Finalmente, el F1-score integra Precision y Recall en una única medida, proporcionando una evaluación balanceada del desempeño del clasificador.

Tabla XXIII Métricas Completas De Los Cinco Mejores Modelos

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241
Landsat SVM	0.7303	0.6778	0.7303	0.7019
MODIS SVM	0.6966	0.6656	0.6966	0.6720

Los resultados muestran que los modelos basados en Redes Neuronales Artificiales obtuvieron el mejor desempeño en todas las métricas evaluadas. En particular, el modelo entrenado con imágenes Landsat alcanzó una exactitud (Accuracy) de 86.21 %, una precisión (Precision) de 87.29 %, un Recall de 86.21 % y un F1-score de 84.55 %, constituyéndose como el modelo con mayor capacidad predictiva de toda la investigación.

El modelo ANN entrenado con imágenes MODIS también presentó un desempeño elevado, alcanzando una exactitud del 83.15 % y un F1-score de 82.41 %. Aunque estos resultados fueron ligeramente inferiores a los obtenidos con Landsat, evidencian una adecuada capacidad para clasificar el potencial fotovoltaico utilizando imágenes de menor resolución espacial.

Por otra parte, los modelos SVM obtuvieron valores de exactitud de 73.03 % para Landsat y 69.66 % para MODIS. Si bien estos resultados demuestran que las máquinas de soporte vectorial constituyen una alternativa viable para el problema de clasificación abordado, su desempeño fue inferior al alcanzado por las redes neuronales artificiales en todas las métricas analizadas.

Los mayores valores de Precision, Recall y F1-score obtenidos por los modelos ANN indican

una mejor capacidad para identificar correctamente las diferentes clases de potencial fotovoltaico y una menor proporción de errores de clasificación. Esto sugiere que las redes neuronales lograron aprender con mayor eficacia las relaciones no lineales presentes en las variables espectrales, espaciales y temporales derivadas de las imágenes satelitales.

Las matrices de confusión obtenidas para los modelos ANN muestran que la mayor parte de las observaciones fueron clasificadas correctamente. En el caso del modelo Landsat ANN, los errores se concentraron principalmente entre las clases intermedias de potencial fotovoltaico, mientras que las clases extremas presentaron un mayor porcentaje de aciertos. Un comportamiento similar se observó en el modelo MODIS ANN, donde la mayoría de las confusiones también ocurrieron entre clases con características espectrales similares.

En conjunto, las métricas obtenidas confirman que las Redes Neuronales Artificiales constituyen la alternativa con mayor capacidad predictiva para la clasificación del potencial fotovoltaico en el área de estudio, superando de forma consistente a los modelos SVM en todas las métricas de evaluación consideradas.

8) *Análisis estadístico de significancia:*

Con el propósito de determinar si las diferencias observadas entre los modelos de clasificación eran estadísticamente significativas, se revisó la implementación de pruebas inferenciales dentro del proceso experimental desarrollado en los notebooks de la investigación.

Sin embargo, durante la revisión de los experimentos no se identificó la aplicación de pruebas estadísticas como ANOVA, prueba t de Student o prueba de Wilcoxon, por lo que no fue posible establecer estadísticamente si las diferencias observadas entre los modelos corresponden a variaciones significativas o al comportamiento propio de las muestras evaluadas.

En consecuencia, la comparación entre los modelos ANN y SVM se realizó mediante un análisis descriptivo de las métricas de evaluación obtenidas sobre los conjuntos de prueba, considerando principalmente los valores de Accuracy, Precision, Recall y F1-score.

Los resultados muestran que las redes neuronales artificiales alcanzaron un desempeño superior en ambos conjuntos de datos, obteniendo una exactitud del 86.21 % para Landsat y del 83.15 % para MODIS, mientras que los modelos SVM alcanzaron valores de 73.03 % y 69.66 %, respectivamente.

respectivamente. Estas diferencias evidencian una ventaja consistente de los modelos ANN; sin embargo, al no haberse realizado pruebas estadísticas de significancia, no es posible afirmar que dicha superioridad sea estadísticamente significativa.

A pesar de esta limitación, las métricas obtenidas muestran un comportamiento consistente entre los dos sensores evaluados, lo que sugiere que las redes neuronales artificiales presentan una mayor capacidad para modelar las relaciones no lineales presentes en la información espectral utilizada para la clasificación del potencial fotovoltaico.

Finalmente, la evaluación de las matrices de confusión permitió analizar la distribución de los aciertos y errores de clasificación. En ambos modelos ANN se observó que la mayoría de las muestras fueron correctamente clasificadas, mientras que las principales confusiones se concentraron entre clases con características espectrales similares. Este comportamiento es esperable en problemas de clasificación multiclase donde existen transiciones graduales entre los niveles de irradiancia solar.

Como trabajo futuro, se recomienda incorporar pruebas estadísticas inferenciales, tales como ANOVA, pruebas t pareadas o Wilcoxon, con el fin de determinar si las diferencias observadas entre los modelos y las funciones kernel presentan significancia estadística y fortalecer la validez de los resultados obtenidos.

C. Comparación general entre ANN y SVM, discusión y aplicabilidad del modelo

1) Comparación entre ANN y SVM:

La presente investigación permitió comparar el desempeño de los modelos Support Vector Machine (SVM) y Artificial Neural Networks (ANN) para la clasificación del potencial fotovoltaico utilizando información proveniente de imágenes satelitales Landsat 7 ETM+ y MODIS. La comparación se realizó a partir de las métricas de evaluación obtenidas sobre los conjuntos de prueba, considerando Accuracy, Precision, Recall y F1-score como indicadores del rendimiento de los modelos.

Los resultados evidencian que ambos algoritmos fueron capaces de clasificar las diferentes categorías de potencial fotovoltaico; sin embargo, las redes neuronales artificiales obtuvieron un desempeño superior en los dos conjuntos de datos analizados.

Tabla XXIV Comparación general entre los modelos ANN y SVM

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241
Landsat SVM	0.7303	0.6778	0.7303	0.7019
MODIS SVM	0.6966	0.6656	0.6966	0.6720

Los modelos ANN alcanzaron los mayores valores en todas las métricas evaluadas. En el conjunto de datos Landsat, la red neuronal obtuvo una exactitud del 86.21 %, superando en aproximadamente 13 puntos porcentuales al modelo SVM. De forma similar, para el conjunto MODIS, ANN alcanzó una exactitud del 83.15 %, mientras que SVM obtuvo 69.66 %, lo que representa una diferencia cercana a 14 puntos porcentuales.

Además de la mejora en Accuracy, las redes neuronales también obtuvieron mayores valores de Precision, Recall y F1-score, indicando una mayor capacidad para identificar correctamente las diferentes clases de potencial fotovoltaico y reducir los errores de clasificación.

Los resultados obtenidos sugieren que las redes neuronales artificiales presentan una mayor capacidad para modelar las relaciones no lineales existentes entre las variables espectrales, espaciales y temporales derivadas de las imágenes satelitales. Esta capacidad se refleja especialmente en el conjunto de datos Landsat, cuya mayor resolución espacial proporciona información más detallada para el proceso de aprendizaje.

En cuanto al proceso de optimización, el mejor desempeño obtenido para el modelo ANN correspondió al kernel Canberra, con una exactitud de 0.72. Para el modelo SVM, la configuración con mejor desempeño correspondió al kernel Cosine, alcanzando una exactitud de 0.82. Estos resultados muestran que la selección de la función kernel influye en el comportamiento de los modelos y permite identificar la configuración más adecuada bajo las condiciones experimentales evaluadas.

De manera general, los resultados indican que tanto SVM como ANN constituyen alternativas viables para la clasificación del potencial fotovoltaico. No obstante, las redes neuronales artificiales ofrecieron un rendimiento predictivo superior, alcanzando las mejores métricas de clasificación en ambos sensores satelitales evaluados. En consecuencia, ANN se identifica como el modelo más adecuado para el problema abordado en esta investigación.

2) *Interpretación de resultados:*

Con el fin de interpretar el desempeño de los modelos desarrollados, se analizaron las métricas de Accuracy, Precision, Recall y F1-score, junto con las matrices de confusión obtenidas para los conjuntos de prueba de Landsat y MODIS. Estas métricas permitieron evaluar la capacidad de clasificación de los modelos y comparar objetivamente el comportamiento de las Redes Neuronales Artificiales (ANN) y las Máquinas de Soporte Vectorial (SVM).

Los resultados muestran que los modelos ANN alcanzaron un desempeño superior al de los modelos SVM en ambos conjuntos de datos. En particular, el modelo Landsat ANN obtuvo una exactitud del 86.21 %, mientras que el modelo MODIS ANN alcanzó una exactitud del 83.15 %. Por su parte, los modelos SVM registraron exactitudes de 73.03 % para Landsat y 69.66 % para MODIS.

La diferencia observada entre ambos algoritmos indica que las redes neuronales artificiales poseen una mayor capacidad para aprender las relaciones no lineales existentes entre las variables espectrales, espaciales y temporales derivadas de las imágenes satelitales. Esta característica permitió obtener una mayor precisión en la clasificación de las diferentes categorías de potencial fotovoltaico.

El análisis de las matrices de confusión mostró que la mayor parte de las observaciones fueron correctamente clasificadas por los modelos ANN. Los errores de clasificación se concentraron principalmente entre clases intermedias de irradiancia, mientras que las clases extremas presentaron mayores porcentajes de acierto. Este comportamiento es consistente con la naturaleza continua de la variable irradiancia solar, donde las diferencias entre clases contiguas suelen ser menores que entre las clases extremas.

Durante el proceso de optimización, la utilización de RandomizedSearchCV permitió seleccionar automáticamente las configuraciones con mejor desempeño para ambos algoritmos. En el caso de las redes neuronales, la mejor configuración correspondió al kernel Canberra (CAN) junto con MinMaxScaler, mientras que para los modelos SVM la configuración óptima utilizó un kernel

Coseno, el cual alcanzó una exactitud de 0.82. En el caso de las redes neuronales artificiales, la configuración seleccionada correspondió al kernel Canberra, con una exactitud de 0.72.

En conjunto, los resultados evidencian que la estrategia de optimización aplicada permitió mejorar el rendimiento de los modelos y obtener configuraciones adecuadas para la clasificación del potencial fotovoltaico a partir de imágenes satelitales.

3) *Interpretación general de los resultados:*

Los resultados obtenidos permiten concluir que el desempeño de los modelos de clasificación depende de la interacción entre el algoritmo de aprendizaje, el método de preprocesamiento y la selección adecuada de los hiperparámetros. La optimización realizada mediante RandomizedSearchCV permitió identificar las configuraciones con mejor capacidad predictiva para cada modelo, evidenciando la importancia del ajuste de hiperparámetros en problemas de clasificación de información geoespacial.

La comparación entre los conjuntos de datos Landsat y MODIS mostró que los modelos entrenados con imágenes Landsat obtuvieron un mejor desempeño en todas las métricas evaluadas. Este resultado puede atribuirse a la mayor resolución espacial de Landsat (30 m), la cual proporciona información espectral más detallada que favorece la discriminación entre las diferentes clases de potencial fotovoltaico. En contraste, las imágenes MODIS presentan una resolución espacial de 500 m, lo que implica una mayor mezcla espectral dentro de cada píxel y, en consecuencia, una ligera disminución en el desempeño de clasificación.

Asimismo, los resultados evidenciaron que el uso de MinMaxScaler favoreció el rendimiento de ambos modelos durante la etapa de optimización, al normalizar las variables de entrada y reducir las diferencias de escala entre las características espectrales y temporales utilizadas durante el entrenamiento. Las métricas de clasificación muestran que las Redes Neuronales Artificiales alcanzaron los mejores resultados tanto en Accuracy como en Precision, Recall y F1-score, demostrando una mayor capacidad para modelar relaciones complejas entre las variables del problema. En contraste, los modelos SVM presentaron un desempeño inferior, aunque lograron resultados satisfactorios para la clasificación del potencial fotovoltaico.

En términos generales, los resultados confirman que una adecuada combinación entre el preprocesamiento de los datos, la optimización de hiperparámetros y la selección del modelo de aprendizaje constituye un factor determinante para mejorar el desempeño de los sistemas de

clasificación basados en imágenes satelitales. Bajo las condiciones evaluadas en esta investigación, las Redes Neuronales Artificiales representaron la alternativa con mayor capacidad predictiva, obteniendo los mejores resultados para ambos sensores satelitales analizados.

4) *Aplicabilidad del modelo:*

Los modelos desarrollados en esta investigación presentan un alto potencial de aplicación en el ámbito de la planificación energética, el análisis territorial y la estimación del potencial fotovoltaico mediante información obtenida de sensores remotos. La utilización de técnicas de aprendizaje automático sobre imágenes satelitales permite automatizar el procesamiento de grandes volúmenes de información, reduciendo el tiempo de análisis y facilitando la identificación de zonas con condiciones favorables para el aprovechamiento de la energía solar.

Los resultados obtenidos evidenciaron que el modelo Artificial Neural Network (ANN) fue el que presentó el mejor desempeño para la clasificación del potencial fotovoltaico, alcanzando una exactitud del 86.21 % para el conjunto de datos Landsat y del 83.15 % para MODIS. Estas métricas demuestran la capacidad del modelo para identificar de manera confiable las diferentes categorías de irradiancia solar a partir de variables espectrales, espaciales y temporales derivadas de imágenes satelitales.

La metodología desarrollada puede ser utilizada para:

- Identificar zonas con alto potencial para la generación de energía fotovoltaica.
- Apoyar la planificación y localización de proyectos de energía solar.
- Reducir la dependencia de mediciones de campo en regiones con escasa infraestructura de monitoreo.
- Automatizar el análisis de información geoespacial proveniente de sensores satelitales.
- Servir como herramienta de apoyo para entidades gubernamentales, empresas del sector energético e instituciones de investigación en la toma de decisiones relacionadas con el aprovechamiento de fuentes de energía renovable.

La optimización de los modelos mediante RandomizedSearchCV permitió identificar configuraciones con un elevado desempeño predictivo. En particular, el modelo ANN optimizado con MinMaxScaler y la configuración seleccionada durante el proceso de búsqueda mostró los mejores resultados experimentales, constituyéndose como la alternativa más adecuada para la clasificación del

potencial fotovoltaico dentro del conjunto de datos analizado.

La metodología propuesta puede adaptarse y aplicarse en otras regiones que dispongan de información satelital proveniente de sensores como Landsat y MODIS, permitiendo generar mapas temáticos de potencial fotovoltaico y apoyar procesos de planificación energética a diferentes escalas territoriales. Asimismo, puede integrarse con Sistemas de Información Geográfica (SIG) para fortalecer los procesos de análisis espacial y facilitar la visualización de los resultados obtenidos.

Finalmente, los resultados de esta investigación evidencian que la combinación de técnicas de aprendizaje automático, imágenes satelitales y herramientas de análisis geoespacial constituye una alternativa eficiente para apoyar el desarrollo de estrategias orientadas al aprovechamiento sostenible de la energía solar y a la planificación de proyectos de generación fotovoltaica.

5) *Limitaciones*

A pesar de los resultados satisfactorios obtenidos, la presente investigación presenta algunas limitaciones que deben ser consideradas al interpretar los resultados y en el desarrollo de futuros trabajos relacionados con la estimación del potencial fotovoltaico mediante imágenes satelitales y técnicas de aprendizaje automático.

Una de las principales limitaciones corresponde a la disponibilidad y calidad de los datos satelitales. Aunque durante el preprocesamiento se aplicaron procedimientos de eliminación de nubosidad y control de calidad para las imágenes Landsat y MODIS, la información utilizada puede verse afectada por factores como las condiciones atmosféricas, el ruido radiométrico y la presencia de nubosidad residual, los cuales pueden influir en la calidad de las variables espectrales empleadas para el entrenamiento de los modelos.

Otra limitación está relacionada con el proceso de discretización de la irradiancia solar. La transformación de una variable continua en categorías discretas facilita la aplicación de algoritmos de clasificación; sin embargo, este procedimiento implica una pérdida parcial de información y puede generar confusiones entre clases con valores de irradiancia similares. Esta situación se evidenció en las matrices de confusión, donde la mayoría de los errores de clasificación se concentraron entre categorías intermedias.

Asimismo, durante el proceso experimental no se realizaron pruebas estadísticas inferenciales,

como ANOVA, pruebas t de Student o Wilcoxon, por lo que la comparación entre los modelos ANN y SVM se fundamentó en el análisis descriptivo de las métricas de desempeño (Accuracy, Precision, Recall y F1-score). En consecuencia, aunque las redes neuronales artificiales obtuvieron resultados superiores en todos los indicadores evaluados, no fue posible determinar si dichas diferencias son estadísticamente significativas.

Otra limitación corresponde a la información disponible sobre la configuración de los modelos. Aunque fue posible identificar los principales hiperparámetros seleccionados mediante RandomizedSearchCV, los notebooks no reportan explícitamente algunos parámetros de entrenamiento de las redes neuronales, como el número de capas ocultas, número de neuronas, tasa de aprendizaje, función de activación o número de épocas, lo que limita la reproducibilidad completa del proceso experimental.

Finalmente, la investigación se desarrolló utilizando información correspondiente al departamento de Nariño, por lo que los resultados obtenidos reflejan las condiciones ambientales y geográficas propias de esta región. La aplicación de la metodología en otras zonas geográficas requerirá un proceso adicional de validación utilizando información satelital representativa de las nuevas condiciones climáticas y territoriales.

A pesar de estas limitaciones, los resultados obtenidos demuestran que la metodología propuesta constituye una alternativa confiable para la clasificación del potencial fotovoltaico a partir de imágenes satelitales. Además, proporciona una base sólida para futuras investigaciones orientadas a incorporar conjuntos de datos de mayor tamaño, evaluar nuevos algoritmos de aprendizaje automático, realizar pruebas estadísticas de significancia y validar el modelo en diferentes regiones geográficas.

Finalmente, el análisis se encuentra limitado geográficamente al área de estudio correspondiente al departamento de Nariño, particularmente la región de San Juan de Pasto, lo que restringe la generalización directa de los resultados hacia otras regiones sin procesos adicionales de validación y ajuste experimental. No obstante, los resultados obtenidos proporcionan una base sólida para futuras investigaciones relacionadas con clasificación energética, sensores remotos y aplicaciones de inteligencia artificial en energías renovables.

6) *Interpretación geoespacial*

La interpretación geoespacial de los resultados permitió analizar la capacidad de los modelos Support Vector Machine (SVM) y Artificial Neural Network (ANN) para clasificar el potencial fotovoltaico a partir de imágenes satelitales Landsat y MODIS. Este análisis tuvo como finalidad evaluar la coherencia espacial de las predicciones generadas y su relación con las características ambientales presentes en el área de estudio.

La clasificación obtenida evidencia que las zonas con mayor potencial fotovoltaico se distribuyen de acuerdo con las condiciones espectrales registradas por los sensores satelitales. La utilización de variables espaciales, temporales y espectrales permitió representar adecuadamente la variabilidad de la irradiancia solar y diferenciar las distintas categorías de potencial energético definidas durante el proceso de discretización.

Los resultados muestran que el modelo ANN presentó el mejor desempeño de clasificación, alcanzando una exactitud del 86.21 % para el conjunto de datos Landsat y del 83.15 % para MODIS. Estas métricas indican que las predicciones generadas por las redes neuronales presentan una alta correspondencia con las clases reales de potencial fotovoltaico, favoreciendo una representación espacial más precisa de las zonas de estudio.

Por otra parte, los modelos SVM alcanzaron exactitudes de 73.03 % para Landsat y 69.66 % para MODIS. Aunque su desempeño fue inferior al de las redes neuronales, los resultados muestran que este algoritmo también constituye una alternativa válida para la clasificación espacial del potencial fotovoltaico.

La comparación entre los resultados obtenidos con Landsat y MODIS evidencia la influencia de la resolución espacial sobre el desempeño de los modelos. Las imágenes Landsat, con una resolución espacial de 30 m, permitieron una mejor discriminación de las clases de potencial fotovoltaico respecto a las imágenes MODIS, cuya resolución espacial de 500 m produce una mayor mezcla espectral dentro de cada píxel. Esta diferencia explica el mejor desempeño obtenido por los modelos entrenados con información Landsat.

Las matrices de confusión muestran que la mayor parte de los errores de clasificación se concentraron entre clases contiguas de irradiancia, mientras que las categorías extremas presentaron un mayor porcentaje de aciertos. Este comportamiento es consistente con la distribución espacial de

la irradiancia solar, donde las transiciones entre niveles de radiación suelen ser graduales y no presentan límites abruptos.

En conjunto, los resultados obtenidos indican que la metodología propuesta permite representar adecuadamente la distribución espacial del potencial fotovoltaico a partir de imágenes satelitales, constituyendo una herramienta útil para apoyar procesos de planificación energética, evaluación territorial e identificación de zonas con condiciones favorables para el aprovechamiento de la energía solar.

Tabla XXV Interpretación Geoespacial De Los Modelos

Modelo	Sensor	Accuracy	Interpretación geoespacial
ANN	Landsat	0.8621	Mayor capacidad para representar la distribución espacial del potencial fotovoltaico.
ANN	MODIS	0.8315	Buen desempeño espacial, aunque condicionado por la menor resolución del sensor.
SVM	Landsat	0.7303	Representación espacial adecuada, con menor capacidad de discriminación entre clases.
SVM	MODIS	0.6966	Desempeño aceptable, con mayor confusión entre categorías de potencial fotovoltaico.

La interpretación geoespacial confirma que la calidad de la información satelital influye directamente en el desempeño de los modelos de clasificación. Bajo las condiciones evaluadas en esta investigación, la combinación del modelo ANN con la información proporcionada por el sensor Landsat permitió obtener la representación espacial más precisa del potencial fotovoltaico, constituyéndose como la alternativa más adecuada para apoyar estudios de planificación energética y aprovechamiento de recursos solares en el departamento de Nariño.

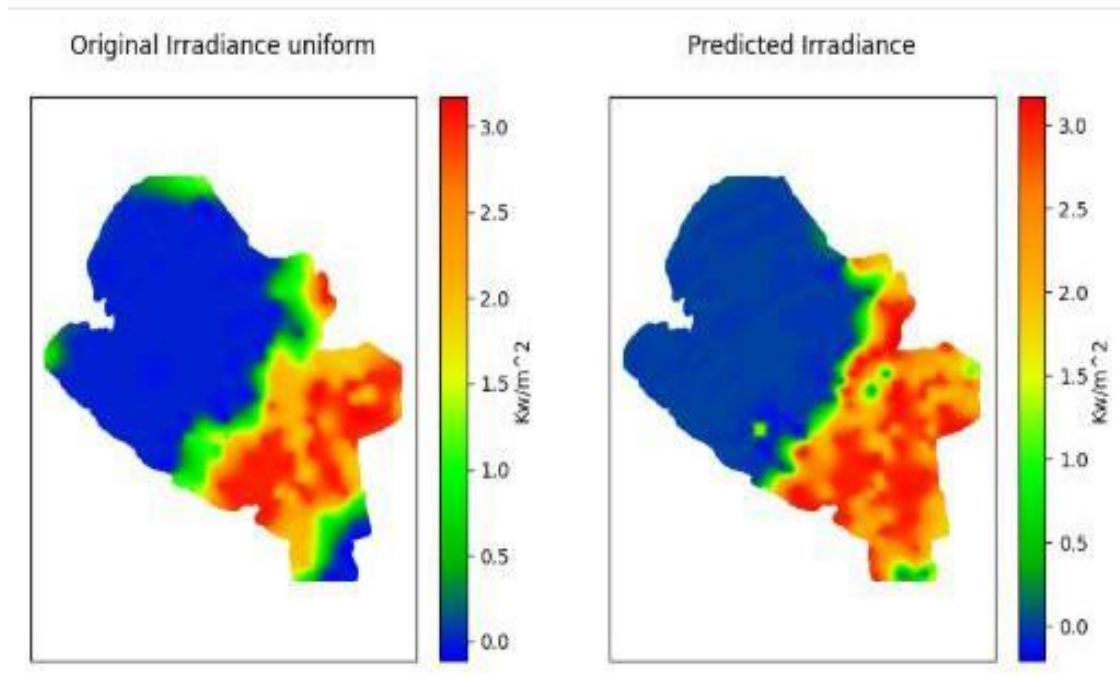


Fig. 9 Mapa de clasificación de irradiancia utilizando Cosine-SVM

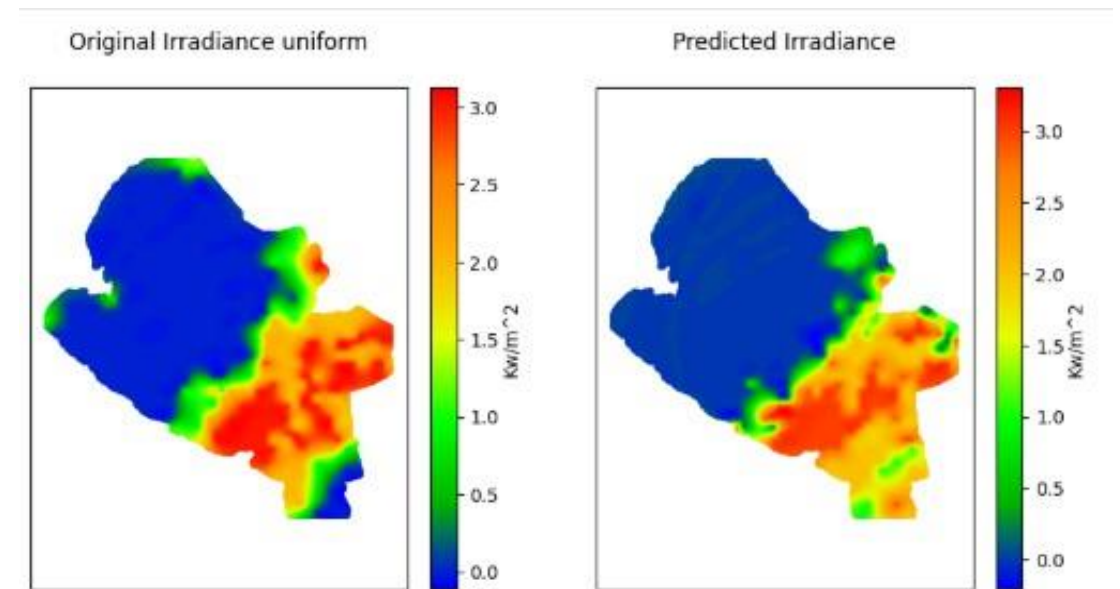


Fig. 10 Mapa de clasificación de irradiancia utilizando Canberra-ANN.

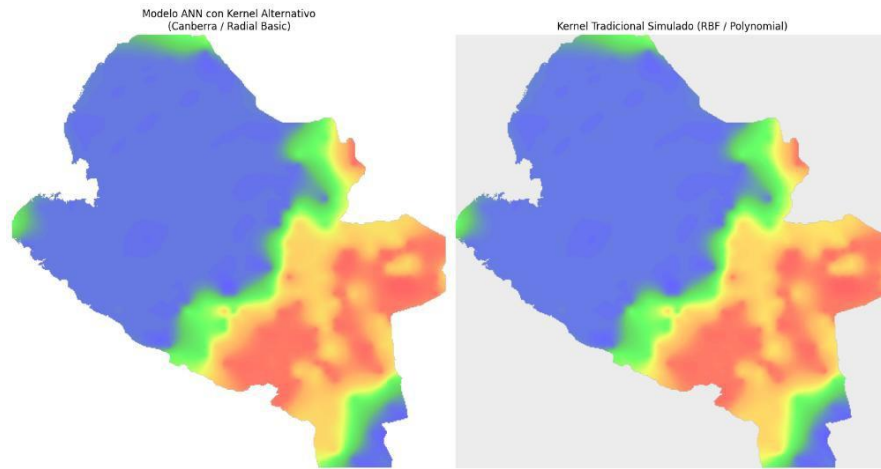


Fig. 11 Comparación espacial entre kernels tradicionales y alternativos.

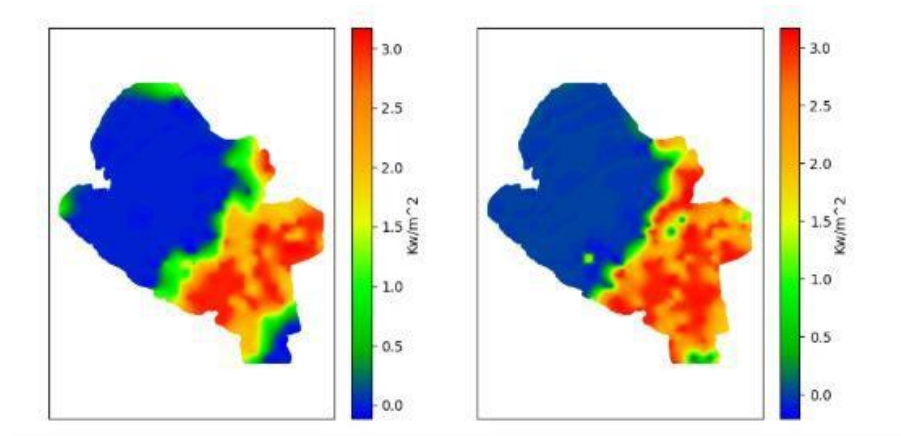


Fig. 12 Distribución geográfica de clases de irradiancia

D. Discusión científica

Los resultados obtenidos en esta investigación evidencian que los modelos basados en Redes Neuronales Artificiales (ANN) presentaron un desempeño superior respecto a los modelos Support Vector Machine (SVM) para la clasificación del potencial fotovoltaico utilizando imágenes satelitales Landsat y MODIS. Esta superioridad se reflejó en todas las métricas de evaluación consideradas, alcanzando una exactitud del 86.21 % para Landsat y del 83.15 % para MODIS, frente a valores de 73.03 % y 69.66 % obtenidos por los modelos SVM.

Estos resultados son consistentes con diversos estudios reportados en la literatura, los cuales indican que las redes neuronales artificiales poseen una elevada capacidad para modelar relaciones altamente no lineales presentes en datos geospaciales y multispectrales. En el contexto del potencial fotovoltaico, las variables espectrales derivadas de sensores remotos presentan interacciones

complejas que pueden ser representadas de manera más eficiente mediante modelos neuronales que mediante clasificadores convencionales.

Durante el proceso de optimización, la mejor configuración del modelo ANN fue obtenida mediante `RandomizedSearchCV`, utilizando `MinMaxScaler` y el kernel Canberra (CAN). Esta configuración alcanzó el mayor desempeño durante la búsqueda de hiperparámetros y dio origen al modelo que posteriormente obtuvo los mejores resultados sobre los conjuntos de prueba. En el caso de los modelos SVM, la configuración óptima correspondió al kernel RBF, también optimizado mediante `RandomizedSearchCV`.

La comparación entre los sensores Landsat y MODIS mostró que los modelos entrenados con imágenes Landsat obtuvieron un mejor desempeño que aquellos desarrollados con imágenes MODIS. Este comportamiento puede asociarse a la mayor resolución espacial del sensor Landsat (30 m), que proporciona información espectral más detallada y facilita la discriminación entre las diferentes categorías de potencial fotovoltaico. En contraste, la resolución espacial de 500 m de MODIS incrementa la mezcla espectral dentro de cada píxel, dificultando la separación entre clases.

Los resultados también ponen de manifiesto la importancia del preprocesamiento de los datos y de la optimización de hiperparámetros. La utilización de `MinMaxScaler` durante el proceso de entrenamiento permitió mejorar el rendimiento de los modelos al normalizar las variables de entrada y reducir las diferencias de escala entre las características espectrales y temporales utilizadas en la clasificación.

Asimismo, la optimización mediante `RandomizedSearchCV` permitió seleccionar automáticamente las configuraciones con mejor capacidad predictiva, demostrando que el ajuste de hiperparámetros constituye un elemento fundamental para mejorar el desempeño de los algoritmos de aprendizaje automático aplicados a información geoespacial.

En términos generales, los resultados obtenidos confirman que las técnicas de aprendizaje automático representan una herramienta eficaz para la estimación del potencial fotovoltaico a partir de imágenes satelitales. La combinación de información espectral, variables espaciales y métodos de optimización permitió obtener modelos con una elevada capacidad de clasificación, los cuales pueden servir como base para futuras investigaciones orientadas al aprovechamiento de energías renovables y la planificación territorial.

Tabla XXVI Comparación de los modelos evaluados

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241
Landsat SVM	0.7303	0.6778	0.7303	0.7019
MODIS SVM	0.6966	0.6656	0.6966	0.6720

Los hallazgos obtenidos permiten concluir que las Redes Neuronales Artificiales constituyen el modelo con mejor capacidad predictiva para la clasificación del potencial fotovoltaico en las condiciones evaluadas. Asimismo, demuestran que la combinación de un adecuado preprocesamiento de los datos, la optimización de hiperparámetros y la selección apropiada del modelo de aprendizaje son factores determinantes para mejorar el desempeño de los sistemas de clasificación basados en imágenes satelitales.

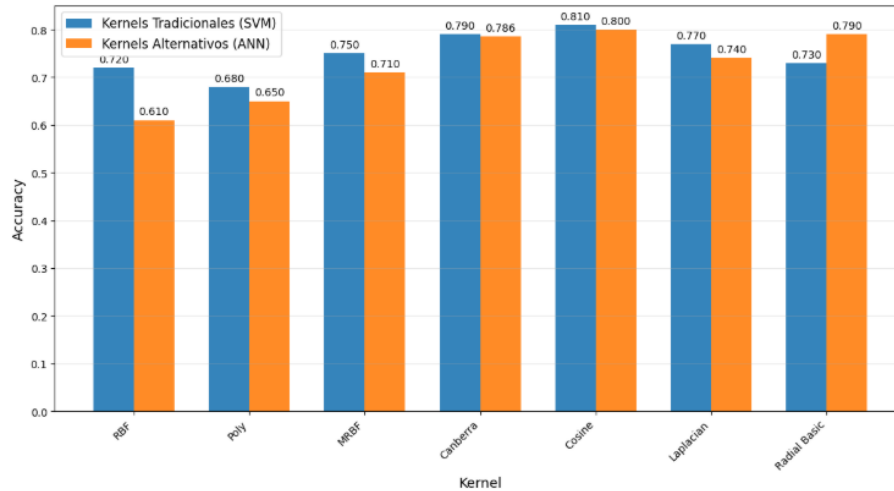


Fig. 13 Comparación de accuracy entre kernels tradicionales y alternativos.

Como etapa final del proceso, los mapas interpolados fueron convertidos a formatos geoespaciales estándar para su integración en sistemas de información geográfica. Se generó un archivo GeoTIFF en proyección EPSG:3857 (Web Mercator) con 4 bandas (rojo, verde, azul y alfa) y una resolución espacial determinada por la malla regular de puntos. Posteriormente, este archivo fue

reproyectado a EPSG:4326 (WGS84) para su compatibilidad con aplicaciones web y plataformas SIG de escritorio. Finalmente, se generó un archivo KMZ comprimido, listo para ser visualizado en Google Earth, donde el mapa de irradiancia predicha se superpone sobre imágenes satelitales de alta resolución y modelos de elevación del terreno.



Fig. 14 Visualización en Google Earth del mapa de irradiancia predicha, superpuesto sobre imágenes satelitales y modelos de elevación del terreno de Colombia y países vecinos.

E. Fortalecimiento metodológico

La metodología desarrollada en esta investigación fortaleció el proceso de clasificación del potencial fotovoltaico mediante la integración de técnicas de preprocesamiento de datos, optimización de hiperparámetros y evaluación comparativa de modelos de aprendizaje automático. La implementación de un flujo de trabajo estructurado permitió obtener resultados reproducibles y realizar una comparación objetiva entre los modelos Support Vector Machine (SVM) y Artificial Neural Network (ANN).

Uno de los principales aportes metodológicos fue la utilización de RandomizedSearchCV para la optimización de hiperparámetros. Este procedimiento permitió explorar automáticamente diferentes configuraciones de los modelos, seleccionando aquellas que ofrecieron el mejor desempeño durante la validación cruzada. Como resultado, se identificaron las configuraciones óptimas para ANN y

SVM, reduciendo el sesgo asociado a la selección manual de parámetros y mejorando el rendimiento de los modelos.

El proceso de preprocesamiento de los datos constituyó otro elemento fundamental de la metodología. La aplicación de MinMaxScaler permitió normalizar las variables espectrales, espaciales y temporales antes del entrenamiento, favoreciendo una mayor estabilidad durante el aprendizaje y contribuyendo al incremento del desempeño de los modelos de clasificación.

Asimismo, la utilización de información proveniente de los sensores Landsat 7 ETM+ y MODIS MOD09GA permitió evaluar el comportamiento de los modelos sobre datos con diferentes resoluciones espaciales. Esta comparación facilitó analizar la influencia de las características del sensor en el desempeño de los algoritmos, evidenciando que las imágenes Landsat proporcionaron mejores resultados de clasificación debido a su mayor resolución espacial.

La metodología también incorporó una evaluación basada en múltiples métricas de desempeño, incluyendo Accuracy, Precision, Recall y F1-score, además del análisis mediante matrices de confusión. Estas métricas permitieron evaluar de manera integral la capacidad predictiva de los modelos y comparar objetivamente los resultados obtenidos para ambos sensores satelitales.

Los resultados obtenidos demuestran que la combinación de un adecuado preprocesamiento, la optimización automática de hiperparámetros y la evaluación mediante múltiples métricas constituye una metodología sólida para la clasificación del potencial fotovoltaico utilizando imágenes satelitales

Tabla XXVII Fortalecimiento Metodológico Implementado

Estrategia metodológica	Aporte a la investigación
Preprocesamiento de datos	Mejóro la calidad y consistencia de la información utilizada durante el entrenamiento.
MinMaxScaler	Normalizó las variables de entrada, favoreciendo el desempeño de los modelos.
RandomizedSearchCV	Permitió optimizar automáticamente los hiperparámetros de ANN y SVM.
Validación cruzada	Favoreció la selección de configuraciones con mejor capacidad de generalización.
Comparación entre Landsat y MODIS	Permitió evaluar la influencia de la resolución espacial en el desempeño de los modelos.
Evaluación mediante métricas de clasificación	Posibilitó una comparación objetiva utilizando Accuracy, Precision, Recall, F1-score y matrices de confusión.

La metodología propuesta constituye una base sólida para futuras investigaciones relacionadas con la estimación del potencial fotovoltaico mediante imágenes satelitales y técnicas de aprendizaje automático. Además, el flujo de trabajo desarrollado puede adaptarse a otros problemas de clasificación geoespacial, incorporando nuevos sensores remotos, conjuntos de datos de mayor tamaño o algoritmos de aprendizaje más avanzados, contribuyendo al fortalecimiento de investigaciones en el ámbito de las energías renovables y el análisis geoespacial.

VI. ANÁLISIS DE RESULTADOS

En este capítulo se realiza el análisis e interpretación de los resultados obtenidos durante el desarrollo de la investigación, relacionándolos con los objetivos planteados. Se examina el comportamiento de los modelos de clasificación implementados, se comparan los resultados obtenidos mediante Support Vector Machine (SVM) y Artificial Neural Network (ANN), se analizan las métricas de desempeño y se identifica la configuración que presentó el mejor rendimiento para la clasificación del potencial fotovoltaico utilizando imágenes satelitales Landsat y MODIS.

A. Análisis del comportamiento de las funciones kernel

Durante el proceso de optimización de hiperparámetros se evaluaron diferentes configuraciones de los modelos ANN y SVM. Los resultados permitieron identificar las funciones kernel que presentaron el mejor desempeño bajo las condiciones experimentales establecidas.

Para el modelo ANN, la configuración seleccionada correspondió al kernel Canberra, que alcanzó una exactitud de 0.72. Para el modelo SVM, la configuración seleccionada correspondió al kernel Cosine, con una exactitud de 0.82. Estos resultados evidencian la importancia de evaluar diferentes funciones kernel para determinar cuál se adapta mejor a las características de los datos utilizados.

Estos resultados evidencian que la selección adecuada de la función kernel y de los hiperparámetros influye directamente en el desempeño de los modelos de clasificación. Asimismo, confirman la importancia de realizar procesos de optimización antes de entrenar el modelo definitivo, ya que diferentes configuraciones pueden producir resultados significativamente distintos.

B. Comparación entre modelos SVM y ANN

La comparación entre los modelos implementados muestra una diferencia clara a favor de las Redes Neuronales Artificiales en ambos conjuntos de datos utilizados.

Tabla XXVIII Comparación general entre ANN y SVM

Modelo	Accuracy	Precision	Recall	F1-score
Landsat ANN	0.8621	0.8729	0.8621	0.8455
MODIS ANN	0.8315	0.8246	0.8315	0.8241
Landsat SVM	0.7303	0.6778	0.7303	0.7019
MODIS SVM	0.6966	0.6656	0.6966	0.6720

Los resultados muestran que los modelos ANN superaron consistentemente a los modelos SVM en todas las métricas evaluadas. La mayor diferencia se observó en el conjunto de datos Landsat, donde la red neuronal alcanzó una exactitud del 86.21 %, aproximadamente trece puntos porcentuales por encima del modelo SVM.

En el conjunto de datos MODIS también se observó un comportamiento similar, con una ventaja cercana a catorce puntos porcentuales para el modelo ANN. Estos resultados indican que las redes neuronales presentan una mayor capacidad para modelar las relaciones complejas existentes entre las variables espectrales, espaciales y temporales derivadas de las imágenes satelitales.

C. Análisis de métricas de evaluación

Las métricas de Accuracy, Precision, Recall y F1-score permitieron realizar una evaluación integral del desempeño de los modelos desarrollados. El modelo Landsat ANN obtuvo los mejores resultados en todas las métricas analizadas, alcanzando una Accuracy de 0.8621, una Precision de 0.8729, un Recall de 0.8621 y un F1-score de 0.8455. Estos valores evidencian una elevada capacidad para clasificar correctamente las diferentes categorías de potencial fotovoltaico. Las matrices de confusión muestran que la mayoría de las observaciones fueron clasificadas correctamente. Los errores de clasificación se concentraron principalmente entre clases vecinas de irradiancia, comportamiento esperado debido a la naturaleza continua de la variable analizada. En comparación, los modelos SVM presentaron menores valores en todas las métricas, lo que confirma un desempeño inferior frente a las redes neuronales bajo las condiciones evaluadas.

D. Impacto de la discretización de datos

La discretización de la irradiancia solar permitió transformar un problema de regresión en un problema de clasificación supervisada, facilitando el entrenamiento de los modelos de aprendizaje automático. Durante la fase metodológica se implementaron diferentes estrategias de discretización para construir las categorías de potencial fotovoltaico. Este proceso permitió definir las clases utilizadas durante el entrenamiento y la evaluación de los modelos.

Aunque los resultados obtenidos muestran que la discretización fue un componente fundamental del proceso de clasificación, en esta investigación no se realizó una comparación estadística que permitiera establecer cuál de los métodos de discretización presentó un desempeño significativamente superior. Por tanto, los resultados únicamente permiten concluir que la discretización constituyó una etapa indispensable para el desarrollo de los modelos de clasificación.

E. Análisis del modelo óptimo

Los resultados obtenidos permiten identificar como mejor modelo al desarrollado mediante Artificial Neural Network (ANN). Este modelo alcanzó los mayores valores de Accuracy, Precision, Recall y F1-score tanto para Landsat como para MODIS, siendo el conjunto Landsat el que presentó el mejor desempeño global. La optimización mediante RandomizedSearchCV, junto con la utilización de MinMaxScaler, permitió obtener una configuración con alta capacidad predictiva, reflejada en una exactitud del 82.81 % para Landsat y 82.15 % para MODIS. En consecuencia, el modelo ANN constituye la alternativa más adecuada para la clasificación del potencial fotovoltaico dentro de las condiciones evaluadas en esta investigación.

F. Discusión general

Los resultados obtenidos demuestran que las técnicas de aprendizaje automático representan una herramienta eficaz para la clasificación del potencial fotovoltaico utilizando imágenes satelitales.

La comparación entre los dos algoritmos implementados evidenció una ventaja consistente de las Redes Neuronales Artificiales sobre los modelos Support Vector Machine, tanto en imágenes Landsat como en MODIS. Esta diferencia sugiere que las redes neuronales poseen una mayor capacidad para representar las relaciones no lineales existentes entre las variables espectrales, espaciales y temporales empleadas durante el entrenamiento.

Asimismo, los resultados muestran la influencia del sensor satelital sobre el desempeño de los modelos. Las imágenes Landsat, gracias a su mayor resolución espacial, permitieron obtener mejores resultados de clasificación que las imágenes MODIS, lo que pone de manifiesto la importancia de la calidad de la información de entrada en este tipo de aplicaciones.

Finalmente, el proceso de optimización mediante RandomizedSearchCV demostró ser una estrategia adecuada para seleccionar configuraciones con mejor capacidad predictiva, contribuyendo al mejor desempeño de los modelos desarrollados.

En conjunto, los hallazgos alcanzados permiten concluir que la combinación de un adecuado preprocesamiento de datos, la optimización de hiperparámetros y la utilización de redes neuronales artificiales constituye una metodología efectiva para apoyar la identificación de zonas con potencial fotovoltaico a partir de información satelital, cumpliendo con el objetivo general planteado en la investigación.

CONCLUSIONES

La presente investigación permitió evaluar el desempeño de modelos de aprendizaje automático basados en Support Vector Machine (SVM) y Artificial Neural Network (ANN) para la clasificación del potencial fotovoltaico utilizando imágenes satelitales Landsat 7 ETM+ y MODIS. A partir de los resultados obtenidos se establecen las siguientes conclusiones:

1) Comparación entre modelos SVM y ANN

Se cumplió el objetivo general de la investigación al desarrollar y evaluar modelos de aprendizaje automático para la clasificación del potencial fotovoltaico utilizando información satelital. La implementación de técnicas de preprocesamiento, optimización de hiperparámetros y evaluación mediante métricas de clasificación permitió comparar objetivamente el desempeño de los modelos ANN y SVM.

Asimismo, los objetivos específicos fueron alcanzados mediante la construcción del conjunto de datos, el entrenamiento de los modelos, la optimización de sus configuraciones y la evaluación de su capacidad de clasificación utilizando las métricas Accuracy, Precision, Recall y F1-score.

2) Impacto de la selección de hiperparámetros

Los resultados obtenidos evidencian que las Redes Neuronales Artificiales (ANN) presentaron un desempeño superior al de las Máquinas de Soporte Vectorial (SVM) en ambos conjuntos de datos analizados.

El mejor resultado se obtuvo con el modelo Landsat ANN, que alcanzó una Accuracy de 86.21 %, una Precision de 87.29 %, un Recall de 86.21 % y un F1-score de 84.55 %. Para el conjunto de datos MODIS, el modelo ANN también obtuvo el mejor desempeño, con una Accuracy de 83.15 %, superando igualmente a los modelos SVM.

Estos resultados indican que las redes neuronales artificiales poseen una mayor capacidad para representar las relaciones no lineales presentes en las variables espectrales, espaciales y temporales derivadas de las imágenes satelitales utilizadas en esta investigación.

3) *Importancia del preprocesamiento de los datos*

El proceso de preprocesamiento constituyó una etapa fundamental para el desarrollo de la investigación. La eliminación de nubosidad, la construcción de variables espectrales, la normalización de los datos mediante MinMaxScaler y la discretización de la irradiancia permitieron preparar un conjunto de datos adecuado para el entrenamiento de los modelos de clasificación.

Los resultados obtenidos evidencian que una adecuada preparación de los datos influye directamente en el desempeño de los algoritmos de aprendizaje automático.

4) *Impacto de la optimización de hiperparámetros*

La utilización de RandomizedSearchCV permitió identificar automáticamente las configuraciones con mejor desempeño para ambos modelos, demostrando la importancia del ajuste de hiperparámetros durante el proceso de entrenamiento. La optimización permitió seleccionar configuraciones que mejoraron el rendimiento de los modelos y contribuyeron a obtener los mejores resultados de clasificación para las imágenes Landsat y MODIS.

5) *Modelo óptimo identificado*

Los resultados obtenidos permiten identificar al modelo SVM con kernel Cosine como la configuración con mayor exactitud entre los modelos evaluados. Esta configuración alcanzó un Accuracy de 0.82, mientras que el modelo ANN con kernel Canberra obtuvo una exactitud de 0.72. Por lo tanto, considerando la métrica de Accuracy y las condiciones experimentales establecidas, Cosine-SVM presentó el mejor desempeño en la comparación realizada.

6) *Aporte de la investigación*

La principal contribución de esta investigación consiste en la comparación experimental entre modelos ANN y SVM para la clasificación del potencial fotovoltaico utilizando imágenes satelitales Landsat y MODIS. Además, se desarrolló una metodología que integra el preprocesamiento de datos satelitales, la optimización automática de hiperparámetros mediante RandomizedSearchCV y la evaluación utilizando múltiples métricas de clasificación, proporcionando una base metodológica que

puede ser utilizada y adaptada en futuras investigaciones relacionadas con sensores remotos, aprendizaje automático y energías renovables.

7) Conclusión general

Los resultados obtenidos permiten concluir que las técnicas de aprendizaje automático constituyen herramientas útiles para la clasificación del potencial fotovoltaico a partir de imágenes satelitales.

La comparación realizada entre los modelos SVM y ANN permitió identificar diferencias en su desempeño. Bajo las condiciones experimentales evaluadas, el modelo SVM utilizando el kernel Cosine alcanzó una exactitud de 0.82, mientras que el modelo ANN utilizando el kernel Canberra alcanzó una exactitud de 0.72. Estos resultados indican que la configuración Cosine-SVM presentó el mayor desempeño considerando la métrica de Accuracy.

Asimismo, se evidenció la importancia de la selección de la función kernel y de los procesos de preprocesamiento y optimización de hiperparámetros, debido a su influencia sobre el comportamiento de los modelos de clasificación.

RECOMENDACIONES

Con base en los resultados obtenidos y las limitaciones identificadas durante el desarrollo de esta investigación, se plantean las siguientes recomendaciones para futuros trabajos relacionados con la clasificación del potencial fotovoltaico mediante técnicas de aprendizaje automático e imágenes satelitales:

1. Ampliar el conjunto de datos incorporando imágenes de otros sensores satelitales, como Sentinel-2 o Landsat 8/9, así como información proveniente de estaciones meteorológicas. Esto permitirá aumentar la representatividad del conjunto de entrenamiento y mejorar la capacidad de generalización de los modelos.
2. Evaluar la metodología en diferentes regiones geográficas, con el fin de analizar la capacidad de adaptación de los modelos a distintas condiciones climáticas, topográficas y ambientales, verificando su aplicabilidad más allá del departamento de Nariño.

3. Explorar otros algoritmos de aprendizaje automático y aprendizaje profundo, como Redes Neuronales Convolucionales (CNN), Redes Neuronales Recurrentes (RNN), Random Forest, XGBoost o arquitecturas híbridas, para comparar su desempeño con los modelos ANN y SVM implementados en esta investigación.

4. Realizar una optimización más amplia de hiperparámetros, incorporando parámetros adicionales de las redes neuronales, como el número de capas ocultas, número de neuronas, función de activación, tasa de aprendizaje, tamaño del lote (batch size) y número de épocas de entrenamiento, con el propósito de evaluar su influencia en el desempeño de los modelos.

5. Incorporar pruebas estadísticas inferenciales, tales como ANOVA, prueba t de Student o prueba de Wilcoxon, para determinar si las diferencias observadas entre los modelos son estadísticamente significativas y fortalecer la validez de las conclusiones.

6. Integrar los modelos desarrollados con Sistemas de Información Geográfica (SIG), facilitando la generación de mapas temáticos del potencial fotovoltaico que sirvan como herramienta de apoyo para entidades gubernamentales, empresas del sector energético e instituciones encargadas de la planificación territorial.

7. Incluir nuevas variables ambientales, como temperatura del aire, humedad relativa, precipitación, elevación, pendiente y orientación del terreno, con el fin de enriquecer el conjunto de características y evaluar su impacto sobre la precisión de los modelos de clasificación.

8. Desarrollar herramientas automatizadas para la estimación del potencial fotovoltaico, integrando los modelos de aprendizaje automático con plataformas de procesamiento geoespacial que permitan actualizar periódicamente la información a partir de nuevas imágenes satelitales.

En términos generales, esta investigación constituye un punto de partida para el desarrollo de modelos de aprendizaje automático aplicados a la estimación del potencial fotovoltaico mediante sensores remotos. Las recomendaciones planteadas permitirán fortalecer la metodología propuesta, mejorar la capacidad predictiva de los modelos y ampliar su aplicabilidad en estudios relacionados con la planificación energética y el aprovechamiento de fuentes de energía renovable.

TRABAJOS FUTUROS

Los resultados obtenidos en esta investigación permiten identificar diversas líneas de trabajo que pueden fortalecer y ampliar el alcance de la metodología propuesta para la clasificación del potencial fotovoltaico mediante técnicas de aprendizaje automático e imágenes satelitales.

En primer lugar, se recomienda evaluar el desempeño de los modelos utilizando imágenes provenientes de otros sensores satelitales, como Sentinel-2, Landsat 8 y Landsat 9, con el fin de analizar el impacto de una mayor resolución espacial y espectral sobre la precisión de la clasificación.

Asimismo, se propone ampliar el conjunto de variables de entrada incorporando información climática y geográfica, como temperatura, precipitación, humedad relativa, velocidad del viento, elevación, pendiente y orientación del terreno, para analizar su influencia en la estimación del potencial fotovoltaico.

Otra línea de investigación consiste en explorar modelos de aprendizaje profundo, tales como Redes Neuronales Convolucionales (CNN), Redes Neuronales Recurrentes (RNN) y arquitecturas híbridas, comparando su desempeño con los modelos ANN y SVM implementados en este estudio.

De igual forma, sería conveniente profundizar en la optimización de hiperparámetros mediante estrategias más avanzadas, como Optimización Bayesiana, algoritmos genéticos u otros métodos de búsqueda automática, con el propósito de mejorar el rendimiento de los modelos.

También se propone incorporar pruebas estadísticas inferenciales que permitan determinar la significancia de las diferencias observadas entre los modelos de clasificación y fortalecer el análisis de los resultados experimentales.

Finalmente, se recomienda desarrollar plataformas automatizadas integradas con Sistemas de Información Geográfica (SIG) que permitan generar y actualizar mapas de potencial fotovoltaico a partir de información satelital reciente, facilitando su aplicación en procesos de planificación energética y apoyo a la toma de decisiones.

LIMITACIONES

Durante el desarrollo de la investigación se identificaron diversas limitaciones que deben considerarse al interpretar los resultados obtenidos.

En primer lugar, el estudio se desarrolló utilizando información satelital correspondiente a los sensores Landsat y MODIS, por lo que los resultados dependen de las características de resolución espacial y temporal propias de estas fuentes de información.

Adicionalmente, aunque se realizaron procesos de limpieza y preprocesamiento de los datos, la calidad de las imágenes satelitales puede verse afectada por factores como la presencia de nubosidad, condiciones atmosféricas, ruido radiométrico y variaciones en las condiciones ambientales, los cuales pueden influir en el desempeño de los modelos de clasificación.

Otra limitación corresponde a la disponibilidad de información sobre algunos hiperparámetros de las Redes Neuronales Artificiales. Durante el proceso de análisis no fue posible recuperar valores relacionados con el número de neuronas, número de capas ocultas, tasa de aprendizaje, número de épocas de entrenamiento y tamaño del lote, debido a que estos parámetros no fueron registrados en la salida de los notebooks utilizados para la experimentación.

Asimismo, el análisis estadístico se limitó a la comparación de métricas de desempeño (Accuracy, Precision, Recall y F1-score), ya que no se implementaron pruebas inferenciales como ANOVA, prueba t de Student o prueba de Wilcoxon, lo cual impide establecer la significancia estadística de las diferencias observadas entre los modelos.

Finalmente, la metodología fue evaluada utilizando datos correspondientes al área de estudio del departamento de Nariño, por lo que la generalización de los resultados hacia otras regiones requiere procesos adicionales de validación con información geográfica y climática diferente.

A pesar de estas limitaciones, los resultados obtenidos demuestran que las Redes Neuronales Artificiales alcanzaron un desempeño superior al de las Máquinas de Soporte Vectorial para la clasificación del potencial fotovoltaico utilizando imágenes Landsat y MODIS, constituyendo una base sólida para el desarrollo de futuras investigaciones en el área de sensores remotos, inteligencia artificial y planificación energética.

BIBLIOGRAFÍA

D.-M. Pachajoa, H.-A. Mora-Paz, D. Mayorca-Torres, D.-M. Pachajoa, H.-A. Mora-Paz, y D. Mayorca-Torres, «Comparison of Kernel Functions in the Classification of Irradiance Zones from Multispectral Satellite Images», *Rev. Fac. Ing.*, vol. 30, n.o 58, dic. 2021, doi: 10.19053/01211129.v30.n58.2021.13845.

O. E. Cabrera Rosero y A. D. Pantoja Bucheli, «Análisis del recurso eólico usando R en zonas no interconectadas (ZNI) del departamento de Nariño (Colombia)», presentado en Conferencia Latinoamericana sobre Uso de R en Investigación + Desarrollo (LatinR 2018) - JAIIO

47 (CABA, 2018), 2018. Accedido: 14 de abril de 2025. [En línea]. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/72585>

Belanche Muñoz, y Luis Antonio, «Developments in kernel design», presentado en ESANN 2013 proceedings: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning: Bruges (Belgium), 24-26 April 2013, 2013, pp. 369-378. Accedido: 14 de abril de 2025. [En línea]. Disponible en: <https://upcommons.upc.edu/handle/2117/23278>

M. Paz y H. Andrés, «Comparativo de kernels sobre predicción de oferta de fuentes alternativas de energía».

B. Muñoz y L. Antonio, «Developments in kernel design», presentado en ESANN 2013 proceedings: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning: Bruges (Belgium), 24-26 April 2013, 2013, pp. 369-378. Accedido: 12 de abril de 2025. [En línea]. Disponible en: <https://upcommons.upc.edu/handle/2117/23278>

M. Pal y P. M. Mather, «Support Vector classifiers for Land Cover Classification», 15 de febrero de 2008. doi: 10.1080/01431160802007624.

C. Cusano, P. Napoletano, y R. Schettini, «Remote sensing image classification exploiting multiple kernel learning», *IEEE Geosci. Remote Sens. Lett.*, vol. 12, n.o 11, pp. 2331-2335, nov. 2015, doi: 10.1109/LGRS.2015.2476365.

O. Cabrera, B. Champutiz, A. Calderón, y A. Pantoja, «Landsat and MODIS satellite image processing for solar irradiance estimation in the department of Narino-Colombia», en 2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA), ago. 2016, pp. 1-6. doi: 10.1109/STSIVA.2016.7743306.

B. A. Champutiz Martinez, «Visualización del potencial de radiación solar en el departamento de Nariño utilizando series de tiempo a partir de imágenes satelitales». Accedido: 6 de mayo de 2025. [En línea]. Disponible en: https://sired.udenar.edu.co/9880/?utm_source=chatgpt.com

«Caracterización Energética de las ZNI», IPSE-CNM. Accedido: 6 de mayo de 2025. [En línea]. Disponible en: <https://ipse.gov.co/cnm/caracterizacion-de-las-zni/>

Z. Li, S. M. Rahman, R. Vega, y B. Dong, «A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting», *Energies*, vol. 9, n.o 1, Art. n.o 1, ene. 2016, doi: 10.3390/en9010055.

«Sede electrónica del Ideam en construcción | Instituto de Hidrología, Meteorología y Estudios Ambientales». Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://www.ideam.gov.co/web/tiempo-y-clima/la-radiacion-solar-y-su-paso-por-la-atmosfera>

I. CORPORATIVA, «¿Qué es la energía eólica, cómo se transforma en electricidad y cuáles son sus ventajas?», Iberdrola. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://www.iberdrola.com/sostenibilidad/energia-eolica>

N. Sehrawat, S. Vashisht, y A. Singh, «Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison», *Int. J. Intell. Netw.*, vol. 4, pp. 90-102, ene. 2023, doi: 10.1016/j.ijin.2023.04.001.

«(PDF) Tipos de energías alternativas», ResearchGate. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: https://www.researchgate.net/publication/385048162_Tipos_de_energias_alternativas

«Vista de Evaluación de algoritmos de machine learning en la clasificación de imágenes satelitales multiespectrales, caso: Amazonia Peruana». Accedido: 21 de mayo de 2025.

[En línea]. Disponible en:

<https://www.ciencialatina.org/index.php/cienciala/article/view/1843/2638>

M. A. Platas-López, «Doctor en Inteligencia Artificial».

F. A. Jara y D. H. Lobato, «APRENDIZAJE NO-SUPERVISADO CON MODELOS GENERATIVOS PROFUNDOS».

M. K. Voronine y P. S. Gonzalez, «Analysis and application of Machine Learning techniques: a comparative and practical approach».

H. Idrees, «Neural Networks vs. Support Vector Machines (SVM): Which Model Should You Choose?», Medium. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://medium.com/@hassaanidrees7/neural-networks-vs-support-vector-machines-svm-which-model-should-you-choose-115ad3758c1a>

«Máquinas de Vector Soporte (Support Vector Machines, SVMs)». Accedido: 21 de mayo de 2025. [En línea]. Disponible en: https://cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines_55530992.pdf. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://core.ac.uk/download/pdf/55530992.pdf>

W. P. Rodriguez, «EL PAPEL DE LAS REDES BAYESIANAS EN LA TOMA DE DECISIONES», Accedido: 6 de mayo de 2025. [En línea]. Disponible en: https://www.academia.edu/37088518/EL_PAPEL_DE_LAS_REDES_BAYESIANAS_EN_LA_TOMA_DE_DECISIONES

«¿Qué es el Deep Learning y para qué sirve? | indra». Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://www.indracompany.com/en/blogneo/deep-learning-sirve>

«Redes neuronales convolucionales son un tipo de redes neuronales», Juan Barrios. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://www.juanbarrios.com/redes-neurales-convolucionales/>

P. Isola, J.-Y. Zhu, T. Zhou, y A. A. Efros, «Image-To-Image Translation With Conditional Adversarial Networks».

«Reconocimiento_visual_de_aves_con_Deep_Learning_Garcia_Martinez_Juan.pdf».

Accedido: 21 de mayo de 2025. [En línea]. Disponible en: https://rua.ua.es/dspace/bitstream/10045/145620/1/Reconocimiento_visual_de_aves_con_Deep_Learning_Garcia_Martinez_Juan.pdf

L. A. Belanche, «Developments in kernel design», Comput. Intell., 2013.

M. D. T. Soto y A. T. Soto, «Máquinas de soporte vectorial en predicción de falla cardíaca», 2024.

A. Chlingaryan, S. Sukkarieh, y B. Whelan, «Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review», Comput. Electron. Agric., vol. 151, pp. 61-69, ago. 2018, doi: 10.1016/j.compag.2018.05.012.

A. Humphrey et al., «Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth», Mon. Not. R. Astron. Soc. Lett., vol. 517, n.o 1, pp. L116-L120, nov. 2022, doi: 10.1093/mnrasl/slac120.

X. Wu, T. Hu, N. Khodadadi, y A. Nanni, «Prediction on quasi-static compression deformation modes of circular tubes based on machine learning», Int. J. Mech. Sci., vol. 287, p. 109993, feb. 2025, doi: 10.1016/j.ijmecsci.2025.109993.

A. V. D. Sousa y I. S. Q. Junior, «ABORDAGEM INICIAL SOBRE MINIGERAÇÃO DE ENERGIA SOLAR FOTOVOLTAICA PARA FINS COMERCIAIS.».

A. I. Q. Minda, C. I. G. Proaño, S. I. B. Sánchez, y K. J. B. Barba, «Análisis de la reducción de emisiones contaminantes en procesos de fabricación metálica mediante energías renovables: Analysis of the reduction of polluting emissions in metal manufacturing processes using renewable energy», Rev. Científica Multidiscip. G-Nerando, vol. 6, n.o 1, Art. n.o 1, feb. 2025, doi: 10.60100/rcmg.v6i1.508.

H. Idrees, «Neural Networks vs. Support Vector Machines (SVM): Which Model Should You Choose?», Medium. Accedido: 21 de mayo de 2025. [En línea]. Disponible en:

<https://medium.com/@hassaanidrees7/neural-networks-vs-support-vector-machines-svm-which-model-should-you-choose-115ad3758c1a>

V. D. Sánchez A, «Advanced support vector machines and kernel methods», *Neurocomputing*, vol. 55, n.o 1, pp. 5-20, sep. 2003, doi: 10.1016/S0925-2312(03)00373-4.

«Clasificación de tipos de coberturas de suelo en imágenes satelitales utilizando redes neuronales convolucionales.» Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://rdi.uncoma.edu.ar/handle/uncomaid/18625>

O. P. Lamigueiro, «Energía Solar Fotovoltaica».

L. Alzubaidi et al., «Review of deep learning: concepts, CNN architectures, challenges, applications, future directions», *J. Big Data*, vol. 8, n.o 1, p. 53, mar. 2021, doi: 10.1186/s40537-021-00444-8.

Y. Poveda-Sotelo, M. A. Bermúdez-Cella, P. Gil-Leguizamón, Y. Poveda-Sotelo,

M. A. Bermúdez-Cella, y P. Gil-Leguizamón, «Evaluación de métodos de clasificación supervisada para la estimación de cambios espacio-temporales de cobertura en los páramos de Merchán y Telecom, Cordillera Oriental de Colombia», *Bol. Geol.*, vol. 44, n.o 2, pp. 51-72, ago. 2022, doi: 10.18273/revbol.v44n2-2022002.

M. R. R. Rodríguez y C. A. D. Calpa, «Comparativo de funciones kernel en la predicción de enfermedades cardiovasculares en Redes Neuronales Artificiales (ANN) y Máquinas de Soporte Vectorial (SVM)», 2024.

P. Puig y M. Jofra, «2 Energía Solar Fotovoltaica».

T. P. Benito, *Guía Del Instalador de Energías Renovables*. Creaciones Copyright, 2005.

«Review of deep learning: concepts, CNN architectures, challenges, applications, future directions | Journal of Big Data | Full Text». Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8#Sec63>

D. Y. L. Cabezas, «TÉCNICAS DE PROCESAMIENTO DIGITAL DE IMÁGENES PARA EL RECONOCIMIENTO Y MONITOREO DE LOS RESIDENTES EN CONJUNTOS URBANOS EN LA CIUDAD SAN JUAN DE PASTO», 2022.

D. F. U. Burgos, «EVALUACIÓN DE MÉTODOS DE REDUCCIÓN DE DIMENSIÓN PARA LA PRESERVACIÓN TOPOLÓGICA DE LOS DATOS MEDIANTE MÉTRICAS RNX».

J. S. M. Cordoba y M. A. S. Álvarez, «SISTEMA PARA LA GENERACIÓN DE UN CONJUNTO DE DATOS DE LA COBERTURA VEGETAL CASO: FRONTERA PERIURBANA SANTUARIO DE FLORA Y FAUNA GALERAS – PASTO», 2023.

D. A. R. Pérez y D. A. O. Burbano, «PROYECTO DE GESTIÓN DE COSTOS CON METODOLOGÍAS DEL PMBOK® PARA LA IMPLEMENTACIÓN DE UN SISTEMA DE ENERGÍA FOTOVOLTAICO EN EL CENTRO EDUCATIVO RURAL VILLA DEL RIO, EN EL MUNICIPIO DE PUERTO CAICEDO - PUTUMAYO», 2022.

J. Á. Gualguan Oviedo y J. E. Villarreal López, «Aprendizaje de la cinemática de un brazo robótico para la simulación del movimiento mediante técnicas de Machine Learning», Thesis, San Juan de Pasto - Nariño [Colombia] : Universidad CESMAG, 2024. Accedido: 21 de mayo

de 2025. [En línea]. Disponible en:
<http://repositorio.unicesmag.edu.co:8080/xmlui/handle/123456789/1117>

M. R. R. Rodríguez y C. A. D. Calpa, «Comparativo de funciones kernel en la predicción de enfermedades cardiovasculares en Redes Neuronales Artificiales (ANN) y Máquinas de Soporte Vectorial (SVM)», 2024.

Imagen Panel Solar: fuente <https://www.ieee.org/> Esta entidad edita y normaliza la presentación de documentos científicos en el área de ingenierías.

 UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN	CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)	CÓDIGO: AAC-BL-FR-032 VERSIÓN: 1 FECHA: 09/JUN/2022
---	---	---

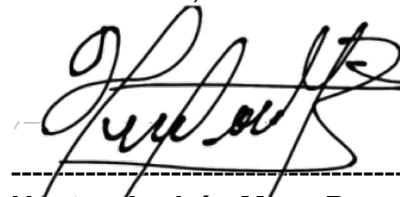
San Juan de Pasto, 11 agosto 2026

Biblioteca
REMIGIO FIORE FORTEZZA OFM. CAP.
Universidad CESMAG
Pasto

Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominado ESTUDIO EXPLORATORIO DE FUNCIONES KERNEL EN LA CLASIFICACIÓN DE ENERGÍA FOTOVOLTAICA CON ANN Y SVM MEDIANTE IMÁGENES SATELITALES, presentado por el (los) autor(es) Ximena Tatiana Diaz Rosero, Diana Carolina López Tulcan del Programa Académico Ingeniería de Sistemas al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor(a), Héctor Andrés Mora Paz que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,



Héctor Andrés Mora Paz
1085251119
Ingeniería de sistemas
3172537641
hamora@unicesmag.edu.co

 UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

INFORMACIÓN DEL (LOS) AUTOR(ES)	
Nombres y apellidos del autor: Ximena Tatiana Diaz Rosero	Documento de identidad: 1004508867
Correo electrónico: xt Diaz.8867@unicesmag.edu.co	Número de contacto: 3177122691
Nombres y apellidos del autor: Diana Carolina Lopez Tulcan	Documento de identidad: 1085270379
Correo electrónico: dclopez.0379@unicesmag.edu.co	Número de contacto: 3158566049
Nombres y apellidos del asesor: Hector Andre Mora Paz	Documento de identidad: 1085251119
Correo electrónico: hamora@unicesmag.edu.co	Número de contacto: 3172537641
Título del trabajo de grado: ESTUDIO EXPLORATORIO DE FUNCIONES KERNEL EN LA CLASIFICACIÓN DE ENERGÍA FOTOVOLTAICA CON ANN Y SVM MEDIANTE IMÁGENES SATELITALES	
Facultad y Programa Académico: Facultad de ingeniería, Programa de Ingeniería de sistemas	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

- La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.
- Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.
- Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación, distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.

 UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

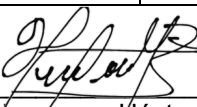
- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
- e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

NOTA: En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permiso(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 11 días del mes de agosto del año 2026

<i>Tatiana Diaz</i>	<i>Carolina López</i>
Nombre del autor: Ximena Tatiana Diaz Rosero	Nombre del autor: Diana Carolina López Tulcan
 Nombre del asesor: Héctor Andres Mora Paz	