

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Estudio comparativo de algoritmos de forecasting en la predicción de irradiancia
mediante imágenes satelitales multiespectrales

Laura Sofia Gomez Rosero y Ingridt Carolina Ordoñez Lopez

Programa de ingeniería de sistemas, facultad de ingeniería, 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 sistemas de la Universidad CESMAG. Correo electrónico:
ingenieriadesistemas@unicesmag.edu.co

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Estudio comparativo de algoritmos de forecasting en la predicción de irradiancia
mediante imágenes satelitales multiespectrales

Laura Sofia Gomez Rosero y Ingridt Carolina Ordoñez Lopez

Ingeniería de sistemas, facultad de ingeniería, universidad CESMAG

Asesor. Mg. Héctor Andrés Mora Paz

06 de agosto de 2026

Nota de aceptación

“Va la aceptación definida desde la universidad”.

Nota de exclusión

DEDICATORIA

Dedico el presente proyecto de grado a mis padres, Jaime Ordoñez y Cecilia Narvaez quienes, con su amor incondicional, esfuerzo y sacrificio han sido el pilar fundamental en mi formación. Gracias por su apoyo constante, por sus enseñanzas y por brindarme los valores que me han permitido llegar hasta este momento. Este logro también les pertenece.

A mi hermano, que hoy me acompaña desde el cielo, cuya memoria ha sido fuente de inspiración y fortaleza en cada paso de este camino. Aunque su presencia física no esté, su amor y recuerdo viven en mi corazón y han sido un impulso para no rendirme.

A mi pareja, por su comprensión, paciencia y acompañamiento durante este proceso. Gracias por su apoyo incondicional, por motivarme en los momentos difíciles y por celebrar conmigo cada avance alcanzado. Su presencia ha sido esencial para culminar esta etapa con éxito

Carolina O.

Dedico este proyecto a mis padres, Juan Carlos Gómez y Amanda Rosero, por ser mi guía constante, mi refugio y la fuerza que impulsa cada uno de mis logros. Agradezco profundamente su apoyo incondicional, cada esfuerzo realizado, su amor infinito y cada sacrificio silencioso que han hecho por mí. Gracias por nunca soltar mi mano, por creer en mí incluso en los momentos más desafiantes y por acompañarme con paciencia y fortaleza en cada paso de mi vida. Para ustedes, que han sido mi mayor inspiración y el mejor ejemplo, dedico este logro con todo mi corazón. Son y serán siempre el pilar más importante en mi camino.

Laura G.

AGRADECIMIENTOS

Expresamos nuestro más sincero agradecimiento a Dios, por brindarnos la vida, la fortaleza, la sabiduría y la perseverancia necesarias para culminar con éxito esta importante etapa de nuestra formación académica.

En segundo lugar, expresamos nuestro reconocimiento a nuestros docentes Hector Mora y Dalila Pachajoa, por su orientación, acompañamiento académico y valiosa asesoría técnica, los cuales contribuyeron de manera significativa al desarrollo y culminación de este trabajo de grado.

De igual manera, manifestamos nuestro profundo agradecimiento a nuestros padres, por su apoyo incondicional, su esfuerzo constante y por ser el pilar fundamental en nuestro desarrollo personal y profesional.

Finalmente, expresamos nuestro agradecimiento mutuo, valorando el compromiso, la colaboración y la disposición constante de cada una, aspectos que fueron clave para el cumplimiento de los objetivos planteados.

TABLA DE CONTENIDO

	Pag
INTRODUCCIÓN	14
I. PROBLEMA DE INVESTIGACIÓN.....	15
A. OBJETO O TEMA DE ESTUDIO	15
B. LÍNEA DE INVESTIGACIÓN.....	15
C. PLANTEAMIENTO DEL PROBLEMA.....	15
D. FORMULACIÓN DEL PROBLEMA	16
E. OBJETIVOS	16
1) Objetivo general.....	16
2) Objetivos específicos.....	16
F. JUSTIFICACIÓN.....	17
G. DELIMITACIÓN	17
II. MARCO TEÓRICO.....	19
A. ANTECEDENTES.....	19
B. SUPUESTOS TEÓRICOS.....	20
C. VARIABLES DEL ESTUDIO	27
D. DEFINICIÓN NOMINAL DE LAS VARIABLES	27
E. DEFINICIÓN NOMINAL DE LAS VARIABLES	29
F. FORMULACIÓN DE HIPÓTESIS	34
III. METODOLOGÍA.....	35
A. PARADIGMA	35
B. ENFOQUE	35
C. MÉTODO	35
D. TIPO DE INVESTIGACIÓN	35
E. DISEÑO DE LA INVESTIGACIÓN	35
F. POBLACIÓN.....	36
G. MUESTRA.....	37
H. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN	37

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

K. INSTRUMENTO DE RECOLECCIÓN DE DATOS.....	38
L. METODOLOGÍA DE DESARROLLO.....	39
IV. RESULTADOS DE LA INVESTIGACIÓN	49
V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS	64
CONCLUSIONES	67
RECOMENDACIONES	68
BIBLIOGRAFÍA	69
ANEXOS	74

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

LISTA DE FIGURAS

	Pag
FIG. 1.ENERGÍAS LIMPIAS.....	20
FIG. 2. SATÉLITE.....	21
FIG. 3. IMÁGENES SATELITALES	22
FIG. 4. . IMAGEN SATELITAL MULTIESPECTRAL.....	23
FIG. 5. IMAGEN SENSORES SATÉLITES	24
FIG. 7. ALGORITMO DE APRENDIZAJE.....	25
FIG. 8. PROCESO	27
FIG. 9. ALGORITMO	30
FIG. 10. REGRESIÓN Y CLASIFICACIÓN.....	31
FIG. 11. ADQUISICIÓN DE IMÁGENES	39
FIG. 12. MER DE LAS TABLAS DE REFLECTANCIA E IRRADIANCIA EN LAS BASES DE DATOS LANDSAT Y MODIS	40
FIG. 13. CONJUNTO CARACTERÍSTICO	42
FIG. 14. CONJUNTOS CARACTERÍSTICOS POR BANDAS	45
FIG. 15. MAPA DE NARIÑO.....	46
FIG. 16. ARQUITECTURA DEL MODELO DE RED NEURONAL.....	46
FIG. 17. REGISTROS OBTENIDOS DEL DATASET INICIAL.....	51
FIG. 18.FECHAS GENERADAS PARA LA SERIE TEMPORAL.....	52
FIG. 19. VARIOGRAMAS APLICADOS A LAS MUESTRAS DE IRRADIANCIA	53
FIG. 20. REPRESENTACIÓN VISUAL DEL PROCESO DE INTERPOLACIÓN.....	54
FIG. 21. REPRESENTACIÓN CADA BANDA DEL ESPECTRO ELECTROMAGNÉTICO	54
FIG. 22. ARCHIVOS CSV GENERADOS PARA CADA FECHA DEL PERIODO DE ESTUDIO	55
FIG. 23. . GRÁFICA DEL COMPORTAMIENTO DEL MODELO.....	57
FIG. 24. MAPA DE NARIÑO.....	57
FIG. 25.MAPA DE CALOR DE NARIÑO	58
FIG. 26. IMAGEN DE ALGORITMO DE INTERPOLACIÓN	59
FIG. 27.MAPA DE LA IRRADIANCIA PREDICHA PARA NARIÑO EN FORMATO RASTER EN FORMATO PNG	59
FIG. 28. MAPA DE LA IRRADIANCIA PREDICHA PARA NARIÑO EN FORMATO RASTER .TIF PROYECCIONES DE MERCATOR 3857 EN METROS VISUALIZADA EN LA APLICACIÓN GEOTIFF.IO.....	60
FIG. 29. MAPA DE LA IRRADIANCIA PREDICHA PARA NARIÑO EN FORMATO RASTER .KMZ EN PROYECCIONES DE MERCATOR 4326 EN ÁNGULOS VISUALIZADA EN LA APLICACIÓN GOOGLE EARTH	60
FIG. 30. MAPA DE LA IRRADIANCIA PREDICHA PARA NARIÑO EN FORMATO RASTER .TIF EN PROYECCIONES DE MERCATOR 4326 EN ÁNGULOS VISUALIZADA EN LA APLICACIÓN WEB CON API DE GOOGLE MAPS.....	61
FIG. 31. GRAFICA COMPORTAMIENTO DE R^2	62
FIG. 32. GRAFICA COMPORTAMIENTO MSE	63
FIG. 33.GRÁFICA COMPARATIVA DEL R^2 Y MSE	63

LISTA DE TABLAS

TABLA I. METODOLOGÍA PRELIMINAR.....	36
TABLA II. DESCRIPCIÓN DEL CONJUNTO DE DATOS.....	49
TABLA III. SEGMENTOS TEMPORALES.....	49
TABLA IV. DATASETS INICIALES	50
TABLA V. CONJUNTO DE DATOS	52
TABLA VI. CONJUNTO DE DATOS.....	55
TABLA VII. SUBCONJUNTO DE DATOS	55
TABLA VIII. VENTANAS TEMPORALES.....	56
TABLA IX. COMPARACIÓN DE MÉTRICAS DE CALIDAD	56
TABLA X. EXPERIMENTOS REALIZADOS	62
TABLA XI. EJEMPLO DE CONSULTAS SQL PARA OBTENCIÓN DE SEGMENTOS TEMPORALES DE LANDSAT.....	74
TABLA XII. MODELOS UTILIZADOS PARA LA GENERACIÓN DE LAS BANDAS ESPECTRALES.....	74

LISTA DE ECUACIONES

	Pag
ECUACIÓN 1. SVM	30
ECUACIÓN 2. CLASIFICACIÓN	30
ECUACIÓN 3. REGRESIÓN	30
ECUACIÓN 4. KNN	31
ECUACIÓN 5. FORGET GATE	31
ECUACIÓN 6. INPUT GATE	32
ECUACIÓN 7. CELL STATE	32
ECUACIÓN 8. OUTPUT GATE.....	32
ECUACIÓN 9. UPDATE GATE.....	32
ECUACIÓN 10. RESET GATE.....	32
ECUACIÓN 11. NUEVO ESTADO OCULTO	32
ECUACIÓN 12. ACTUALIZACIÓN ESTADO OCULTO	32
ECUACIÓN 13. NORMALIZACIÓN	33
ECUACIÓN 14. ESTANDARIZACIÓN	33
ECUACIÓN 15. ONE HOT	33
ECUACIÓN 16. IMPUTACIÓN.....	33
ECUACIÓN 17. PARTICIÓN DE LA REFLECTANCIA POR RANGO DE FECHAS	41
ECUACIÓN 18. PARTICIÓN DE LA IRRADIANCIA POR RANGO DE FECHAS.....	41
ECUACIÓN 19. CRUCE DE LOS SEGMENTOS DE RADIACIÓN E IRRADIANCIA.	41
ECUACIÓN 20. TRANSFORMACIÓN DE LOS MESES QUE CONFORMAN A LA FECHA DE CAPTURA.	43
ECUACIÓN 21. TRANSFORMACIÓN DE LOS DÍAS QUE CONFORMAN A LA FECHA DE CAPTURA.	43
ECUACIÓN 22. TRANSFORMACIÓN DE LOS AÑOS QUE CONFORMAN A LA FECHA DE CAPTURA	43
ECUACIÓN 23. EXPRESIÓN GENERAL USADA EN MODELADO MATEMÁTICO / MACHINE LEARNING.....	43

GLOSARIO

Este glosario reúne y define las palabras clave que se emplean en esta investigación. Este estudio, que se ubica en la intersección de la inteligencia artificial, el procesamiento de imágenes y la energía solar, incorpora definiciones de algoritmos de aprendizaje automático, métodos para procesar imágenes satelitales, criterios para evaluar modelos y términos asociados a la irradiancia solar. Siendo una ayuda para aquellos que están interesados en examinar y comparar modelos de predicción utilizados en energías renovables.

Imágenes satelitales: Imágenes visuales de la superficie del planeta que son captadas por sensores montados en satélites. Estas imágenes tienen la capacidad de cubrir distintas bandas del espectro electromagnético, incluyendo la infrarroja, las microondas y la visible. Esto posibilita su utilización en meteorología, monitoreo ambiental, predicción de irradiancia solar y agricultura[1].

Multiespectrales: característica de los datos adquiridos en varias bandas del espectro electromagnético, lo que posibilita un análisis minucioso de las propiedades físicas y químicas de los elementos estudiados. En el ámbito de las imágenes satelitales, estas bandas pueden contener datos infrarrojos, visibles y de microondas, que se emplean en investigaciones de clima y medio ambiente[1].

Irradiancia: Es la cantidad de energía solar que impacta sobre un área específica en un periodo de tiempo determinado, medida en vatios por metro cuadrado (W/m^2). Es un parámetro esencial para la evaluación del potencial de producción de energía fotovoltaica.

Transformación de datos: Conjunto de métodos que se emplean para transformar datos de un formato a otro con el propósito de optimizar su calidad, simplificar su análisis o adecuarlos a modelos de aprendizaje automático. Incorpora procedimientos como la normalización, la estandarización, el reemplazo de escalas y la eliminación de valores anómalos, los cuales son esenciales para mejorar el desempeño de los algoritmos predictivos[1].

Predicción de series de tiempo: Método de análisis que se emplea para predecir valores futuros basándose en patrones pasados de datos secuenciales. Se utiliza en diferentes áreas, como la meteorología, la economía y la previsión de oferta solar energética. Se lleva a cabo en el aprendizaje automático por medio de modelos autorregresivos y redes neuronales recurrentes (RNN).

RESUMEN

El incremento de la demanda energética y los efectos del cambio climático han impulsado el desarrollo de fuentes renovables, especialmente la energía solar fotovoltaica. Aunque la radiación solar representa la principal fuente energética del sistema climático terrestre, factores como las emisiones de gases de efecto invernadero intensifican el calentamiento global [1]. lo que hace necesario optimizar el aprovechamiento de energías limpias. Debido a las limitaciones y altos costos asociados a las redes tradicionales de medición, la estimación de la irradiancia solar mediante imágenes satelitales y técnicas de Machine Learning se ha convertido en una alternativa eficiente para el análisis espacio-temporal del recurso solar.

Investigaciones previas han utilizado imágenes provenientes de los satélites Landsat y MODIS para construir bases de datos geospaciales y desarrollar modelos supervisados, incluyendo redes neuronales artificiales y máquinas de soporte vectorial, obteniendo resultados satisfactorios [2]. En este contexto, el presente estudio analiza imágenes satelitales multiespectrales para el departamento de Nariño, donde evalúa el desempeño de distintos modelos de aprendizaje automático en la estimación de irradiancia solar y a partir de esos resultados, construye mapas especializados orientados a mejorar la precisión predictiva y apoyar futuras investigaciones sobre el aprovechamiento de la energía solar en la región.

Palabras Clave: Irradiancia, imágenes satelitales multiespectrales, redes neuronales convolucionales, cambio climático, fuentes renovables.

ABSTRACT

The increase in energy demand and the effects of climate change have driven the development of renewable energy sources, especially solar photovoltaic energy. Although solar radiation represents the main energy source of the Earth's climate system, factors such as greenhouse gas emissions intensify global warming [1], making it necessary to optimize the use of clean energy sources. Due to the limitations and high costs associated with traditional measurement networks, the estimation of solar irradiance through satellite imagery and Machine Learning techniques has become an efficient alternative for the spatiotemporal analysis of solar resources.

Previous research has used images from Landsat and MODIS satellites to build geospatial databases and develop supervised models, including artificial neural networks and support vector machines, achieving satisfactory results [2]. In this context, the present study analyzes multispectral satellite images for the department of Nariño, evaluating the performance of different machine learning models in the estimation of solar irradiance. Based on these results, specialized maps are generated to improve predictive accuracy and support future research on the use of solar energy in the region.

Keywords: Irradiance, multispectral satellite images, convolutional neural networks, climate change, renewable energy sources.

INTRODUCCIÓN

El cambio climático es una realidad que afecta cada vez más nuestra vida diaria. Si la temperatura del planeta aumenta más de 2 °C, podríamos enfrentar consecuencias graves como la subida del nivel del mar, incendios forestales, escasez de alimentos, menos acceso a agua potable e incluso conflictos sociales [3]. Una de las formas más efectivas de enfrentar este problema es cambiando nuestras fuentes de energía hacia opciones limpias, como la solar, la eólica o la geotérmica [3]. Diversos estudios advierten que, sin acciones estructuradas y sostenidas para reducir las emisiones, los impactos climáticos tenderán a intensificarse en las próximas décadas.

Para aprovechar bien la energía solar, es necesario saber en qué lugares se recibe más radiación del sol. Sin embargo, instalar estaciones de medición en todo el territorio es muy costoso. En departamentos como Nariño, por ejemplo, se necesitarían más de 500 estaciones para tener datos precisos [4]. Por eso, es importante buscar nuevas formas de estimar esta información sin depender solo de estaciones físicas. Además, la planificación energética requiere información con alta resolución espacial, lo que incrementa aún más los costos de infraestructura y mantenimiento de redes de monitoreo tradicionales.

Aquí es donde entran las imágenes satelitales y los algoritmos de Machine Learning (ML). A través de imágenes tomadas por satélites como Landsat o MODIS, es posible observar cómo se refleja la luz solar en la superficie terrestre. Estas imágenes, combinadas con modelos de aprendizaje automático, permiten estimar cuánta energía solar llega a cada zona. Estos enfoques permiten generar estimaciones en zonas donde no existen mediciones directas, mediante el aprendizaje de patrones entre variables históricas, reflectancia y condiciones atmosféricas. Asimismo, el análisis multiespectral aporta mayor riqueza de información para mejorar la capacidad predictiva de los modelos. En otros países ya se han hecho avances usando estas técnicas [5], pero en Colombia y especialmente en Nariño no se habían explorado su uso con suficiente profundidad.

Este proyecto buscó aportar en esa línea de investigación. Se enfocó en evaluar si el uso de imágenes satelitales de alta resolución mejoraba las predicciones de irradiancia solar. Asimismo, se compararon diferentes modelos de inteligencia artificial, como redes neuronales y técnicas de Forecasting, para determinar cuáles ofrecían un mejor desempeño con este tipo de datos. También se consideró la necesidad de analizar el equilibrio entre precisión de los modelos y costo computacional, con el fin de proponer soluciones técnicamente viables y escalables. El objetivo fue desarrollar herramientas que permitieran estimar la generación de energía solar en Nariño con mayor precisión y a menor costo, promoviendo así el uso de energías limpias en la región.

I. PROBLEMA DE INVESTIGACIÓN

A. *OBJETO O TEMA DE ESTUDIO*

Esta investigación se llevó a cabo en el departamento de Nariño y tuvo como propósito optimizar la predicción de la irradiancia solar mediante el uso de imágenes satelitales multiespectrales y modelos de aprendizaje automático [2]. El enfoque se centró en evaluar el desempeño de diferentes algoritmos de Forecasting, analizando tanto la precisión de sus estimaciones como la eficiencia en su procesamiento. Para ello, se utilizaron datos meteorológicos e imágenes históricas. El proyecto se desarrolló entre 2025 y mediados de 2026, y se espera que una vez finalizado e implementado apoye la toma de decisiones en sectores como la agricultura, la gestión ambiental y la planificación energética [1].

B. *LÍNEA DE INVESTIGACIÓN*

Gestión, desarrollo e innovación en las Tecnologías de la Información y la Comunicación.

Sub línea de investigación

Inteligencia Artificial

C. *PLANTEAMIENTO DEL PROBLEMA*

El clima terrestre ha experimentado transformaciones mundiales durante su evolución, y esta variabilidad se ve afectada tanto por elementos internos como externos. Sin embargo, la época industrial, se caracteriza por un significativo incremento en la temperatura mundial. Este aumento ha causado la concentración de gases en la atmósfera, incidiendo de manera considerable en la temperatura del planeta [6]. Cada país es responsable de determinar su contribución a la respuesta mundial al cambio climático, y tomar medidas de mitigación, adaptación y cooperación [7]. Estas medidas están propuestas en el Acuerdo de París donde se destaca principalmente la disminución del uso de energías fósiles y nucleares.

Por otro lado, está la captación de energía que se alinea a políticas de oferta energética que van atadas a una política estatal, para ello es necesario determinar en qué zonas es más viable colocar estaciones con receptores energéticos teniendo en cuenta el impacto social, ecológico y por supuesto la oferta energética de la zona [1]. Además, en el contexto colombiano, la dependencia de fuentes fósiles es limitada, lo que hace necesario fomentar el uso de energías renovables.

El cambio climático y la necesidad de energías renovables han impulsado la búsqueda de modelos más precisos para predecir la irradiancia solar, ya que esta predicción es fundamental para la eficiencia y distribución de los sistemas fotovoltaicos. Sin embargo, uno de los retos fue que no se conocía el equilibrio entre las métricas de calidad y el tiempo de

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

predicción en los algoritmos de *Forecasting* utilizados para la predicción de la irradiancia. Las investigaciones previas se centraron principalmente en enfoque como regresión con la investigación de Cabrera [4] y Mora [2] y en clasificación con los estudios realizados de Pachajoa [1], todo esto sin considerar métricas específicas para *Forecasting*.

A esto se le sumó la dificultad para acceder a un dataset que cumpliera con los criterios espaciales y temporales adecuados para el entrenamiento y evaluación de los modelos de predicción de *Forecasting*. Los factores asociados al muestreo, la nubosidad y las condiciones atmosféricas, junto con las fallas en los sistemas de medición, limitaron la recopilación de estos datos.

Como consecuencia, los planes de energización que se habían generado estaban basados en datos históricos o del presente, lo que impide una proyección estratégica hacia el futuro. Esto generaba dificultades en la planificación de generación y distribución de energía fotovoltaica. Además, se evidencian planes de contingencia inadecuados ante fenómenos ambientales como incendios y sequías, lo que llevó a comprometer el cumplimiento de compromisos enmarcados en el Acuerdo de París y UPME.

En este contexto nuestra investigación buscó resaltar la importancia de comparar y evaluar algoritmos de *Forecasting* para la predicción de irradiancia, utilizando imágenes satelitales multiespectrales, mejorando así la precisión y eficiencia de los modelos.

D. FORMULACIÓN DEL PROBLEMA

¿Qué modelo de *Forecasting* es más eficiente en la predicción de oferta energética fotovoltaica?

E. OBJETIVOS

1) Objetivo general

Determinar el mejor equilibrio entre la calidad de las predicciones y la velocidad de procesamiento de un conjunto de algoritmos de *Forecasting* para la predicción de irradiancia mediante imágenes satelitales multiespectrales y técnicas de Machine Learning.

2) Objetivos específicos

- Componer un dataset de imágenes satelitales multiespectrales de reflectancia e irradiancia solar, para el entrenamiento de algoritmos de *Forecasting* aplicando técnicas de adquisición, procesamiento y transformación de datos.
- Diseñar un conjunto de experimentos de entrenamiento de algoritmos de *Forecasting* para la adquisición de métricas de calidad de predicción y velocidad de procesamiento a partir de un dataset de imágenes satelitales multiespectrales.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

- Clasificar la relación entre las métricas de calidad y el tiempo de inferencia de los experimentos realizados para determinar los mejores balances de cada configuración mediante una caracterización de algoritmos de Forecasting recomendada por expertos.

F. JUSTIFICACIÓN

Hoy por hoy, las energías renovables se han convertido en una prioridad global debido a la necesidad de sustituir la energía fósil y reducir el impacto ambiental [7]. La transición energética ha impulsado a países como Colombia a la inversión de tecnologías alternativas para la producción de energía con recursos renovables, aportando a las metas de cambio climático y promoviendo el uso de fuentes de energía como la solar y la eólica [8].

Los modelos tradicionales de predicción energética presentaron limitaciones al no capturar con suficiente precisión la variabilidad espacial y temporal de la irradiancia solar. En este contexto, los algoritmos de Machine Learning aplicados a Forecasting permitieron optimizar la precisión de estas predicciones mediante el análisis de imágenes satelitales multiespectrales. Satélites como Landsat y MODIS de la NASA demostraron ser clave en la generación de modelos de extrapolación e interpolación para visualizar la distribución de la irradiancia [2].

Este estudio validó que el uso de algoritmos avanzados pudo mejorar la precisión y eficiencia computacional en la predicción de la irradiancia solar. Su aporte se centró en el análisis del desempeño de diferentes modelos para optimizar la planificación energética y fomentar el desarrollo de estrategias basadas en generación distribuida con fuentes renovables [2].

En regiones como Nariño, donde la escasez de petróleo ha afectado el suministro energético, se han implementado planes como el PERS (Planes de Energización Rural Sostenible), orientados a mejorar el acceso a energías limpias en comunidades aisladas [4]. Además, Colombia, en cumplimiento con los compromisos adquiridos en el Acuerdo de París, busca reducir sus emisiones de gases de efecto invernadero y avanzar en la descarbonización del sector eléctrico.

En este sentido, la implementación de modelos de Forecasting basados en imágenes satelitales representó una herramienta clave para fortalecer estas iniciativas y optimizar la toma de decisiones en la transición hacia una matriz energética más sostenible [2].

G. DELIMITACIÓN

Este proyecto se realizó mediante experimentos sobre bases de datos obtenidas de imágenes Landsat y MODIS para el departamento de Nariño [4], Colombia, cuyo rango temporal va desde los años 2000 a 2015 para Landsat y 2005 a 2015 para MODIS, significando ello que en total se cuenta con 446 registros, obtenidos mediante un script SQL de una base de datos que se complementa de la siguiente manera:

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

- 21 Gb para Landsat
- 79 Gb para MODIS

Los modelos fueron entrenados usando redes neuronales artificiales: MLP, LSTM, GRU, Atencional MODELS, algoritmos tradicionales de Machine Learning como máquinas de soporte vectorial, árboles de decisión, random forest, ARIMA para evaluar los modelos se utilizaron métricas de Forecasting.

II. MARCO TEÓRICO

La predicción de la irradiancia solar ha sido un tema ampliamente estudiado debido a su importancia en la generación de energía renovable. Diferentes investigaciones han demostrado que el análisis de datos atmosféricos y ambientales mediante imágenes satelitales permiten estimar con mayor precisión la cantidad de radiación solar en una determinada zona. Estas imágenes multiespectrales contienen información clave sobre factores como la nubosidad, la humedad y la temperatura, los cuales influyen directamente en la irradiancia. A lo largo del tiempo, se han desarrollado distintos enfoques para procesar estos datos, mejorando la capacidad de predicción y optimización en el sector energético [2].

A. ANTECEDENTES

Antecedentes internacionales

A nivel internacional, Kim et al. (2024) desarrollaron un modelo para la predicción de la irradiancia solar superficial utilizando imágenes satelitales multiespectrales provenientes de satélites geoestacionarios y Deep Learning. El estudio demostró que la incorporación de información sobre la nubosidad y la relación espacio-temporal capturadas en las imágenes satelitales mejora la precisión de las predicciones de irradiancia, esto aumenta la estimación confiable del recurso solar y la planificación de sistemas fotovoltaicos [9].

En Ecuador, Acosta et al. (2022) utilizó el algoritmo XGBoost el cual optimizó con algoritmos genéticos para predecir la generación de energía fotovoltaica, logrando así mejores resultados al compararlos con los modelos tradicionales. También destaca la importancia de poder contar con variables meteorológicas y datos históricos confiables, especialmente en contextos regionales [10].

En Corea del Sur, Kim et al. (2020) desarrollaron un modelo de predicción de generación fotovoltaica usando imágenes satelitales del satélite GEO-KOMPSAT-2A. Implementaron un modelo E3D-LSTM para predecir la cantidad de nubosidad, mejorando la precisión en la predicción de la energía solar disponible. Este enfoque permitió una estimación más precisa de la generación fotovoltaica al considerar la nubosidad en el futuro [11].

En Marruecos, El Ouariachi et al. (2023) realizaron un estudio comparativo de algoritmos de Machine Learning para predecir la producción de energía solar. Se evaluaron seis modelos: regresión de vectores de soporte (SVR), redes neuronales artificiales (RNA), árboles de decisión (DT), bosques aleatorios (RF), modelos aditivos generalizados (GAM) y potenciación de gradiente extremo (XGBOOST). Los resultados mostraron que el modelo de RNA ofrecía la mayor precisión para predecir la producción diaria de energía en una planta solar [12].

Antecedentes nacionales

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

En Colombia, Ladino Nieto (2025) desarrolló un modelo de predicción multifuente de irradiancia solar que integra datos satelitales del producto GOES ABI L2 con modelos de aprendizaje profundo para pronosticar la irradiancia global horizontal a partir de información tabular y de imágenes, encontrando que la fusión de múltiples fuentes de datos mejora el desempeño predictivo frente a esquemas de persistencia tradicionales [13].

En Colombia, Velasco Pérez et al. (2024) investigaron sobre la predicción a corto plazo de irradiancia solar utilizando imágenes satelitales del satélite GOES-16 y arquitecturas de redes neuronales convolucionales conjunto con LSTM, mostrando que los enfoques de aprendizaje profundo permiten capturar características espaciales y temporales relevantes para mejorar las estimaciones del recurso solar [14].

Antecedentes regionales

A nivel local, en Nariño, Cabrera et al. (2016) emplearon imágenes satelitales y algoritmos de regresión para calcular la irradiación solar en zonas rurales del departamento. Además[4], Mora (2019) utilizó imágenes satelitales provenientes de Landsat y MODIS para implementar técnicas de aprendizaje automático (SVM y ANN) con funciones kernel, lo que permitió una mejor precisión en la predicción de la oferta energética mediante la optimización de hiper parámetros [2].

B. SUPUESTOS TEÓRICOS

1. Fuentes de energías limpias

Las fuentes de energía alternativa son generalmente llamadas a aquellas energías que son usadas para transformarse en electricidad y cuyo origen primario es diferente al de las fuentes fósiles como energías que afectan el medio ambiente como el petróleo y carbón; y las tradicionales fuentes hidroeléctricas. De igual forma, debido al desarrollo tecnológico industrial, las energías alternativas tienen poca participación en el mercado de energía mundial [15]. A continuación, se definirán las energías alternativas más relevantes.



*Fig. 1. Energías limpias
Fuente [16]*

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Energía solar: la energía solar está directamente relacionada con procesos químicos, físicos y biológicos, es decir, está asociada a la vida en la Tierra [1]. Los cambios en esta energía influyen en la atmósfera y el clima del planeta a través de amplios periodos de tiempo. En este estudio se muestran los recursos bibliográficos segmentados en elementos teóricos y proyecciones para la generación de patrones de energía solar fotovoltaica.

Irradiancia solar: es la cantidad de energía proveniente del sol, por unidad de área, que recibe todo punto por encima de la atmósfera terrestre, en cada instante de tiempo; por lo tanto, es una cantidad que varía tanto espacial como temporalmente, de acuerdo a la posición de un punto en particular sobre la Tierra y de la posición de ésta respecto del Sol [17]. La irradiancia es un factor importante que influye en los procesos químicos, físicos y biológicos de la Tierra. Esta emitida por las diferentes capas de la atmósfera solar afecta a la atmósfera superior y también al clima de la Tierra [18].

Satélite: es un objeto que orbita (da vueltas) alrededor de un planeta. Hay centenares de satélites naturales, o lunas, en nuestro sistema solar, pero, desde 1957, también se han lanzado al espacio miles de satélites artificiales (fabricados por el hombre). Estos tienen usos muy diversos, como captar imágenes del Sol, la Tierra y otros planetas, o explorar el espacio para estudiar los agujeros negros, y las estrellas y galaxias remotas. También están los satélites de comunicaciones, los satélites meteorológicos y la Estación Espacial Internacional [19].



Fig. 2. Satélite
Fuente:[20]

Imágenes satelitales: las imágenes satelitales son fotografías de la superficie de la Tierra tomadas desde el espacio por un satélite artificial. Estas imágenes son útiles para estudiar el clima, la vegetación, la contaminación, la geología, la topografía y la planificación urbana, entre otros. También se utilizan para hacer mapas y para monitorear desastres naturales, como terremotos, inundaciones y huracanes. Gracias a las imágenes satelitales, se visualizará la Tierra desde una perspectiva diferente y obtener información muy valiosa [21].

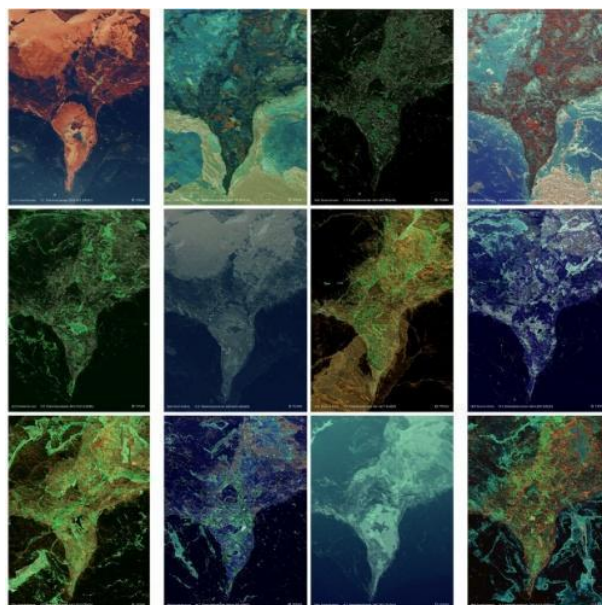


Fig. 3. Imágenes satelitales
Fuente:[22]

Captura de imágenes satelitales: Las imágenes satelitales se obtienen a través de sensores remotos que están ubicados en satélites artificiales que giran alrededor del planeta. Estos sensores detectan la radiación electromagnética que emite o refleja la superficie de la Tierra, y que es originada principalmente por la energía del sol. Cuando la radiación solar llega a la superficie de la Tierra y se refleja en varias longitudes de onda, según las propiedades físicas de los objetos (el agua, el suelo o la vegetación).

Después, los sensores del satélite captan la energía reflejada y la convierten en señales digitales que se transmiten a estaciones en la Tierra para ser procesadas. Cada valor anotado se refiere a un píxel de la imagen, que contiene información espectral concreta y facilita el análisis de variables del medio ambiente, como lo son la nubosidad, la temperatura y la reflectancia. Este tipo de adquisición es esencial para el análisis multiespectral porque posibilita la obtención de datos de la superficie terrestre en diversas bandas del espectro electromagnético [23],[24].

Imágenes satelitales multiespectrales: imágenes del mismo objeto representadas cada una de ellas con diferentes longitudes de onda. Están formadas por bandas que no necesariamente son contiguas unas a otras. El procesamiento y la visualización de las imágenes satelitales tienen como propósito facilitar su uso y comprensión, y estar en condiciones de procesar las bandas espectrales de manera que se pueda obtener una imagen visual adecuada para lo propuesto en esta investigación [1].

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

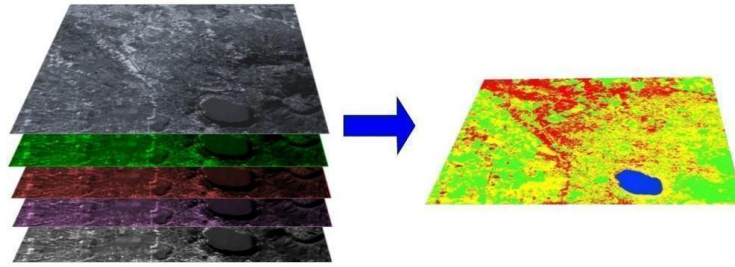


Fig. 4. . Imagen satelital multispectral
Fuente: [25]

Sensores satelitales (pasivos, activos, multispectrales): Los sensores satelitales son dispositivos instalados en los satélites encargados de captar la información de la superficie terrestre mediante la detección de energía electromagnética. Estos sensores permiten obtener datos en diferentes longitudes de onda, lo cual es fundamental para el análisis de variables ambientales y climáticas.

Los sensores pasivos y los sensores activos son los dos tipos de sensores que existen, en términos generales. Los sensores pasivos recopilan la radiación natural que los objetos emiten o reflejan, y la radiación solar es su fuente de energía primordial. Satélites como Landsat y MODIS, que recogen información en diferentes bandas del espectro electromagnético, emplean estos tipos de sensores para posibilitar el análisis multispectral de la superficie terrestre. Por otra parte, los sensores activos, que funcionan como los sistemas de radar, envían su propia señal de energía a la superficie y registran la respuesta devuelta. Esto les posibilita operar en situaciones con nubosidad o sin luz solar.

En esta clasificación, los sensores multispectrales son esenciales porque recogen información en varias bandas espectrales, como la del infrarrojo cercano, la visible y otras zonas del espectro. Esta habilidad posibilita el reconocimiento de rasgos particulares de la superficie del planeta, como la vegetación, la cobertura de nubes y la humedad, elementos cruciales para calcular la irradiancia solar.[26]

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

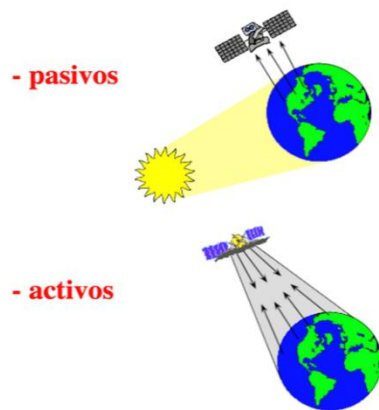


Fig. 5. Imagen sensores satélites
Fuente: [27]

Procesamiento de imágenes satelitales: El procesamiento de imágenes satelitales comprende un conjunto de técnicas y procedimientos aplicados a las imágenes obtenidas por sensores remotos con el fin de mejorar su calidad, extraer información relevante y prepararlas para su análisis. Este proceso es fundamental en estudios ambientales y energéticos, ya que permite transformar los datos crudos capturados por los satélites en información útil para la toma de decisiones. Inicialmente, las imágenes satelitales requieren una fase de corrección, la cual puede ser geométrica y radiométrica.

La corrección geométrica permite ajustar la posición espacial de la imagen para que coincida con coordenadas reales de la superficie terrestre, mientras que la corrección radiométrica busca eliminar errores asociados a los sensores o a las condiciones de captura, garantizando la consistencia de los valores registrados. Posteriormente, se aplican técnicas de preprocesamiento como la eliminación de ruido y la corrección atmosférica, que permiten reducir las distorsiones causadas por factores como la nubosidad, partículas en suspensión o interferencias del sensor. Estas técnicas mejoran la calidad de la información espectral, lo cual es esencial para el análisis de variables como la reflectancia y la irradiancia solar.

Adicionalmente, cuando se requiere analizar grandes extensiones geográficas, se emplean técnicas de composición o mosaico de imágenes, las cuales consisten en la unión de múltiples escenas satelitales para generar una representación continua de una región específica. Este procedimiento es especialmente relevante en estudios regionales, como el análisis de la irradiancia en el departamento de Nariño [28].

Representación de datos a partir de imágenes satelitales: La representación de datos a partir de imágenes satelitales consiste en transformar la información contenida en las imágenes en estructuras organizadas que puedan ser utilizadas por algoritmos de análisis y modelos de aprendizaje automático. Los píxeles que conforman las imágenes satelitales

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

contienen un valor numérico conocido como número digital (DN, por sus siglas en inglés), que representa la fuerza de la radiación electromagnética captada por el sensor en una banda espectral específica.

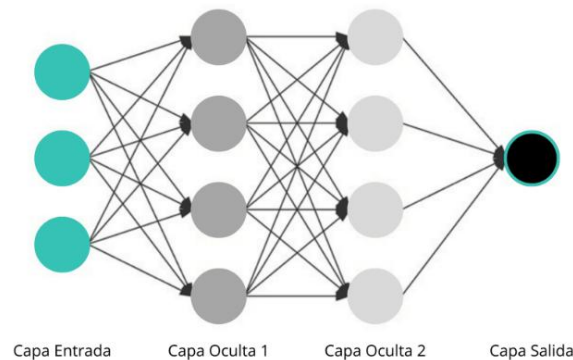
Para emplearlos en modelos computacionales, es necesario extraer estos valores y organizarlos en tablas. Este procedimiento supone transformar la imagen en una matriz de datos. En ella, cada fila puede ser un píxel o una zona de interés, y cada columna representa variables como las diversas bandas espectrales, índices derivados o información temporal. Así, se generan conjuntos de datos estructurados que posibilitan el uso de métodos de análisis estadístico y Machine Learning.

Además, en numerosas ocasiones, los valores digitales son convertidos a variables físicas más comprensibles, como la irradiancia o la reflectancia, mediante procedimientos de normalización y calibración. Esto posibilita incrementar la calidad de los datos y hacer más fácil su interpretación en modelos predictivos. Además, es posible incluir datos extra, como variables temporales o meteorológicas, para mejorar el conjunto de información que se emplea en el entrenamiento de los modelos [29],[30].

TriTier: se define como un sistema compuesto por un conjunto de sensores instalados a nivel terrestre, cuyo propósito es medir y generar datos de irradiancia solar. Estos datos se obtienen a partir de mediciones directas en superficie, lo que permite contar con información precisa sobre la radiación disponible en una zona específica. En el contexto de esta investigación, los datos de Tritier se emplean como conjunto objetivo o referencia, ya que proporcionan valores reales de irradiancia frente a los cuales se pueden comparar y validar los modelos predictivos.[1],[2]

2. Algoritmos de aprendizaje automático

El aprendizaje automático (ML en inglés) es un tipo de algoritmo que se mejora automáticamente a sí mismo basado en la experiencia, no por un programador que escribe un mejor algoritmo. El algoritmo gana experiencia al procesar más y más datos, y luego modificándose basado en las propiedades de los datos [32].



*Fig. 6. Algoritmo de aprendizaje
Fuente:[32]*

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Forecasting: consiste en la estimación y monitorización de variables futuras para un producto utilizando datos históricos. Hoy en día, la previsión combina el análisis de datos, el aprendizaje automático, el modelado estadístico y el juicio de expertos [33]. Por ello, desde un enfoque energético solar, el Forecasting va desde las variables de entrada hasta las variables de salida [34]. Por lo tanto, el Forecasting de irradiancia puede tomar varios enfoques.

Deep Learning: el aprendizaje profundo es un tipo de métodos de aprendizaje automático basados en el aprendizaje de representaciones de datos. Usa una red neuronal artificial que se compone de niveles jerárquicos. En el nivel inicial la red aprende algo simple y luego envía esta información al siguiente nivel. El siguiente nivel combina, y compone una información un poco más compleja, y se lo pasa al 3º nivel, y así sucesivamente [1].

Redes neuronales artificiales: está formada por neuronas artificiales conectadas entre sí y agrupadas en diferentes niveles denominadas capas. Las neuronas de la primera capa recibe como entrada los datos reales que alimentan a la red neuronal. Es por eso por lo que la primera capa se conoce como capa de entrada. La salida de la última capa es el resultado visible de la red, por lo que la última capa se conoce como la capa de salida. Las capas que se sitúan entre la capa de entrada y la capa de salida se conocen como capas ocultas ya que desconocemos tanto los valores de entrada como los de salida [1].

MLP (Multilayer Perceptrón): una red neuronal está construida por un conjunto de nodos interconectados que se denominan neuronas y se organizan en capas. Una red neuronal básica tiene una capa de entrada (Input layer), una capa intermedia (Hidden layer) y una capa de salida (Output layer) [35] se refiere a una clase de red neuronal artificial compuesta por diversas capas neuronales. Las neuronas del MLP generalmente emplean funciones de activación no lineales, lo que facilita a la red el aprendizaje de patrones complejos en la información. Los MLP son esenciales en el aprendizaje automático ya que tienen la capacidad de aprender relaciones no lineales en los datos, transformándolos en poderosos modelos para actividades como la clasificación, la regresión y la identificación de patrones [36].

LSTM (Long Short-Term Memory): son una forma especializada de la arquitectura de RNN. Las RNN utilizan información anterior para mejorar el rendimiento de una red neuronal con entradas actuales y futuras. Las LSTM funcionan bien con datos de secuencias y series temporales para tareas de clasificación y regresión [37].

GRU (Gated Recurrent Unit): son una arquitectura de Red Neuronal Recurrente creada para manejar de manera eficiente datos secuenciales, como secuencias de tiempo. Las GRU buscan solucionar el problema del gradiente evanescente que podría impactar a las GRU [38].

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Atencional MODELS: mecanismo de atención es una red neuronal que simula la atención del cerebro. Se ha aplicado con éxito a la traducción automática, el análisis de video y otros campos relacionados. La aplicación del mecanismo de atención a la red neuronal profunda permite que esta se enfoque adaptativamente en las características de entrada más importantes para la salida actual y mitigue la interferencia de otras características. Utilizando el vector de salida de la capa oculta LSTM [39].

Machine Learning: el aprendizaje automático es una rama en evolución de los algoritmos computacionales, diseñados para emular la inteligencia humana aprendiendo del entorno. Se consideran la herramienta clave en la nueva era del llamado big data [40].

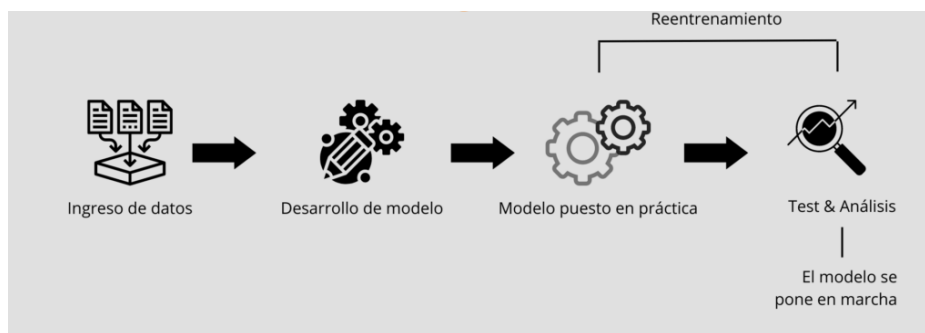


Fig. 7. Proceso
Fuente:[41]

C. VARIABLES DEL ESTUDIO

Variables Dependientes

- R^2 (Coeficiente de determinación)
- Tiempo de predicción

Variables Independientes

Algoritmos Machine Learning Tradicionales

- SVM (Support Vector Machine)
- Random forest
- KNN (K-nearest neighbors)
- MLP (Multilayer Perceptron)

Deep Learning

- LSTM (Long Short-Term Memory)
- GRU (Unidad Recurrente Cerrada)
- Técnicas de preprocesamiento

D. DEFINICIÓN NOMINAL DE LAS VARIABLES

Variables Dependientes

-

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

- **Coefficiente de determinación:** (R^2) es una medida estadística que cuantifica la proporción de la varianza de la variable dependiente explicada por las variables independientes en un modelo de regresión [42].
- **Distancia de Warping Dinámico:** (DTW) es un algoritmo de programación dinámica que mide la similitud entre dos secuencias temporales con posibles variaciones de velocidad. Minimiza la distancia acumulada al alinear los puntos, permitiendo estiramientos o compresiones temporales [43].
- **Tiempo de predicción:** predice resultados futuros analizando patrones en datos históricos con al menos dos ciclos temporales (periodos divididos, como meses en años) [44].

Variables Independientes

Algoritmos Machine Learning Tradicionales

- **SVM:** es un algoritmo de aprendizaje supervisado utilizado para clasificación y regresión. Busca encontrar el hiperplano óptimo que mejor separe las clases en el espacio de características, maximizando el margen entre las clases más cercanas (vectores de soporte) [45].
- **Random forest:** el bosque aleatorio es un algoritmo de machine learning de uso común, que combina el output de múltiples árboles de decisión para llegar a un único resultado. Su facilidad de uso y flexibilidad han impulsado su adopción, ya que maneja tanto problemas de clasificación como de regresión [46].
- **KNN:** K-nearest neighbors es un clasificador de aprendizaje supervisado no paramétrico, que emplea la proximidad para realizar clasificaciones o predicciones sobre la agrupación de un punto de datos individual [47].
- **MLP:** el perceptrón multicapa (Multilayer Perceptron MLP) propuesto por Rumelhart es el componente principal de una red neuronal, proporciona la base para la mayoría de las aplicaciones de las redes neuronales [48].

Deep Learning

- **LSTM:** la memoria a corto-largo plazo es una técnica informática que se utiliza para almacenar información dentro de una red neuronal durante un periodo de tiempo más largo. Esto es especialmente importante en el procesamiento de datos secuenciales [49].
- **GRU:** una Unidad Recurrente Cerrada (GRU) es un tipo de arquitectura de red neuronal recurrente (RNN) que se utiliza en el campo del aprendizaje profundo. Las GRU son

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

particularmente eficaces para procesar secuencias de datos en tareas como la predicción de series temporales [50].

- **ConvLSTM2D:** La arquitectura ConvLSTM2D es una variante de las redes neuronales recurrentes que combina las capacidades de las redes convolucionales (CNN) para la extracción de características espaciales con la habilidad de las redes LSTM (Long Short-Term Memory) para modelar dependencias temporales.

A diferencia de las LSTM tradicionales que funcionan con vectores unidimensionales, emplea operaciones convolucionales en sus estados internos. Esto posibilita que la estructura espacial de los datos se mantenga a través del tiempo. Así, las entradas, las salidas y los estados ocultos se muestran como matrices de dos dimensiones, lo que favorece la detección de patrones dinámicos en datos espaciales [51].

La arquitectura ConvLSTM2D puede incluir varias capas de este tipo, junto con procesos de normalización y regularización que ayudan a optimizar la estabilidad del entrenamiento y disminuir el sobreajuste. Además, es posible que se agreguen las capas convolucionales finales, que son las responsables de producir la salida del modelo.

- **Técnicas de preprocesamiento:** el preprocesamiento de datos incluye varias técnicas, como la limpieza de los datos para manejar los valores que faltan, la eliminación de los valores atípicos, el escalamiento de las características, la codificación de las variables categóricas y la división de los datos en conjuntos de entrenamiento y pruebas. Estas técnicas son clave para garantizar que los datos tienen un formato coherente y utilizable para los algoritmos de ML [52].

E. DEFINICIÓN NOMINAL DE LAS VARIABLES

Variables Dependientes

- **Coefficiente de determinación:** el R^2 siempre está entre 0 y 100%.
0% indica que el modelo no explica ninguna porción de la variabilidad de los datos de respuesta en torno a su media.
100% indica que el modelo explica toda la variabilidad de los datos de respuesta en torno a su media. [53]
- **Tiempo de predicción:** el periodo requerido por un modelo para efectuar una predicción puede manifestarse en diferentes unidades de tiempo, tales como segundos, milisegundos, microsegundos o, en ciertas situaciones, minutos, en función de la complejidad del procedimiento.

Variables Independientes

Algoritmos Machine Learning Tradicionales

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

- **SVM:**

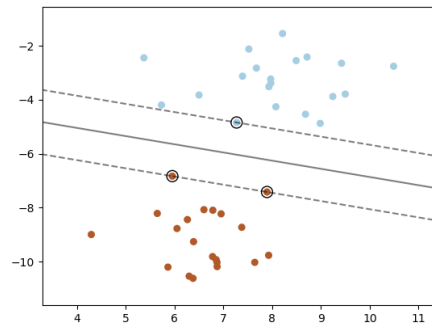
Ecuación 1. SVM

$$wx + b = 0 \quad (1)$$

Donde w es el vector de peso, x es el vector de entrada y b es el término de sesgo.

$$(wx + b)y_j \geq a$$

y luego se maximiza el margen, lo que se representa como: $\max \gamma = \|w\| a$, donde a es el margen proyectado sobre w [54].



*Fig. 8. Algoritmo
Fuente:49*

- **RANDOM FOREST**

Clasificación:

Ecuación 2. Clasificación

$$f(x) = \text{modo}\{hb(x): b = 1, 2, \dots, B\} \quad (2)$$

$f(x)$: predicción final (la clase que más se repite).

$hb(x)$: predicción del árbol número b .

B : número total de árboles.

Regresión:

Ecuación 3. Regresión

$$f(x) = \sum_{b=1}^{B_1} B_h b(x) \quad (3)$$

$f(x)$: predicción final (promedio de todas las predicciones de los árboles).

$hb(x)$: predicción del árbol número b .

$B1$: número total de árboles[55].

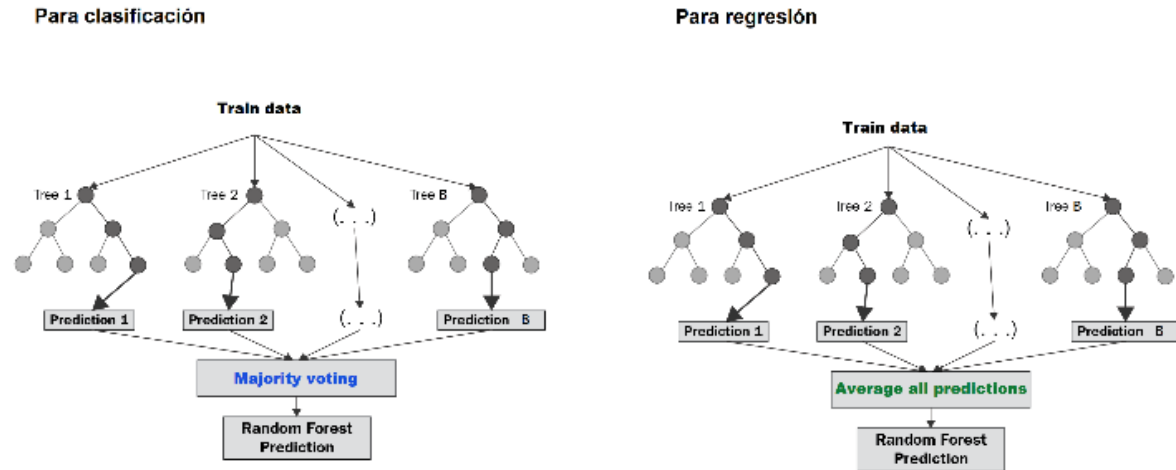


Fig. 9. Regresión y clasificación
Fuente:[56]

- **KNN:**

Ecuación 4. KNN

$$f(x) = k1i = 1\sum kyi \quad (4)$$

$f(x)$: predicción final para el punto x . (Puede ser la clase elegida o el promedio de valores.)

x : punto nuevo que quieres clasificar o predecir.

kli : cada uno de los puntos de entrenamiento (vecinos).

kyl : valor real (etiqueta o número) del punto. (Clase si es clasificación, número si es regresión.)

k : número de vecinos más cercanos que se consideran [57].

- **MLP:** las neuronas de las capas adyacentes están completamente conectadas entre sí. Cada conexión tiene un peso asociado, que determina su fuerza. Estos pesos se aprenden durante el proceso de entrenamiento.

Deep Learning

- **LSTM:**

1. **Puerta de olvido (Forget Gate):** decide qué parte de la memoria previa olvidamos:

Ecuación 5. Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

2. **Puerta de entrada (Input Gate):** decide qué parte de la nueva información almacenamos:

Ecuación 6. Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

3. **Actualizar la memoria (Cell State):** actualiza la memoria antigua con la nueva información:

Ecuación 7. Cell State

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (7)$$

4. **Puerta de salida (Output Gate):** decide qué parte de la memoria mostramos como salida:

Ecuación 8. Output Gate

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (8)$$

[58]

- **GRU:**

1. **Puerta de actualización (Update Gate):** controla cuánto del estado anterior se conserva.

Ecuación 9. Update Gate

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (9)$$

2. **Puerta de reseteo (Reset Gate):** controla cuánto del estado anterior olvidamos.

Ecuación 10. Reset Gate

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (10)$$

3. **Cálculo del candidato a nuevo estado oculto:** genera un nuevo contenido candidato basado en la entrada actual y el estado anterior modificado.

Ecuación 11. Nuevo Estado Oculto

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t] + b_h) \quad (11)$$

4. **Actualización final del estado oculto**

Ecuación 12. Actualización estado oculto

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (12)$$

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Elementos usados:

σ : función sigmoide (output entre 0 y 1).

$\tanh$: tangente hiperbólica (output entre -1 y 1).

$*$: multiplicación elemento a elemento (Hadamard).

h_{t-1} : estado oculto anterior.

x_t : entrada actual.

W, b : matrices de pesos y vectores de sesgos aprendibles.

● TÉCNICAS DE PREPROCESAMIENTO:

1. **Normalización min – máx**: escala todos los valores de una característica entre 0 y 1.

Ecuación 13. Normalización

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (13)$$

x : Valor original.

x_{min}, x_{max} : mínimo y máximo de la característica.

2. **Estandarización (Z-score normalization)**: convierte los datos para que tengan media 0 y desviación estándar 1.

Ecuación 14. Estandarización

$$x' = \frac{x - \mu}{\sigma} \quad (14)$$

μ : media de los datos.

Σ : desviación estándar.

3. **One-Hot Encoding**: convierte una variable categórica en un vector binario. Si tenemos una categoría CCC que puede tomar 3 valores: {Rojo, Azul, Verde}, se convierte en:

Ecuación 15. One hot

$$\text{Rojo} \rightarrow (1, 0, 0), \text{Azul} \rightarrow (0, 1, 0), \text{Verde} \rightarrow (0, 0, 1) \quad (15)$$

4. **Imputación simple de datos faltantes (Media)**: cuando falta un valor x_i , se puede imputar usando la media:

Ecuación 16. Imputación

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j \quad (16)$$

Promediando todos los demás valores existentes [59].

F. FORMULACIÓN DE HIPÓTESIS

Hipótesis de investigación

Al menos uno de los algoritmos de Forecasting logra un coeficiente de determinación mayor a un modelo aleatorio.

Hipótesis nula

Los algoritmos de Forecasting no logran un coeficiente de determinación mayor a un modelo aleatorio.

Hipótesis alterna

Los algoritmos de Forecasting explican significativamente más la variabilidad de los datos que los métodos tradicionales, según el coeficiente de determinación.

III. METODOLOGÍA

A. PARADIGMA

El proyecto es de paradigma positivista ya que tuvo como objetivo comprobar una hipótesis por medios estadísticos, así como determinar parámetros de una determinada variable mediante expresiones numéricas [60]. Este paradigma se basó en métodos estadísticos, tanto descriptivos como inferenciales. Su objetivo fue generar un conocimiento que sea sistemático, comprobable, comparable, medible y replicable. El objeto de estudio se enfocó en fenómenos observables, ya que solo estos pueden ser medidos, analizados y sometidos a control experimental.

B. ENFOQUE

La perspectiva empírico-analítica representó el fundamento de la investigación, puesto que se enfocó en la realidad de los sucesos que pueden ser observados, cuantificados y evaluados. Este procedimiento facilitó el contraste riguroso de las hipótesis a través de la demostración científica, lo que contribuyó a establecer si una hipótesis era auténtica o falsa. Este principio se manifestó en el propósito de este estudio, en el que se llevan a cabo varios experimentos para crear modelos y confirmar o refutar las hipótesis propuestas [1].

C. MÉTODO

El método de la investigación fue de carácter científico, puesto que se basó en la búsqueda de la verdad mediante la evaluación de algoritmos de Forecasting, con el fin de determinar en qué conjunto de datos presentan un mejor rendimiento. Para el desarrollo del estudio, se adoptó la metodología CRISP-DM, la cual permitió estructurar el proceso analítico en fases definidas, desde la comprensión del problema hasta la evaluación de los resultados, garantizando un enfoque ordenado y orientado a la obtención del conocimiento.

D. TIPO DE INVESTIGACIÓN

La investigación fue de tipo experimental porque se buscó comprobar cuál algoritmo de Forecasting predecía con mejor precisión la irradiancia solar utilizando imágenes satelitales. Se realizaron varios experimentos que probaron distintos modelos, midiendo su precisión y el tiempo que demoran en generar las predicciones. Se logró así identificar cuál modelo ofrecía los mejores resultados en términos de exactitud y eficiencia.

E. DISEÑO DE LA INVESTIGACIÓN

El diseño de la investigación se fundamentó en una metodología experimental, ya que se realizaron diferentes pruebas de rendimiento y precisión sobre un conjunto de algoritmos de *Forecasting* aplicados al análisis de imágenes satelitales multiespectrales. El proyecto se articuló con base en un proceso estructurado que abarca desde la adquisición y preparación

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

del conjunto de datos hasta la evaluación comparativa de modelos de predicción y su visualización espacial. Las fases del proceso fueron las siguientes.

TABLA I. METODOLOGÍA PRELIMINAR

Fase	Actividad	Resultados esperados	Recursos
1. Recolección y construcción del dataset	- Descargar imágenes satelitales Landsat y MODIS para Nariño (2000–2015). - Preprocesar datos: limpieza, eliminación de nubes, interpolación temporal. - Combinar reflectancia con datos de irradiancia.	Conjunto de datos estructurado y depurado.	PostgreSQL, Python, QGIS, Google Colab, bibliotecas de preprocesamiento (Pandas, Rasterio, NumPy).
2. Entrenamiento y evaluación de algoritmos de Forecasting	- Entrenar modelos ML/DL: MLP, LSTM, GRU, Random Forest, SVM, ARIMA. - Evaluar mediante métricas R^2, DTW y tiempo de inferencia.	Modelos entrenados con métricas de evaluación.	Scikit-Learn, TensorFlow, Keras, Google Colab (TPU/GPU), Pandas, Matplotlib.
3. Análisis y clasificación de resultados	- Comparar resultados por algoritmo según métricas y eficiencia. - Clasificar modelos según tipo y comportamiento.	Tabla comparativa y análisis de desempeño.	Python, herramientas estadísticas, hojas de cálculo, bibliotecas de visualización.

Finalmente, se desarrolló un documento que incluyó toda la información para facilitar la interpretación de los resultados por parte de investigadores, tomadores de decisiones y comunidades interesadas en la planificación energética basada en fuentes renovables. Este sistema contribuyó al fortalecimiento de iniciativas de generación distribuida y transición energética en el departamento de Nariño.

F. POBLACIÓN

Esta investigación se centró en el análisis de los métodos de *Forecasting*, lo que implicó que no fuera necesario definir una población o una muestra. Esto se debió a que la naturaleza del proyecto requirió un enfoque basado en datos, donde la información se convirtió en uno de los pilares fundamentales. Para lograrlo, se recurrió al uso de *Big Data* y a repositorios de datos creados por investigadores para diversos fines. En este sentido, la investigación no fue la excepción: los métodos de *Forecasting* que se evaluaron se aplicaron a diferentes conjuntos de datos, lo que significó que no se utilizaron datos específicos, sino que se aprovecharon diversas bases de datos ya establecidas.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

G. MUESTRA

Las muestras se obtuvieron a partir del procesamiento de los mismos datos, mediante el uso de técnicas de suavizado de promedios o a través de la selección de muestras aleatorias. La población estuvo constituida por todos los registros pertinentes a la generación de energía solar y eólica en el departamento de Nariño que resultaron relevantes para el análisis.

H. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

Este proyecto no necesitó la recolección de información, ya que existían bases de datos que habían sido utilizadas por los investigadores para probar los métodos de Forecasting propuestos por la comunidad científica y académica.

I. VALIDEZ DE LAS TÉCNICAS DE RECOLECCIÓN DE LA INFORMACIÓN

La validez de la utilidad de las bases de datos de los satélites Landsat y MODIS se examina desde múltiples puntos de vista, que apoyan los temas de estudio enfocados en la irradiancia solar, los cuales están relacionados a continuación:

Calidad y resolución de datos

- **Landsat:** ofrece imágenes espaciales y espectrales de alta resolución, que facilitan la supervisión exacta de la superficie terrestre y la valoración de la reflectancia en distintas bandas del espectro electromagnético. Es esencial para calcular eficazmente la irradiancia solar.
- **MODIS:** proporciona información en resoluciones más extensas, aunque con una alta frecuencia temporal (cada 1-2 días), lo que facilita el estudio dinámico de las variaciones en la irradiancia a través de diversas condiciones climáticas y estacionales.

Base de datos establecidas y reconocidas

- Las investigaciones previamente citadas han evidenciado la fiabilidad de los datos de Landsat y MODIS en estudios vinculados a la agricultura, el medio ambiente y la planificación urbana, lo que fortalece su utilidad para calcular la irradiancia solar.

Capacidad de integración de datos

- Los datos de Landsat y MODIS pueden ser integrados con otras fuentes de datos, por ejemplo: meteorológicos, para mejorar la precisión de las predicciones de irradiancia. Esta capacidad de fusión de datos permite realizar análisis más completos y robustos.

Desarrollo de modelos predictivos

- Las bases de datos utilizadas se estructuraron para los periodos 2003 a 2014 para Landsat y 2005 a 2015 para MODIS como se mencionó anteriormente ofrecen grandes volúmenes de datos históricos, los cuales son fundamentales para el desarrollo y validación de

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

modelos predictivos de Machine Learning. Esto es especialmente importante cuando se busca predecir patrones de irradiancia basados en variables causales.

Aplicaciones multidisciplinarias

- La adaptabilidad de los datos de estos satélites posibilita su uso en diferentes campos, tales como el cambio climático, la administración de recursos naturales y la planificación energética, evidenciando así su importancia y capacidad de adaptación a distintos contextos de estudio.

J. CONFIABILIDAD DE LAS TÉCNICAS DE RECOLECCIÓN

Base de datos

- Ambas bases de datos son producidas y mantenidas por agencias de renombre como la NASA y el Servicio Geológico de los EE. UU. (USGS). Su reputación y uso extendido en diversas investigaciones científicas confirman la fiabilidad de los datos recolectados.

Acceso y compartición de datos

- Los datos de Landsat y MODIS son accesibles públicamente y están disponibles para investigadores a nivel mundial. Lo que fomenta la colaboración y el intercambio de conocimientos entre la comunidad científica, además de permitir la replicación de estudios y la validación de resultados.

Validación cruzada

- Existen varios estudios que han utilizado datos de Landsat y MODIS para la estimación de irradiancia solar, y sus resultados han sido validados comparándolos con mediciones in situ. Esta validación cruzada proporciona confianza adicional en la aplicabilidad de estos instrumentos.

Lo anterior permitió concluir que la combinación de alta calidad de datos, reconocimiento internacional, capacidad de integración y validación cruzada hace que las bases de datos de Landsat y MODIS sean instrumentos válidos y aplicables para esta investigación. Además, su accesibilidad ayudó a los investigadores avanzar en la comprensión del potencial solar y en la implementación de energías renovables.

K. INSTRUMENTO DE RECOLECCIÓN DE DATOS

La recolección de imágenes fue llevada a cabo por la Administración Nacional de Aeronáutica y del Espacio (NASA) utilizando los satélites Landsat y MODIS. Después de preprocesar los datos con el algoritmo Automatic Cloud Cover Assessment para Landsat y el producto MOD09GA para MODIS [61], se obtuvieron los valores digitales de imagen. Estos valores digitales se calcularon aplicando una ecuación en la que I representa el valor digital del píxel en la posición (i, j) para la banda k , mientras que DN es el número digital registrado por el sensor. Posteriormente, se realizó la conversión de DN a radiancia utilizando una fórmula específica, donde la radiancia es el resultado final y $Gain$ y $Offset$ son los

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

coeficientes correspondientes al sensor. Finalmente, se llevó a cabo la conversión de la irradiancia a reflectancia.

L. METODOLOGÍA DE DESARROLLO

1. Comprensión del problema

El uso de imágenes satelitales multiespectrales y técnicas de inteligencia artificial ha surgido como alternativa prometedora para la construcción de mapas de irradiación y modelos de predicción. Sin embargo, no se ha establecido con claridad qué algoritmo de Forecasting presenta un mejor desempeño en precisión y eficiencia.

Dado que la predicción de irradiancia no solo requiere exactitud, si no también tiempos de procesamiento adecuados para su aplicación, surge la necesidad de evaluar y comparar diferentes enfoques de modelado, para así lograr un balance óptimo entre la calidad de predicción y la velocidad de inferencia.

2. Adquisición y preparación de datos (Landsat y MODIS)

Como parte del desarrollo de la investigación, se construyó un conjunto de datos a partir de información obtenida de imágenes satelitales multiespectrales. Para ello, se emplearon datos provenientes de los satélites Landsat y MODIS, los cuales contienen información de reflectancia en diferentes bandas espectrales.

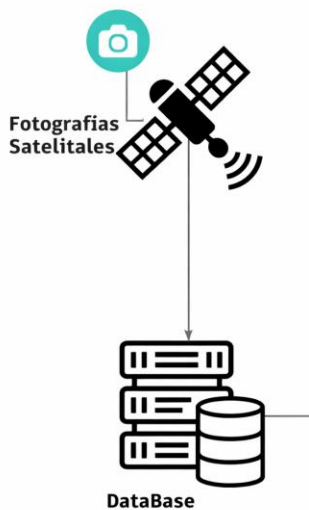


Fig. 10. Adquisición de imágenes
Fuente: [nosotras]

Estos datos fueron integrados y almacenados en un sistema gestor de bases de datos PostgreSQL, con el propósito de facilitar su manipulación, consulta y procesamiento. Las bases de datos incluyeron registros de reflectancia e irradiancia organizados según su resolución espacial y temporal.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Para su procesamiento, las tablas fueron combinadas mediante un enfoque de procesamiento distribuido en cada base de datos, de acuerdo con lo planteado por Mora [2], el cual consiste en fragmentar las tablas y ejecutar los cruces en distintos procesos en lugar de una única transacción.

Para la construcción del conjunto de datos, se implementó un proceso de cruce segmentado entre las tablas de reflectancia e irradiancia almacenadas en la base de datos, el cual permitió integrar la información espacial y temporal asociada a cada registro, considerando atributos como la latitud, la longitud y la fecha de captura.

Con el fin de optimizar el procesamiento de grandes volúmenes de información, se aplicaron estrategias de segmentación temporal que permitieron dividir los datos en intervalos específicos, facilitando así la ejecución eficiente de las consultas y reduciendo la carga computacional.

La estructura de las relaciones entre las tablas de reflectancia e irradiancia, así como su integración, se presenta en la Fig. 12, donde se ilustra el modelo entidad-relación utilizado para el almacenamiento de la información.

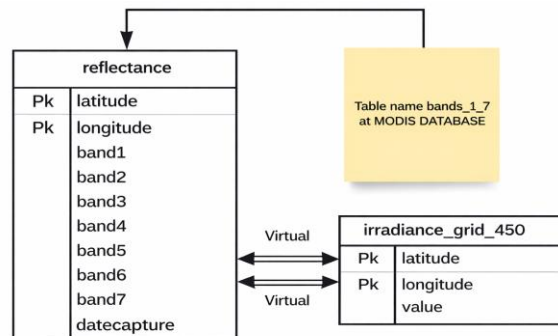


Fig. 11. MER de las tablas de reflectancia e irradiancia en las bases de datos Landsat y MODIS
Fuente: [1]

A partir de este enfoque, se generaron consultas SQL orientadas a la extracción de subconjuntos de datos correspondientes a intervalos temporales específicos. Estas consultas permitieron construir datasets segmentados que posteriormente fueron utilizados en las etapas de procesamiento y modelado. Un ejemplo de estas consultas se presenta en el Anexo A.

A partir de este proceso, se generaron los datasets correspondientes, los cuales fueron exportados en múltiples archivos en formato CSV para facilitar su uso en etapas posteriores.

Finalmente, el conjunto de datos fue almacenado en un repositorio en línea con el propósito de garantizar su reutilización y disponibilidad para investigaciones futuras. El acceso puede realizarse mediante el siguiente enlace:

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

<https://drive.google.com/drive/folders/1C5Lpb7UxRCo6hrQRkcJo3IHWUXGOMtcD?usp=sharing>

Este proceso fue formalizado mediante las ecuaciones de álgebra relacional presentadas en las ecuaciones (17), (18) y (19).

Ecuación 17. Partición de la Reflectancia por rango de fechas

$$R_i = \pi p_k, band_1, band_2, \dots, band_7 \sigma \text{ datecapture between } D_1 \text{ and } D_2 (R) \quad (17)$$

Ecuación 18. Partición de la Irradiancia por rango de fechas

$$I_i = \pi p_k, irradiationValue \sigma \text{ datecapture between } D_1 \text{ and } D_2 (I) \quad (18)$$

Ecuación 19. Cruce de los segmentos de Radiación e irradiancia.

$$Ds_i = \pi p_k, band_1, \dots, irradiationValue \sigma \text{ datecapture between } D_1 \text{ and } D_2 (R_i \times I_i) \quad (19)$$

Donde:

R representa la relación de reflectancia.

I representa la relación de irradiancia.

Pk representa la llave primaria, compuesta por los atributos (latitude, longitude) y la fecha de captura.

D₁ y **D₂** :corresponden a la fecha inicial y la fecha final respectivamente del intervalo de análisis

Ds_i : identifica el conjunto de datos asociado

El proceso de construcción del conjunto de características incluyó la descarga de imágenes satelitales, su integración para conformar la cobertura espacial del departamento de Nariño y la aplicación de técnicas de preprocesamiento orientadas a la reducción de ruido y la corrección de inconsistencias.

Posteriormente, la información fue transformada a un formato tabular, permitiendo la obtención de valores numéricos asociados al número digital de cada píxel, los cuales representan la reflectancia en distintas bandas espectrales.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Adicionalmente, se incorporaron datos de irradiancia recolectados desde la plataforma TriTier, basada en sensores terrestres, lo que permitió complementar la información satelital con mediciones in situ, de acuerdo con lo descrito en el **Marco Teórico (Sección B. Supuestos teóricos)**.

3. Etapa de preprocesamiento y transformación de datos:

En esta etapa, los archivos generados en formato CSV fueron procesados en un entorno de computación en la nube utilizando Google Colab, con el fin de gestionar de manera eficiente grandes volúmenes de información.

Inicialmente, se implementaron funciones de estandarización de los datos, incluyendo la normalización de nombres de columnas, la conversión de tipos de datos a formatos numéricos adecuados y la eliminación de valores nulos o inconsistentes.

Conjunto de datos de imágenes

La construcción de los conjuntos de características y de la variable objetivo a partir de los datos provenientes de los satélites Landsat y MODIS implicó un proceso orientado al tratamiento de vacíos en la cobertura espacial de los datos como se muestra en la fig. 13, los cuales dificultan la generación de representaciones espaciales continuas.

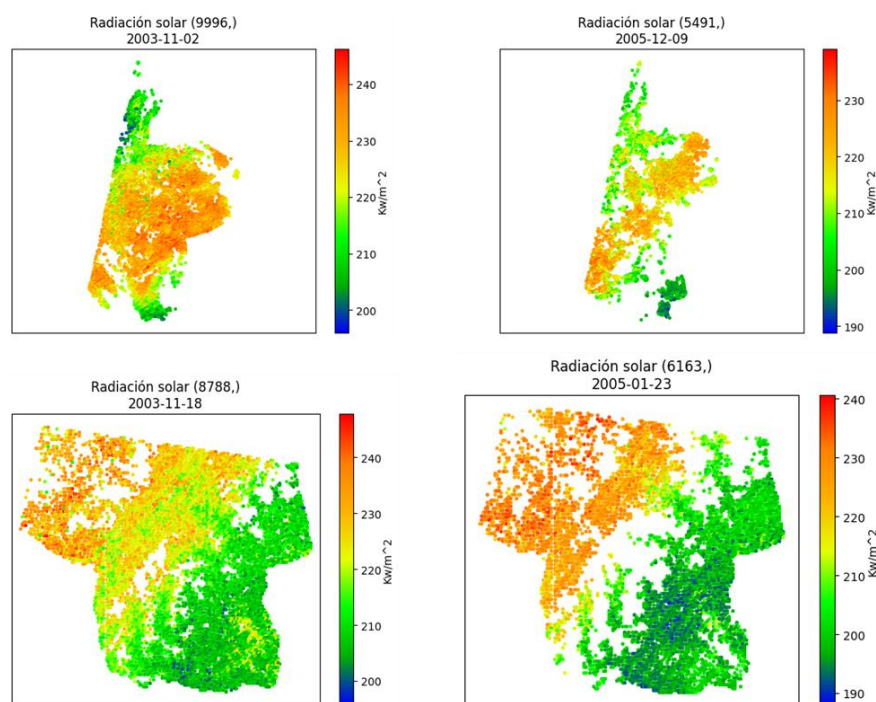


Fig. 12. Conjunto característico
Fuente [1]

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Para abordar esta problemática, se definió un enfoque basado en la generación de imágenes a partir de los valores de las bandas espectrales y su posterior modelado mediante técnicas de aprendizaje automático. En este proceso, cada banda espectral fue estimada a partir de un conjunto de variables compuesto por la latitud, la longitud y atributos temporales derivados de la fecha.

Estos atributos temporales fueron transformados mediante funciones trigonométricas con el fin de representar la naturaleza periódica del tiempo, como se muestra en las ecuaciones 20, 21 y 22.

Ecuación 20. Transformación de los meses que conforman a la fecha de captura.

$$monthSin = \sin \left(2\pi \frac{month}{12} \right), monthCos = \cos \left(2\pi \frac{month}{12} \right) \quad (20)$$

Ecuación 21. Transformación de los días que conforman a la fecha de captura.

$$daySin = \sin \left(2\pi \frac{day}{12} \right), dayCos = \cos \left(2\pi \frac{day}{12} \right) \quad (21)$$

Ecuación 22. Transformación de los años que conforman a la fecha de captura

$$yearNorm = \frac{year - (year)}{(year) - (year)} \quad (22)$$

Adicionalmente, se incluyeron las demás bandas espectrales como variables explicativas, con el propósito de capturar relaciones internas entre ellas. Con base en este planteamiento, se construyeron modelos capaces de aproximar el comportamiento espacial de las variables y generar valores en las zonas donde no existía información.

Para mejorar la continuidad espacial de las representaciones generadas, se aplicaron métodos de interpolación como Kriging. Asimismo, se implementaron estrategias de muestreo basadas en funciones sinusoidales con el fin de optimizar el proceso de generación de imágenes.

Finalmente, se generaron conjuntos de imágenes tanto para las variables de entrada (reflectancia) como para la variable objetivo (irradiancia), los cuales fueron organizados y preparados para su uso en los modelos de aprendizaje automático.

Con el fin de modelar la relación entre las variables espaciales, temporales y las bandas espectrales, se definió un enfoque basado en la aproximación de cada banda como una función de un conjunto de variables de entrada. De manera general, esta relación puede expresarse como:

Ecuación 23. expresión general usada en modelado matemático / Machine Learning

$$b_i \approx f(m) \mid m \subseteq B \quad (23)$$

Donde:

B: está conformado por la latitud, la longitud y las transformaciones temporales *monthSin*, *monthCos*, *daySin*, *dayCos* y *yearNorm*, las cuales se derivan de las ecuaciones (21), (22) y (23)

b_i corresponde a una de las siete bandas multiespectrales.

f : representa los algoritmos de Machine Learning.

A partir de este planteamiento, se estructuró un esquema de predicción secuencial de las bandas espectrales, en el cual cada banda es estimada utilizando como entradas variables previamente calculadas, permitiendo así la reconstrucción progresiva del conjunto completo de información espectral.

Para la implementación de este enfoque, se emplearon modelos de regresión como Support Vector Regression (SVR) y K-Nearest Neighbors (KNN), los cuales fueron entrenados para estimar las diferentes bandas espectrales a partir de las variables definidas. Los modelos generados fueron almacenados en archivos para su posterior reutilización en el proceso de generación de imágenes. Un resumen de los modelos implementados y sus configuraciones se presenta en el **Anexo B**.

4. Desarrollo del modelo de predicción

Para el segundo objetivo de la investigación, se creó un modelo de aprendizaje profundo para predecir la irradiancia a partir de secuencias de imágenes tomadas por satélites. Para la construcción de este modelo, se realizaron alrededor de 27 experimentos con el objetivo de evaluar diversas configuraciones y optimizar el rendimiento del modelo.

Durante estos experimentos, se ajustaron parámetros y arquitecturas, considerando métricas como el error cuadrático medio (MSE), utilizado como función de pérdida (*loss*), el error absoluto medio (MAE) y el coeficiente de determinación (R^2), con el fin de establecer criterios de evaluación del desempeño del modelo.

Inicialmente, se realizó la carga y organización de los datos previamente generados en formato CSV, los cuales contienen información geoespacial asociada a latitud, longitud,

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

bandas espectrales e irradiancia. A partir de estos datos se determinaron las coordenadas espaciales y la resolución de la malla, con el fin de estructurar adecuadamente la información.

En coherencia con la metodología planteada en la sección anterior, los datos tabulares fueron transformados a estructuras tipo imagen mediante la organización de los valores en matrices bidimensionales para cada banda espectral, generando así representaciones multicanal. De manera paralela, se construyeron matrices correspondientes a la irradiancia, las cuales constituyen la variable objetivo del modelo.

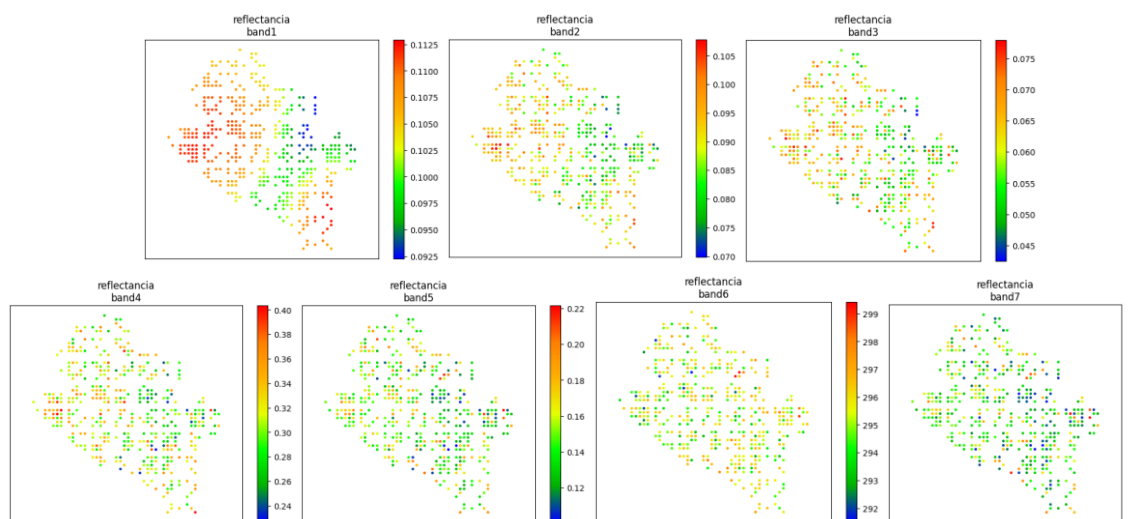


Fig. 13. Conjuntos característicos por bandas

Una vez generadas las imágenes, se realizó la carga masiva de los archivos, los cuales fueron organizados cronológicamente a partir de la fecha asociada a cada registro. Tanto las variables de entrada como la variable objetivo fueron normalizadas en un rango entre 0 y 1, con el fin de estandarizar los datos para el proceso de entrenamiento.

Dado que el problema presenta una naturaleza temporal, se implementó un enfoque basado en series de tiempo mediante la construcción de ventanas temporales. En este contexto, se definieron secuencias de imágenes satelitales correspondientes a instantes previos como entrada del modelo, mientras que la salida corresponde a la predicción de la irradiancia en el instante siguiente.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

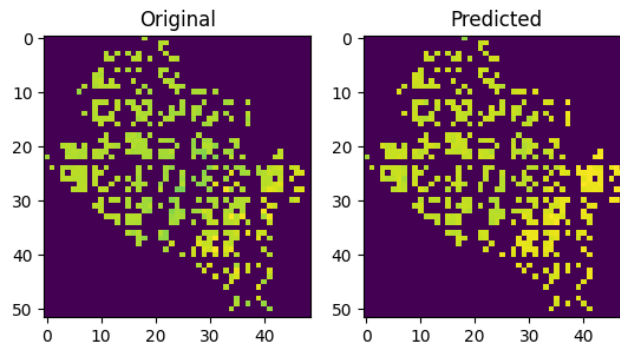


Fig. 14. Mapa de Nariño

Para el modelado de la serie espacio-temporal, se empleó una arquitectura basada en redes neuronales del tipo ConvLSTM2D, seleccionada debido a su capacidad para procesar secuencias de imágenes y capturar dependencias espaciales y temporales de manera conjunta. **tal como se describe en el Marco Teórico (Sección D. Definición nominal de las variables).**

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 15, 52, 49, 8)	0
conv_lstm2d (ConvLSTM2D)	(None, 52, 49, 32)	46,208
batch_normalization (BatchNormalization)	(None, 52, 49, 32)	128
dropout (Dropout)	(None, 52, 49, 32)	0
conv2d (Conv2D)	(None, 52, 49, 16)	4,624
batch_normalization_1 (BatchNormalization)	(None, 52, 49, 16)	64
dropout_1 (Dropout)	(None, 52, 49, 16)	0
conv2d_1 (Conv2D)	(None, 52, 49, 1)	145

Total params: 51,169 (199.88 KB)

Trainable params: 51,073 (199.50 KB)

Non-trainable params: 96 (384.00 B)

Fig. 15. Arquitectura del modelo de red neuronal

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

El modelo fue compilado utilizando el optimizador Adam y la función de pérdida basada en el error cuadrático medio, Asimismo, se definieron métricas como el error absoluto medio para su evaluación.

Durante el entrenamiento, se implementaron estrategias como la detección temprana (*early stopping*), la reducción adaptativa de la tasa de aprendizaje y el almacenamiento del mejor modelo.

Finalmente, el modelo fue evaluado utilizando un conjunto de datos independiente, mediante métricas como el error cuadrático medio (MSE) y el coeficiente de determinación (R^2). Adicionalmente, se consideró la comparación visual entre las imágenes de irradiancia reales y las predichas como criterio de evaluación cualitativa.

5. Evaluación y caracterización de modelos de Forecasting

Como parte del tercer objetivo de la investigación, se definió un enfoque de evaluación orientado a la caracterización de la relación entre las métricas de calidad de los modelos y el tiempo de inferencia.

Se definieron criterios de clasificación del desempeño de los modelos con base en los valores de las métricas. En este sentido, se consideró un nivel de desempeño alto cuando el coeficiente de determinación (R^2) es mayor o igual a 0.75 y el error cuadrático medio (MSE) es menor o igual a 0.0003, indicando un ajuste adecuado y bajo error. Un desempeño medio se estableció para valores de R^2 entre 0.60 y 0.74 y MSE menor o igual a 0.002, mientras que un desempeño bajo corresponde a valores de R^2 inferiores a 0.60 o MSE superiores a 0.002.

Finalmente, se definió un procedimiento orientado a la selección de configuraciones de modelos a partir del análisis conjunto de las métricas de calidad y el tiempo de inferencia, con el fin de establecer criterios para la elección del modelo más adecuado.

6. Despliegue

En la fase de despliegue, siguiendo la metodología CRISP-DM, se realizó la implementación del modelo con el fin de facilitar la visualización de sus resultados.

El experimento **3D-15W**, junto con el modelo generado y el *dataset* correspondiente para su entrenamiento, fue almacenado en un repositorio de GitHub (<https://github.com/LauraGomez25/IrradianciaForecasting>), lo que facilita su acceso, organización y reproducibilidad.

Posteriormente, se generaron productos geoespaciales en formato raster que representan la irradiancia mediante imágenes con referencia geográfica. Estos archivos fueron visualizados y verificados utilizando herramientas como Google Earth y visores compatibles con formato

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

GeoTIFF, con el fin de asegurar su correcta ubicación espacial y coherencia con el área de estudio.

Finalmente, los resultados se pueden ser consultados a través de un visor, lo que permite cumplir con la fase de implementación teniendo en cuenta que en los objetivos no se especifica su entrega sin embargo dando cumplimiento a la metodología se cuenta con ello.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

IV. RESULTADOS DE LA INVESTIGACIÓN

1. Componer un dataset de imágenes satelitales multiespectrales de reflectancia e irradiancia solar, para el entrenamiento de algoritmos de Forecasting aplicando técnicas de adquisición, preprocesamiento y transformación de datos.

El resultado de este objetivo es un **conjunto de datos** que permite analizar el comportamiento de la irradiancia en diferentes zonas del departamento de Nariño y facilita el desarrollo de modelos predictivos orientados al aprovechamiento del recurso solar.

A. Adquisición de base de datos Landsat y MODIS

En esta fase se presentan las características de los conjuntos de datos satelitales provenientes de Landsat y MODIS, utilizados en el desarrollo de la investigación. Estos conjuntos contienen información relacionada con la reflectancia y la irradiancia, organizada en diferentes bandas espectrales y resoluciones temporales.

En la **Tabla 2**, se detallan los elementos más relevantes de la base de datos, incluyendo el tamaño del conjunto de datos, la frecuencia de captura de imágenes, el número de registros asociados a la reflectancia e irradiancia, así como la cantidad de bandas espectrales visibles.

TABLA II. DESCRIPCIÓN DEL CONJUNTO DE DATOS

Base de datos	Tamaño (Gb)	Frecuencia (Díaz)	Filas reflectancia	Filas irradiancia	Bandas
Landsat	20	16	4.071.185	12.319.146	7
MODIS	70	1	373.670.476	335.405.312	7

B. Adquisición de datos satelitales

Como resultado del proceso de integración y segmentación de los datos, se generaron múltiples conjuntos de datos organizados en intervalos temporales, lo que permitió estructurar la información para su posterior análisis.

En la Tabla 3 se presentan las características generales del proceso de segmentación, incluyendo el número de tablas generadas, el periodo cubierto y el tiempo requerido para su procesamiento.

TABLA III. SEGMENTOS TEMPORALES

Elemento	Detalle
Dataset	Dividido en múltiples segmentos temporales
Número de tablas creadas	120
Satélite utilizado	Landsat

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

Periodo cubierto	2004 – 2014
Tiempo de procesamiento	Aproximadamente 3 horas y 20 minutos

A partir de este proceso, se obtuvieron diferentes datasets que contienen información relevante para el análisis de la irradiancia en el departamento de Nariño. En la Tabla 4 se describen los principales archivos generados, incluyendo sus características y tamaño.

TABLA IV. DATASETS INICIALES

Dataset	Tamaño	Características	Procesamiento
gridNariño.csv	4.6MB	Latitud, longitud	Coordenadas del departamento de Nariño
Landsat_all.csv	690.6MB	Latitud, longitud, 7 bandas, fechas	Coordenadas
Landsat_model.csv	66KB	Latitud, longitud, 7 bandas, value	Datos numéricos

La integración de los datos dio como resultado una estructura tabular compuesta por registros que contienen información de las bandas espectrales, coordenadas geoespaciales (latitud y longitud) y valores de irradiancia asociados. Esta estructura constituye la base para el entrenamiento de los modelos de predicción.

Con el fin de ilustrar la estructura final del conjunto de datos, en la Fig. 17 se presenta un ejemplo de los registros obtenidos tras el proceso de integración. En esta representación se observan las variables correspondientes a las coordenadas geoespaciales (latitud y longitud), las bandas espectrales (band1 a band7) y la variable objetivo asociada a la irradiancia.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

	latitude	longitude	band1	band2	band3	band4	band5	band6	band7	value
0	-8789850.0	184950.0	0.105113	0.082716	0.052756	0.315520	0.132641	296.178607	0.044586	203.4
1	-8784900.0	174600.0	0.104840	0.085732	0.053527	0.271463	0.090247	295.969710	0.027292	198.3
2	-8775000.0	164700.0	0.103468	0.079147	0.048932	0.248619	0.078593	296.842689	0.022098	203.1
3	-8775000.0	169650.0	0.108849	0.090665	0.055715	0.338889	0.128227	296.123927	0.039798	199.1
4	-8775000.0	174600.0	0.096526	0.071028	0.042247	0.194788	0.054468	294.407491	0.017712	196.3
...	...	...	...	...	...	...	...	...	...	...
429	-8559900.0	154800.0	0.078925	0.066239	0.047267	0.270656	0.109293	289.724062	0.044993	237.5
430	-8559900.0	169650.0	0.063235	0.047407	0.031352	0.156130	0.061570	286.148731	0.025417	236.1
431	-8559900.0	174600.0	0.078075	0.063393	0.042869	0.230624	0.089819	290.305318	0.035603	234.4
432	-8554950.0	154800.0	0.088854	0.074230	0.049634	0.296663	0.132079	293.226199	0.048379	221.4
433	-8554950.0	169650.0	0.077477	0.069132	0.050686	0.301495	0.129975	290.123712	0.050914	237.6

434 rows × 10 columns

Fig. 16. Registros obtenidos del dataset inicial

C. Caracterización de las bandas del espectro electromagnético

Como resultado del proceso de transformación de las variables temporales, se obtuvo una representación estructurada que permite capturar el comportamiento periódico del tiempo. En particular, las variables correspondientes al mes y al día fueron transformadas mediante funciones sinusoidales, mientras que el año fue normalizado, facilitando su integración en los modelos de aprendizaje automático.

En la Fig. 18 se presenta un ejemplo de los datos generados para la serie temporal. En esta figura se observan las variables originales (fecha, mes, año y día) junto con sus respectivas transformaciones (month_sin, month_cos, day_sin, day_cos y year_norm), evidenciando la correcta codificación de la naturaleza cíclica de las variables temporales.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

	date	month	year	day	month_sin	month_cos	day_sin	day_cos	year_norm
0	2003-01-01	1	2003	1	5.000000e-01	0.866025	0.201299	0.979530	0.0
1	2003-01-04	1	2003	4	5.000000e-01	0.866025	0.724793	0.688967	0.0
2	2003-01-07	1	2003	7	5.000000e-01	0.866025	0.988468	0.151428	0.0
3	2003-01-10	1	2003	10	5.000000e-01	0.866025	0.897805	-0.440394	0.0
4	2003-01-13	1	2003	13	5.000000e-01	0.866025	0.485302	-0.874347	0.0
...	...	...	...	...	...	...	...	...	...

Fig. 17. Fechas generadas para la serie temporal

A partir de esta representación, se estructuró el conjunto de datos para el modelado de las bandas espectrales. En la Tabla 5 se muestra la organización de las variables, donde se define el conjunto de características utilizado y la variable objetivo correspondiente a cada banda.

TABLA V. CONJUNTO DE DATOS

Conjunto de datos	Conjunto objetivo	Conjunto de características
Landsat	b_1	<i>tiempo, latitud, longitud</i>
	b_2	b_1
	b_3	b_1, b_2
	b_4	b_1, b_2, b_3
	b_5	b_1, b_2, b_3, b_4
	b_6	$b_1, b_2, b_3, b_4, b_5, b_7$
	b_7	b_1, b_2, b_3, b_4, b_5

Como se observa en la Tabla 5, la estimación de las bandas espectrales se realiza de manera secuencial, donde cada banda es predicha utilizando como entrada las variables temporales, espaciales y las bandas previamente estimadas. Este enfoque permite construir progresivamente el conjunto completo de información espectral, manteniendo coherencia entre las variables.

D. Aplicación del algoritmo kriging

Tras la generación de los conjuntos de imágenes correspondientes tanto para el conjunto objetivo (Irradiancia) y conjunto de características (Reflectancia), se llevó a cabo una evaluación visual preliminar con el fin de verificar la coherencia y precisión de las imágenes

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

generadas corresponden con la realidad. Esta etapa permitió identificar patrones espaciales iniciales y validar la consistencia de los datos frente a la distribución geográfica de Nariño.

Posteriormente, siguiendo las recomendaciones metodológicas planteadas en estudios previos [3] y [7] se aplicó el algoritmo de interpolación Kriging. Mediante este proceso fue posible completar el mapa del departamento de Nariño, generando un nuevo dataset en el que se almacena la información correspondiente a cada imagen.

En la **Fig 19** se muestran unos de los variogramas experimentales obtenidos a partir de los datos de irradiancia (puntos rojos), junto con el ajuste del modelo esférico (línea continua). Estas gráficas permiten visualizar la estructura de dependencia espacial.

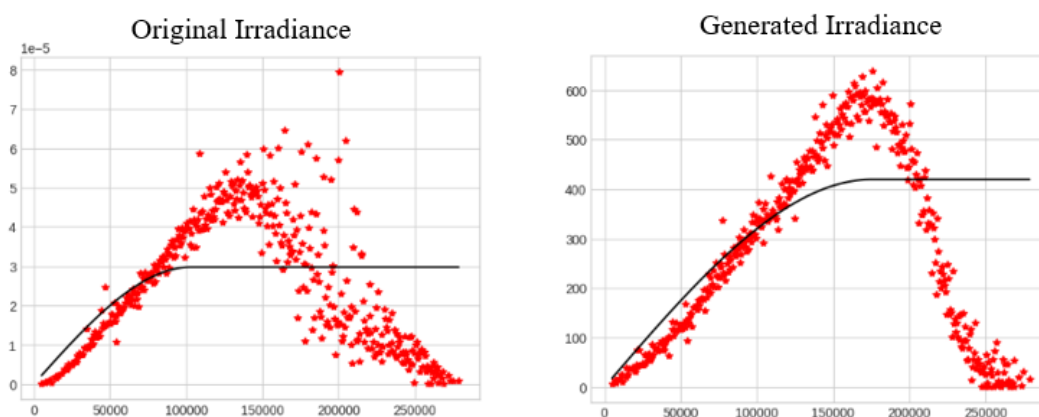


Fig. 18. Variogramas aplicados a las muestras de irradiancia.

El proceso de interpolación no solo facilitó la construcción de imágenes más representativas de la variabilidad espacial de la irradiancia, sino que también generó insumos fundamentales para el análisis posterior del modelo. En la **Fig.20** se ven 2 representaciones del comportamiento espacial en el área de estudio.

En la parte izquierda de la figura se observa la distribución de los datos originales de irradiancia, representados mediante puntos georreferenciados. cada punto muestra la ubicación específica registrada en el dataset, mientras que la escala de colores indica la intensidad de irradiancia medida en cada punto.

En la imagen ubicada a la derecha muestra el resultado obtenido después de aplicar el método de interpolación espacial mediante Kriging. Esta técnica nos permitió estimar los valores de irradiancia en las zonas donde no existen mediciones directas, generando una superficie continua que representa de manera más completa la variabilidad espacial. De esta manera, la interpolación espacial contribuye a mejorar la representación de los datos y nos permite usar las imágenes más completas que posteriormente serán utilizadas para el entrenamiento del modelo de predicción.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

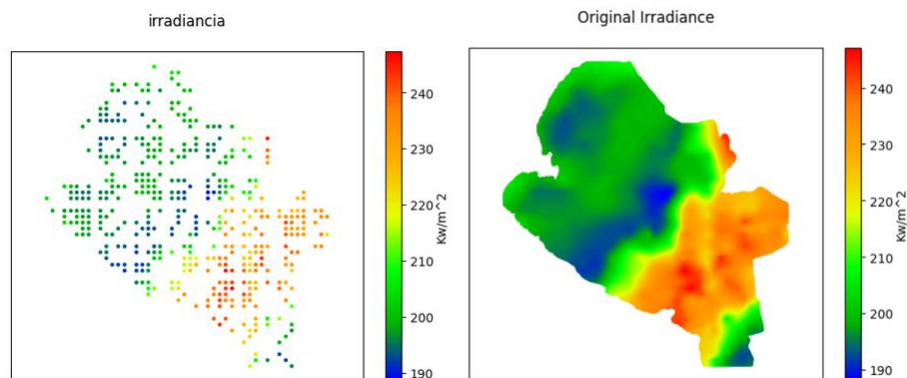


Fig. 19. Representación visual del proceso de interpolación

E. Imágenes interpoladas

Como resultado del procesamiento de los datos, se obtuvieron imágenes multiespectrales que integran la información correspondiente a las siete bandas espectrales junto con la variable de irradiancia. Estas representaciones se estructuran en matrices de tamaño 52×49 , lo que permite preservar la distribución espacial de las variables en el área de estudio.

En las Figuras 21 y 22 se presentan los resultados obtenidos del proceso de generación de imágenes. En la Figura 21 se evidencia la estructura espacial de las bandas del espectro electromagnético, mientras que en la Figura 22 se observa la organización de los archivos generados para cada fecha del periodo de estudio, reflejando la consistencia temporal del conjunto de datos.

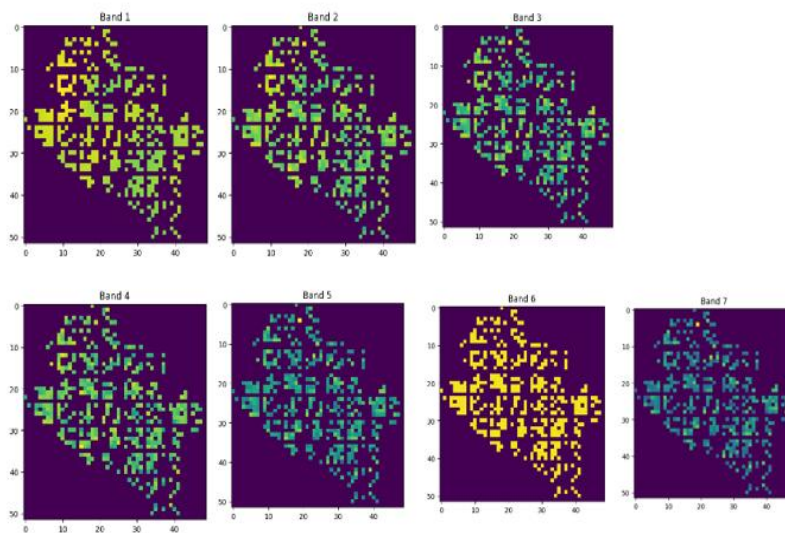
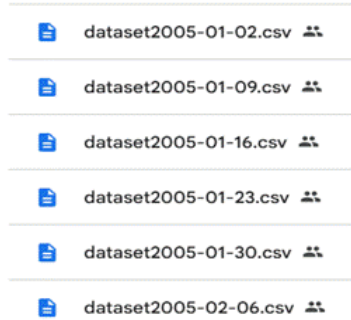


Fig. 20. Representación cada banda del espectro electromagnético

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.



dataset2005-01-02.csv
dataset2005-01-09.csv
dataset2005-01-16.csv
dataset2005-01-23.csv
dataset2005-01-30.csv
dataset2005-02-06.csv

Fig. 21. Archivos CSV generados para cada fecha del periodo de estudio

Como resultado del procesamiento de los datos, se obtuvieron 1446 registros limpios y organizados, garantizando la calidad del conjunto de datos utilizado en los experimentos posteriores.

2. Diseñar un conjunto de experimentos de entrenamiento de algoritmos de Forecasting para la adquisición de métricas de calidad de predicción y velocidad de procesamiento a partir de un dataset de imágenes satelitales multiespectrales.

F. División del conjunto de datos

El dataset previamente descrito fue dividido inicialmente en dos subconjuntos: entrenamiento (80%) y prueba (20%). Posteriormente, el conjunto de entrenamiento fue nuevamente subdividido para incorporar un proceso de validación, quedando distribuido en 64% para entrenamiento, 16% para validación y manteniendo el 20% para prueba.

TABLA VI. CONJUNTO DE DATOS

División del dataset	
Entrenamiento	Prueba
80%	20%

TABLA VII. SUBCONJUNTO DE DATOS

Validación	
Entrenamiento	64%
Validación	16%
Prueba	20%

G. Creación de ventanas temporales

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

La Tabla 7 presenta las diferentes configuraciones de ventanas temporales evaluadas durante los experimentos. Se observa que los modelos utilizaron secuencias de entre 3 y 15 imágenes satelitales como entrada.

Estas configuraciones permiten generar distintos conjuntos de entrenamiento, variando el número de secuencias y las dimensiones de entrada. A partir de estas pruebas, se analizaron posteriormente los resultados en términos de desempeño predictivo.

TABLA VIII. VENTANAS TEMPORALES

Modelo	Parámetro			
	Ventana temporal	Forma de X_t	Forma de y_t	Secuencias
15D - W15	15 imágenes	(1133, 15, 52, 49, 7)	(1133, 52, 49)	1133
W3	3 imágenes	(1145, 3, 52, 49, 7)	(1145, 52, 49)	1145
W15	15 imágenes	(1136, 15, 52, 49, 7)	(1136, 52, 49)	1136
3D - 15W	15 imágenes	(1146, 15, 52, 49, 8)	(1146, 52, 49)	1146

H. Métricas de Calidad

La Tabla IX presentó las métricas de calidad obtenidas para los modelos evaluados.

Entre los modelos analizados, el modelo **3D - 15W** fue seleccionado como el modelo final, debido a que presentó el mejor desempeño. Este modelo obtuvo un MSE de 0.00021, un MAE de 0.00999 y un R^2 de 0.7721, lo que indica que logra explicar aproximadamente el 77% de la variabilidad de los datos de irradiancia.

Estos resultados evidenciaron que este modelo generó las predicciones más precisas, al presentar los valores de error más bajos y el mayor coeficiente de determinación en comparación con los demás modelos evaluados.

TABLA IX. COMPARACIÓN DE MÉTRICAS DE CALIDAD

Modelo	Métrica			
	MSE (loss)	MAE	R^2	Tiempo de inferencia
15D - W15	0.00254	0.01757	0.72016	60 seg
W3	0.00239	0.01574	0.7319	186 seg
W15	0.00248	0.01717	0.7300	585 seg
3D - 15W	0.00021	0.00999	0.7721	750 seg

Como complemento al análisis cuantitativo, la Fig. 23 mostró el comportamiento del modelo durante el entrenamiento y validación. En esta gráfica, la línea azul representa la pérdida de entrenamiento, mientras que la línea naranja representa la pérdida de validación. Se observó

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

que ambas curvas presentaron un comportamiento estable, lo que indicó un adecuado proceso de aprendizaje y una buena capacidad de generalización del modelo.

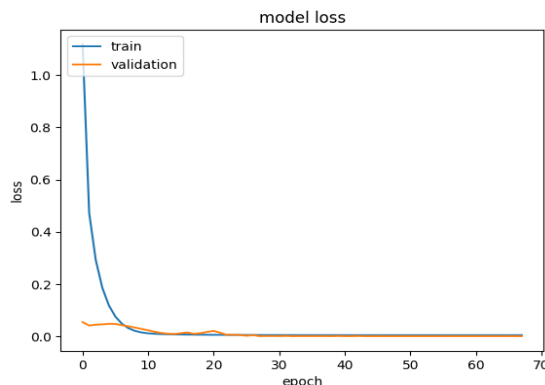


Fig. 22. . Gráfica del comportamiento del modelo

Línea azul: representa la pérdida de entrenamiento que indica el error del modelo al aprender a partir de los datos utilizados durante el entrenamiento en cada época.

Línea naranja: representa la pérdida de validación que muestra el error del modelo al evaluarse sobre un conjunto de datos de validación que no fue utilizado durante el entrenamiento.

Posterior a la evaluación cuantitativa se procedió a examinar la coherencia visual de las predicciones, en ese sentido, en la Fig 24 se observa la última imagen generada del mapa del departamento de Nariño mediante el último registro del conjunto de características para el conjunto de pruebas. En estas representaciones se evidencia que ambos mapas presentan una distribución espacial muy similar, lo cual indicó que el modelo ha logrado capturar correctamente los patrones de irradiancia en las diferentes zonas del área de estudio.

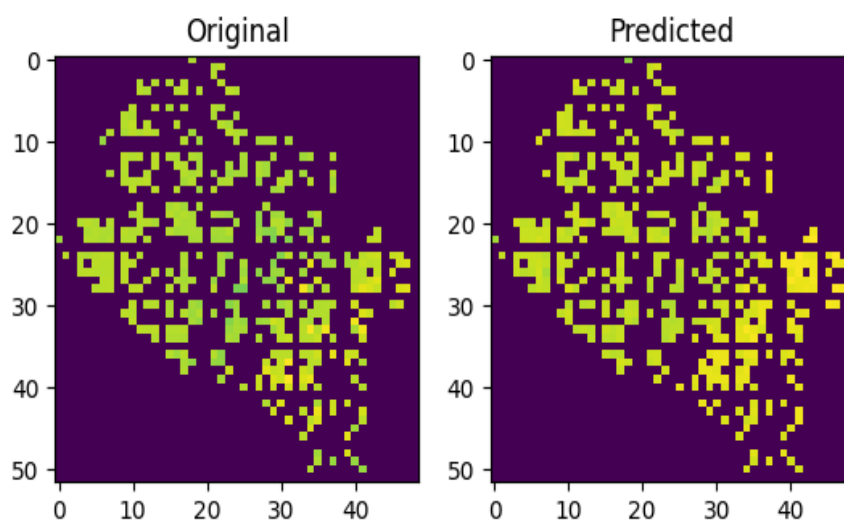


Fig. 23. Mapa de Nariño

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

De igual manera, la Fig. 25 se muestra los mapas de la Fig 24 generados en falso color con los rangos expresados en la literatura [2][1][4], en el cual se identifican diferentes zonas de irradiancia expresadas con los colores azul, verde, amarillo, naranja y rojo, determinando grupos de irradiancia baja, media baja, media, media alta y alta respectivamente.

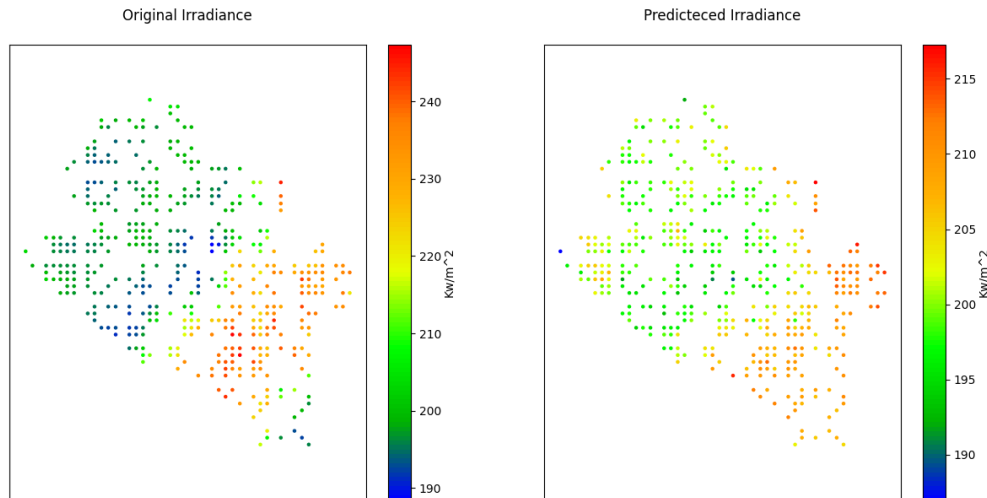


Fig. 24. Mapa de calor de Nariño

En conjunto, esta **comparación visual respalda los resultados numéricos** obtenidos anteriormente, confirmando la **alta capacidad de predicción y generalización** del modelo propuesto.

I. Interpolacion kriging

Posteriormente, las predicciones obtenidas fueron transformadas nuevamente a un rango continuo de valores para inicializar la interpolación como lo muestra la Fig. 26

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

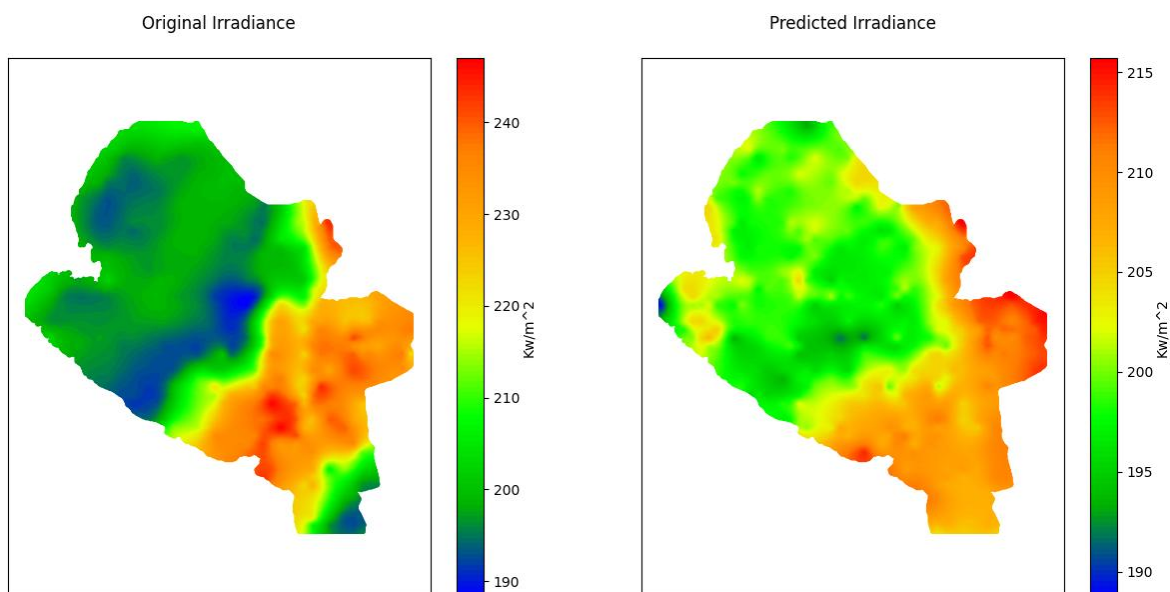


Fig. 25. Imagen de algoritmo de Interpolación

J. Mapa final

Finalmente, convertimos este mapa de irradiancia en una imagen en formato tipo “RASTER”, para así lograr obtener el mapa de Nariño con la respectiva información de irradiancia como se muestra en la Fig. 27

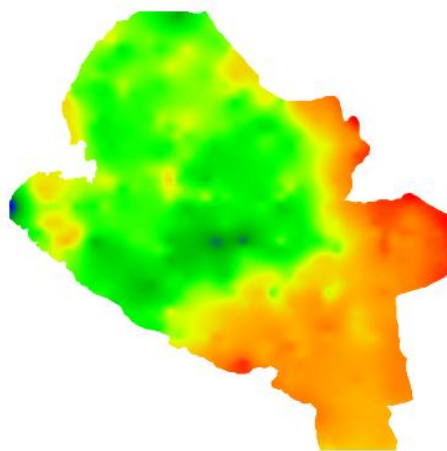


Fig. 26. Mapa de la irradiancia predicha para Nariño en formato RASTER en formato PNG

Este ráster se generó en formato png cómo se muestra en la Fig 27. No obstante, también se generó en formato .tif en proyecciones de Mercator 3857 cuyo despliegue se realizó en la aplicación geotif cómo se muestra en la Fig 28.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

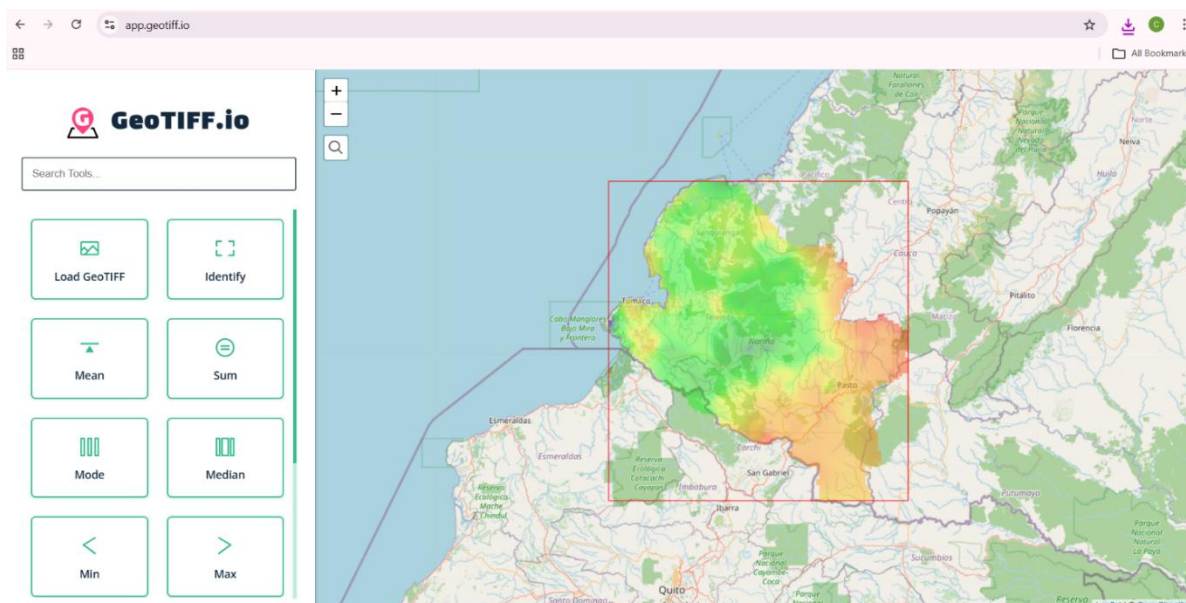


Fig. 27. Mapa de la irradiancia predicha para Nariño en formato RASTER .tif proyecciones de Mercator 3857 en metros visualizada en la aplicación GeoTIFF.IO

De igual forma, se generó en formato .tif en proyecciones de Mercator 4326, el cual es el estándar más utilizado para visores comerciales como Google Earth y Google Maps. Para Google Earth se convirtió el raster a formato kmz ya que no se cuenta de momento con visualización directa de imágenes .tif, en la Fig 29 puede apreciarse el despliegue realizado en Google Earth.

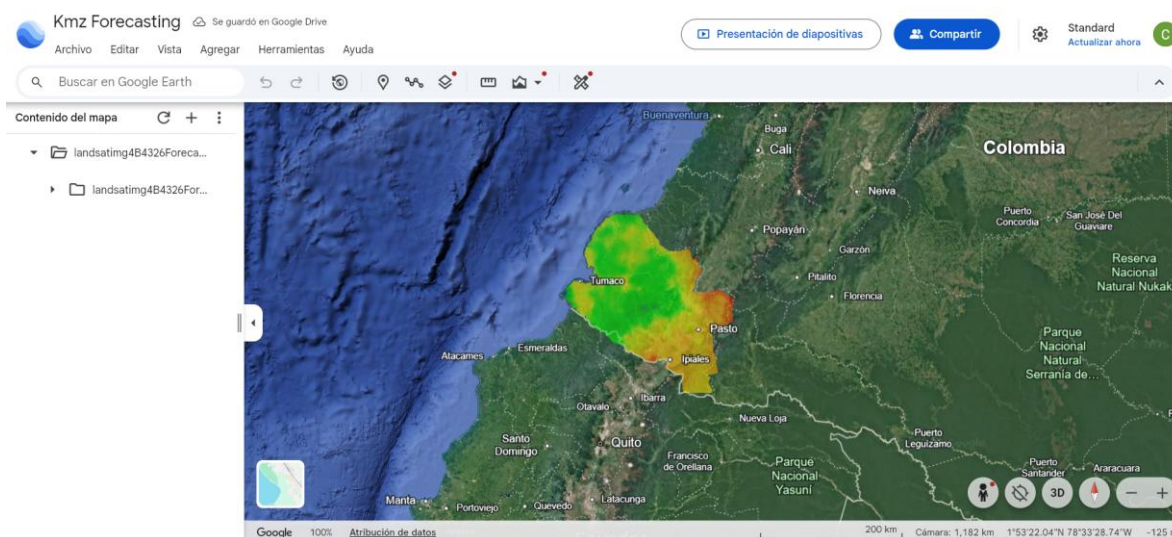


Fig. 28. Mapa de la irradiancia predicha para Nariño en formato RASTER .kmz en proyecciones de Mercator 4326 en ángulos visualizada en la aplicación Google Earth

Por último, se desplegó el raster .tif en coordenadas de Mercator 4326 en una aplicación web utilizando el API de Google maps como se muestra en la Fig 30.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

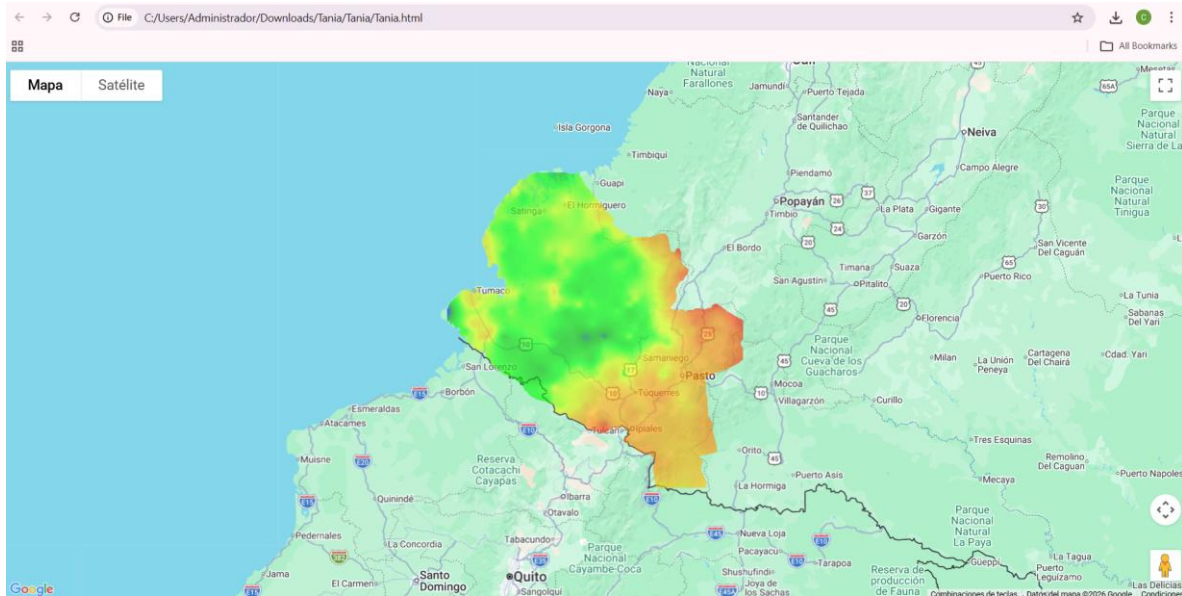


Fig. 29. Mapa de la irradiancia predicha para Nariño en formato RASTER .tif en proyecciones de Mercator 4326 en ángulos visualizada en la aplicación web con API de Google Maps

- 3. Clasificar la relación entre las métricas de calidad y el tiempo de inferencia de los experimentos realizados para determinar los mejores balances de cada configuración mediante una caracterización de algoritmos de Forecasting recomendada por expertos.**

K. COMPARACIÓN EXPERIMENTOS

Para el análisis de los experimentos se consideraron varios aspectos fundamentales. En primer lugar, la frecuencia, definida como la cantidad de días en los que se recolectaron las imágenes. En segundo lugar, la ventana, que corresponde al número de días anteriores utilizados para realizar la predicción. En cuanto a las métricas de evaluación, se emplearon el error cuadrático medio (MSE) y el coeficiente de determinación (R^2) como indicadores de calidad del modelo.

En la Tabla 10 se presentan los resultados obtenidos para las diferentes configuraciones evaluadas, organizadas según la frecuencia y el tamaño de la ventana. En estas se observa el comportamiento de las métricas de desempeño para cada experimento, permitiendo realizar una comparación inicial entre las distintas combinaciones.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

TABLA X. EXPERIMENTOS REALIZADOS

Frecuencia	Ventana	Datos/Modelo	MSE	R2	Tiempo de inferencia
3d	15	1446	0.0002	0.77	750 seg
W(semanas)	3	1145	0.002	0.73	186 seg
W(semanas)	15	1136	0.002	0.73	585 seg
15d	15	1133	0.002	0.70	60 seg
W(semanas)	7	1144	0.002	0.69	16.560 seg
7d	15	1238	0.0002	0.66	468 seg
7d	3	5006	0.0002	0.65	370 seg
3d	3	1458	0.0003	0.64	9.300 seg
3d	7	1424	0.0002	0.61	14.346 seg
15d	3	1148	0.002	0.5	136 seg
15d	7	1148	0.0005	0.21	5.752 seg

A partir de estos resultados tabulares, se construyeron representaciones gráficas con el fin de facilitar la interpretación y comparación de las métricas R^2 y MSE. Las gráficas evidencian una variación significativa en el desempeño de las configuraciones evaluadas, mostrando de manera más clara las tendencias presentes en los datos.

En la gráfica de la Fig 31 se observa una variación significativa en el desempeño de las configuraciones evaluadas. En particular, los mejores resultados en términos de R^2

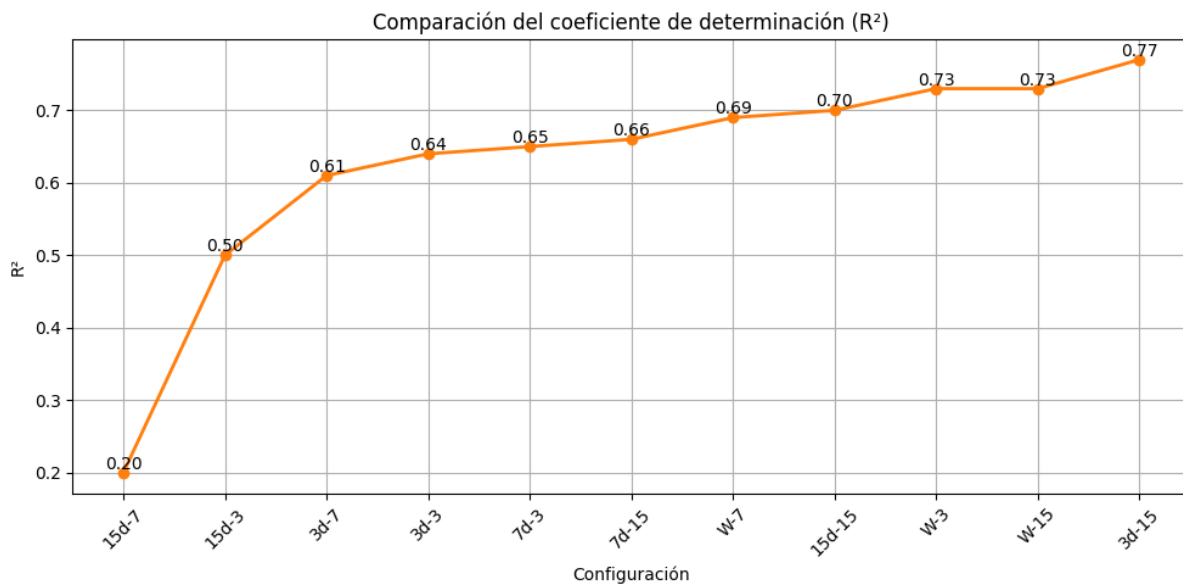


Fig. 30. Grafica comportamiento de R^2

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

En la gráfica de la Fig 32 del MSE, se observa que los modelos con mejor desempeño tienden a presentar errores más bajos, lo que confirma la consistencia entre ambas métricas.

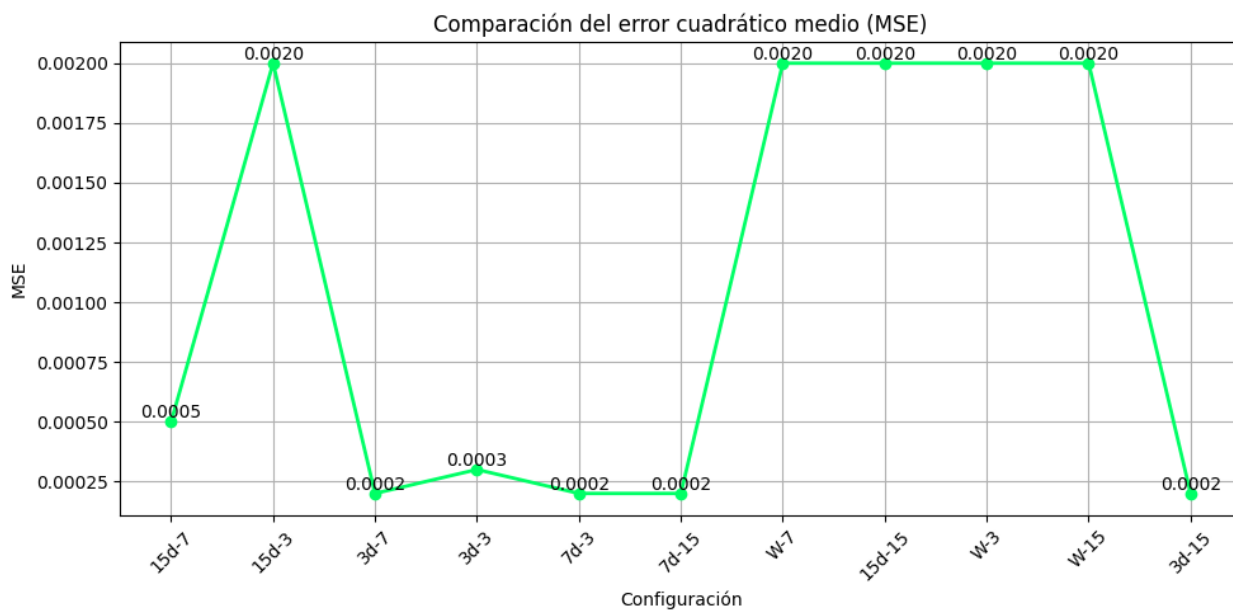


Fig. 31. Grafica comportamiento MSE

De manera conjunta, el análisis permite identificar las configuraciones con mejor balance entre ambas métricas. En este sentido, la configuración correspondiente a una frecuencia de 3 días y una ventana de 15 instantes se destaca como la más adecuada, al presentar el mayor valor de R^2 junto con un MSE reducido.

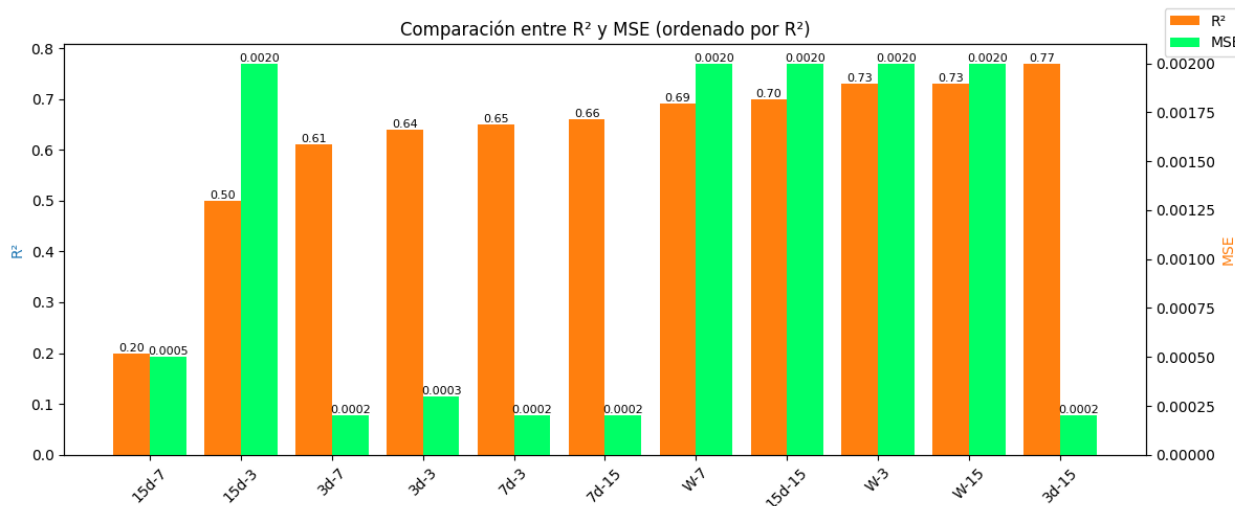


Fig. 32. Gráfica comparativa del R^2 y MSE

V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

A partir de los resultados obtenidos, se evidencio que la implementación de modelos de Deep Learning aplicados a imágenes satelitales multiespectrales conforma una estrategia efectiva para la predicción de irradiancia solar en el departamento de Nariño. Particularmente, la arquitectura ConvLSTM2D demostró un desempeño superior frente a otras configuraciones evaluadas, lo cual valida el enfoque metodológico planteado en la investigación

El modelo con mejor rendimiento, corresponde a una frecuencia de 3 días y una ventana temporal de 15 imágenes (3D - 15W), este alcanzó un coeficiente de determinación R^2 de 0.77, lo que indica que es capaz de explicar aproximadamente del 77% de la variabilidad de los datos de irradiancia. Este resultado ubica al modelo dentro de un rango de desempeño alto para este tipo de aplicación, donde valores de R^2 superiores a 0.75 suelen considerarse adecuados en escenarios de predicción de irradiancia solar basada en datos satelitales.

Este resultado concuerda con lo planteado por Kim et al. [9], quienes evidencian que la incorporación de información espacio-temporal en modelos de Deep Learning mejora significativamente la precisión en la predicción de irradiancia solar. De esta manera, los valores de error obtenidos (MSE de 0.00021 y MAE 0.00999) evidencian una desviación entre los valores reales y los predichos, lo que confirma la precisión del modelo.

El comportamiento de la configuración 3D – 15W puede estar asociado a las condiciones atmosféricas características del departamento de Nariño, donde debido a la nubosidad se presenta cierta persistencia temporal. En este contexto, una frecuencia de 3 días permite capturar de manera más adecuada las variaciones de corto plazo en la nubosidad características del departamento de Nariño, evitando la perdida de información que pueda presentarse en escalas mayores (W, 7, 15). Por su parte, la ventana de 15 imágenes identifica patrones espacio – temporales consistentes. Esta combinación permite representar mejor la dinámica atmosférica, reduciendo el ruido en los datos y mejorando la capacidad predictiva del modelo.

El análisis comparativo entre los diferentes experimentos visualizados en la Tabla.10 permite identificar que el desempeño del modelo tiene relación entre la frecuencia de los datos y el tamaño de la ventana temporal. En este sentido, las configuraciones que incorporan ventanas temporales más amplias logran capturar de manera más efectiva las dependencias espacio-temporales presentes en los datos, mejorando la capacidad predictiva. Este comportamiento coincide con lo expuesto por Velasco Pérez et al. [14], quien destaca que las arquitecturas que integran componentes convolucionales y recurrentes permiten modelar de manera más precisa la dinámica espacio-temporal de la irradiancia.

Se observa concordancia entre estos resultados y la evidencia reportada en investigaciones previas a nivel regional. Mora [2] demostró que el uso de imágenes satelitales provenientes

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

de Landsat y MODIS, combinadas con técnicas de Machine Learning, mejora la precisión en la estimación de oferta energética. De manera similar, Cabrera et al. [4] evidencio que la aplicación de modelos de regresión sobre las imágenes satelitales permitió obtener aproximaciones confiables de la irradiancia solar en zonas rurales. Sin embargo, el análisis se centró en la evaluación de arquitecturas basadas en Deep Learning, particularmente en variaciones del modelo ConvLSTM2D, con el fin de explorar su capacidad para modelar la dinámica espacio – temporal de la irradiancia.

En este sentido, los resultados obtenidos evidencian que este tipo de arquitecturas logran representar de manera efectiva la variabilidad de la irradiancia, lo cual consiste con lo reportado en la literatura, donde los modelos de Deep Learning han mostrado ventajas frente a enfoques tradicionales en problemas que involucran componentes espaciales y temporales.

Asimismo, se evidencia que la construcción y el preprocesamiento de los datasets constituyeron un factor decisivo en el desempeño del modelo, La integración de los datos provenientes del satélite Landsat, junto con la aplicación de técnicas como normalización, transformación de variables temporales y generación de imágenes mediante interpolación (Kriging) permitió optimizar la calidad de los datos, reduciendo así el impacto de la dispersión y la ausencia de datos en ciertas zonas geográficas, garantizando un entrenamiento más estable.

Desde una perspectiva práctica, los resultados obtenidos representan un aporte significativo para la planificación energética, un modelo con un nivel de precisión del 77% permite estimar de manera confiable la distribución espacial de la irradiancia solar, lo cual puede emplearse para identificar zonas óptimas destinadas a la instalación de sistemas fotovoltaicos.

De manera complementaria, la visualización geoespacial de los resultados permitió evidenciar que la irradiancia presenta una distribución heterogénea en el departamento de Nariño. Asimismo, la correspondencia entre la imagen generada y la ubicación geográfica real sugiere que el modelo captura adecuadamente los patrones espaciales, facilitando la identificación de zonas con diferentes niveles de irradiancia.

No obstante, la investigación presenta algunas limitaciones que deben ser consideradas. La disponibilidad y calidad de los datos influyen directamente en el rendimiento del modelo, por lo que un mayor volumen de datos podría mejorar su capacidad de generalización. Asimismo, el procesamiento de aproximadamente 90GB de datos requirió el uso de infraestructura en la nube, específicamente Google Colab, lo que evidencia una dependencia de recursos computacionales elevados. Esta condición, junto con las restricciones del entorno utilizado, limitaron la exploración de arquitecturas más complejas. Finalmente, ampliar el rango temporal del estudio permitirá capturar patrones de largo plazo asociados a fenómenos climáticos.

En conjunto, los resultados obtenidos permiten validar la hipótesis de investigación, demostrando que los algoritmos de Forecasting basados en Deep Learning logran explicar

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

significativamente la variabilidad de los datos de irradiancia en comparación con modelos tradicionales. En particular la arquitectura ConvLSTM2D se consolida como una de las más adecuadas para el modelado de series de tiempo espacio-temporales, aportando una solución robusta y escalable para la predicción del recurso solar.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

CONCLUSIONES

Se construyó un dataset de imágenes satelitales multiespectrales a partir de la información proveniente de los satélites landsat y Modis, integrando datos de reflectancia e irradiancia solar mediante procesos de adquisición, preprocesamiento y transformación. Este conjunto de datos permite representar adecuadamente el comportamiento espacio-temporal de la irradiancia en el departamento de Nariño.

Se determinó que la arquitectura ConvLSTM2D es el modelo más adecuado para la predicción de la irradiancia solar dentro de las configuraciones evaluadas. Particularmente, la combinación de una frecuencia de 3 días y una ventana temporal de 15 imágenes alcanzó el mejor desempeño, evidenciando una elevada capacidad predictiva.

Se estableció que la configuración de los datos, específicamente la frecuencia de captura y el tamaño de la ventana, influye directamente en el rendimiento de los modelos de Forecasting, siendo las ventanas temporales más amplias las que permiten obtener mejores resultados.

El comportamiento de los experimentos se evaluó mediante métricas de calidad, identificando distintos niveles de desempeño (alto, medio y bajo), lo que facilitó determinar el mejor equilibrio entre precisión y estabilidad en las predicciones.

Se comprobó que la aplicación de técnicas de preprocesamiento y transformación de datos contribuye significativamente a mejorar la calidad de la información y el desempeño de los modelos de aprendizaje automático.

Finalmente, se validó la hipótesis de investigación, demostrando que los modelos de forecasting basados en Deep learning permitieron predecir la irradiancia solar con un alto nivel de precisión, consolidándose como una herramienta viable para apoyar la planificación energética y aprovechamiento de fuentes renovables en el departamento de Nariño.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

RECOMENDACIONES

Se recomienda validar el rendimiento de la metodología propuesta mediante el uso de fuentes de datos adicionales, como imágenes satelitales multiespectrales como SENTINEL o LIDIAR, con el fin de evaluar la robustez del modelo frente a diferentes resoluciones espaciales y características espectrales.

Asimismo, es pertinente ampliar el conjunto de métricas de evaluación, incorporando indicadores que permitan un análisis tanto técnico como visual de los resultados obtenidos.

Adicionalmente, se propone evaluar el grado de generalización del modelo mediante la comparación entre mapas de irradiancia actuales y los generados por la metodología, con el objetivo de validar su capacidad de adaptación en distintos escenarios.

Como línea de trabajo futuro, se plantea analizar el desempeño de algoritmos de vanguardia, tales como modelos basados en mecanismos de atención, redes convolucionales (TCN) y redes neuronales de grafos (GNN), con el propósito de medir su capacidad predictiva en este contexto.

De igual forma, se sugiere evaluar que tan claro y comprensible es el funcionamiento del modelo en la generación de mapas de irradiancia, permitiendo identificar los factores que fluyen en sus predicciones.

Finalmente, se recomienda comparar el desempeño del modelo frente a conjuntos de datos sintéticos de mayor resolución, con el fin de analizar su comportamiento en escenarios más complejos.

BIBLIOGRAFÍA

- [1] D.-M. Pachajoa, H. Mora-Paz, y D. Mayorca-Torres, «Comparison of Kernel Functions in the Classification of Irradiance Zones from Multispectral Satellite Images», *Revista Facultad de Ingeniería*, vol. 30, n.º 58, Art. n.º 58, dic. 2021, doi: 10.19053/01211129.v30.n58.2021.13845.
- [2] H. A. Mora-Paz, «Comparativo de kernels sobre predicción de oferta de fuentes alternativas de energía», masterThesis, 2019. Accedido: 2 de abril de 2025. [En línea]. Disponible en: <https://reunir.unir.net/handle/123456789/10020>
- [3] «spanish_paris_agreement.pdf». Accedido: 5 de mayo de 2025. [En línea]. Disponible en: https://unfccc.int/sites/default/files/spanish_paris_agreement.pdf
- [4] 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.
- [5] D. Assouline, N. Mohajeri, y J.-L. Scartezzini, «Quantifying rooftop photovoltaic solar energy potential: A machine learning approach», *Solar Energy*, vol. 141, pp. 278-296, ene. 2017, doi: 10.1016/j.solener.2016.11.045.
- [6] J. H. Jungclaus *et al.*, «Climate and carbon-cycle variability over the last millennium», *Climate of the Past*, vol. 6, n.º 5, pp. 723-737, oct. 2010, doi: 10.5194/cp-6-723-2010.
- [7] S. S. Estrella Restrepo, F. V. Botina Pineda, y D. C. Cuarán Ruano, «Energía renovable en Nariño, posibilidades de la región», *Travesía Emprendedora*, vol. 8, n.º 2, pp. 15-27, 2024.
- [8] «Energías Renovables». Accedido: 11 de abril de 2025. [En línea]. Disponible en: <https://www.metropol.gov.co:443/ambiental/paginas/consumo-sostenible/energias-renovables.aspx>
- [9] M. Attya, O. M. Abo-Seida, H. M. Abdulkader, y A. M. Mohammed, «Advanced solar radiation prediction using combined satellite imagery and tabular data processing», *Sci Rep*, vol. 15, n.º 1, p. 14035, abr. 2025, doi: 10.1038/s41598-025-96109-0.
- [10] M. D. N. Verdezoto, D. A. F. Guerra, J. F. C. Tipan, y J. C. P. Vanegas, «Aplicación de un algoritmo de optimización basado en aprendizaje automático en la predicción de la generación fotovoltaica: Application of a machine learning based optimization algorithm in the prediction of photovoltaic generation», *Boletín Científico Ideas y Voces*, vol. 4, n.º 2, p. ág. 56-71, ago. 2024, doi: 10.60100/bciv.v4i2.144.
- [11] D. S. Kumar, G. M. Yagli, M. Kashyap, y D. Srinivasan, «Solar irradiance resource and forecasting: a comprehensive review», *IET Renewable Power Generation*, vol. 14, n.º 10, pp. 1641-1656, 2020, doi: 10.1049/iet-rpg.2019.1227.
- [12] P. Kumari y D. Toshniwal, «Deep learning models for solar irradiance forecasting: A comprehensive review», *Journal of Cleaner Production*, vol. 318, p. 128566, oct. 2021, doi: 10.1016/j.jclepro.2021.128566.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

- [13] C. A. Arango Navarrete, «Optimización del preprocesamiento de imágenes satelitales para modelos de predicción de irradiancia solar», jul. 2025, Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://hdl.handle.net/1992/77061>
- [14] «Repositorio Institucional Séneca», Repositorio Institucional Séneca. Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://repositorio.uniandes.edu.co/repositorio.uniandes.edu.co>
- [15] 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, p. 55, ene. 2016, doi: 10.3390/en9010055.
- [16] «Las energías renovables como una alternativa para bajar costos y cuidar el medio ambiente». Accedido: 1 de mayo de 2025. [En línea]. Disponible en: <https://www.uac.edu.co/las-energias-renovables-como-una-alternativa-para-bajar-costos-y-cuidar-el-medio-ambiente/>
- [17] «a07.pdf». Accedido: 11 de abril de 2025. [En línea]. Disponible en: https://sisbib.unmsm.edu.pe/bibvirtualdata/publicaciones/geologia/vol10_n19/a07.pdf
- [18] J. M. R. Gómez, F. Carlesso, L. E. Vieira, y L. Da Silva, «A irradiância solar: conceitos básicos», *Rev. Bras. Ensino Fís.*, vol. 40, p. e3312, mar. 2018, doi: <https://doi.org/10.1590/1806-9126-RBEF-2017-0342>.
- [19] «ESA - Space for Kids - Los satélites». Accedido: 22 de abril de 2025. [En línea]. Disponible en: https://www.esa.int/kids/es/Aprende/Tecnologia/El_espacio_util/Los_satelites
- [20] dm2012, «La importancia de los satélites en nuestra vida cotidiana», importancia. Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://www.importancia.cc/satelites/>
- [21] geosolutions, «Imágenes Satelitales - Conceptos Básicos», Geosolutions Consulting. Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://www.geosolutionsconsulting.com/tecnologia/conceptos-basicos-de-las-imagenes-satelitales/>
- [22] «EarthExplorer». Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://earthexplorer.usgs.gov/>
- [23] «¿Qué es la teledetección y para qué se utiliza? | Servicio Geológico de Estados Unidos». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used>
- [24] N. Earth Science Data Systems, «Earth Observation Data Basics | NASA Earthdata». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://www.earthdata.nasa.gov/learn/earth-observation-data-basics>
- [25] «7. Preguntas Frecuentes — documentación de Semi-Automatic Classification Plugin - 5.3.6.1». Accedido: 1 de mayo de 2025. [En línea]. Disponible en: <https://semiautomaticclassificationmanual-v5.readthedocs.io/es/latest/faq.html>
- [26] «MODIS Web». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://modis.gsfc.nasa.gov/about/>

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

- [27] «Introducción a la teledetección / 2. Sistemas espaciales de teledetección». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://static.uvq.edu.ar/mdm/teledeteccion/unidad-2.html>
- [28] «Misiones Landsat | Servicio Geológico de Estados Unidos». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://www.usgs.gov/landsat-missions>
- [29] «Landsat Collection 2 Level-1 Data | U.S. Geological Survey». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://www.usgs.gov/landsat-missions/landsat-collection-2-level-1-data>
- [30] «Algoritmos de Landsat | Google Earth Engine», Google for Developers. Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://developers.google.com/earth-engine/guides/landsat?hl=es-419>
- [31] «What Is Three-Tier Architecture? | IBM». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://www.ibm.com/think/topics/three-tier-architecture>
- [32] «Algoritmos de aprendizaje automático (artículo)», Khan Academy. Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://es.khanacademy.org/computing/ap-computer-science-principles/data-analysis-101/x2d2f703b37b450a3%3Aa-machine-learning-and-bias/a/machine-learning-algorithms>
- [33] «Forecasting: What It Is, How It's Used in Business and Investing», Investopedia. Accedido: 11 de abril de 2025. [En línea]. Disponible en: <https://www.investopedia.com/terms/f/forecasting.asp>
- [34] L. Martín Pomares, «Análisis y predicción de series temporales de irradiancia solar global mediante modelos estadísticos», sep. 2012, Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://hdl.handle.net/20.500.14352/48281>
- [35] M. C. Manchado, «Predicción demanda eléctrica española. Implementación de redes neuronales recurrentes en Python.».
- [36] «Perceptrones multicapa en el aprendizaje automático: Guía completa». Accedido: 11 de abril de 2025. [En línea]. Disponible en: <https://www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning>
- [37] «Introducción a la memoria a corto-largo plazo (LSTM)». Accedido: 11 de abril de 2025. [En línea]. Disponible en: <https://la.mathworks.com/discovery/lstm.html>
- [38] «Explicación de la GRU (Unidad Recurrente Cerrada) | Ultralytics». Accedido: 14 de abril de 2025. [En línea]. Disponible en: <https://www.ultralytics.com/es/glossary/gated-recurrent-unit-gru>
- [39] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, y Y. Du, «Short-Term Photovoltaic Power Forecasting Based on Long Short Term Memory Neural Network and Attention Mechanism», *IEEE Access*, vol. 7, pp. 78063-78074, 2019, doi: 10.1109/ACCESS.2019.2923006.
- [40] I. El Naqa y M. J. Murphy, «What Is Machine Learning?», en *Machine Learning in Radiation Oncology: Theory and Applications*, I. El Naqa, R. Li, y M. J. Murphy, Eds., Cham: Springer International Publishing, 2015, pp. 3-11. doi: 10.1007/978-3-319-18305-3_1.

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

- [41] O. Consulting, «5 ejemplos de Machine Learning y cómo te ayudan en las ventas.», Outliers Consulting. Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://www.outliersconsulting.com/general/5-ejemplos-de-machine-learning-y-como-ayuda-a-tus-ventas/>
- [42] «Coeficiente de determinación», *Wikipedia, la enciclopedia libre*. 22 de febrero de 2024. Accedido: 25 de abril de 2025. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Coeficiente_de_determinaci%C3%B3n&oldid=158364923
- [43] «Dynamic time warping», *Wikipedia*. 11 de diciembre de 2024. Accedido: 25 de abril de 2025. [En línea]. Disponible en: https://en.wikipedia.org/w/index.php?title=Dynamic_time_warping&oldid=1262411913
- [44] «Uso de pronósticos para predecir resultados», Ayuda de Zendesk. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://support.zendesk.com/hc/es/articles/4408832082970-Uso-de-pron%C3%B3sticos-para-predecir-resultados>
- [45] «1.4. Support Vector Machines», scikit-learn. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://scikit-learn/stable/modules/svm.html>
- [46] «¿Qué es un bosque aleatorio? | IBM». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.ibm.com/es-es/think/topics/random-forest>
- [47] «¿Qué es el algoritmo de k vecinos más cercanos? | IBM». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.ibm.com/mx-es/think/topics/knn>
- [48] D. Mercado Polo, L. Pedraza Caballero, y E. Martínez Gómez, «Comparación de Redes Neuronales aplicadas a la predicción de Series de Tiempo», *Prospectiva*, vol. 13, n.º 2, pp. 88-95, jul. 2015, doi: 10.15665/rp.v13i2.491.
- [49] «Long Short-Term Memory: redes con memoria a corto-largo plazo», IONOS Digital Guide. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.ionos.com/es-us/digitalguide/paginas-web/desarrollo-web/long-short-term-memory/>
- [50] «Gated Recurrent Unit», DeepAI. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://deepai.org/machine-learning-glossary-and-terms/gated-recurrent-unit>
- [51] K. Team, «Keras documentation: ConvLSTM2D layer». Accedido: 11 de abril de 2026. [En línea]. Disponible en: https://keras.io/api/layers/recurrent_layers/conv_lstm2d/
- [52] «¿Qué es el preprocesamiento de datos para el aprendizaje automático? | Pure Storage». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.purestorage.com/es/knowledge/what-is-data-preprocessing.html>
- [53] M. B. Editor, «Análisis de Regresión: ¿Cómo Puedo Interpretar el R-cuadrado y Evaluar la Bondad de Ajuste?» Accedido: 26 de abril de 2025. [En línea]. Disponible en:

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

<https://blog.minitab.com/es/analisis-de-regresion-como-puedo-interpretar-el-r-cuadrado-y-evaluar-la-bondad-de-ajuste>

- [54] «What Is Support Vector Machine? | IBM». Accedido: 26 de abril de 2025. [En línea]. Disponible en: <https://www.ibm.com/think/topics/support-vector-machine>
- [55] F. Hernández, *4 Random Forests | Introducción a Machine Learning*. Accedido: 11 de abril de 2026. [En línea]. Disponible en: https://fhernanb.github.io/libro_mod_pred/rand-forests.html
- [56] «5 ejemplos de Machine Learning y cómo te ayudan en las ventas.» Accedido: 1 de mayo de 2025. [En línea]. Disponible en: <https://www.outliersconsulting.com/general/5-ejemplos-de-machine-learning-y-como-ayuda-a-tus-ventas/>
- [57] «Modelado matemático del algoritmo KNN (K-Nearest Neighbors) - Panama Hitek». Accedido: 26 de abril de 2025. [En línea]. Disponible en: <https://panamahitek.com/modelado-matematico-del-algoritmo-knn-k-nearest-neighbors/>
- [58] «9.2. Long Short-Term Memory (LSTM) — Dive into Deep Learning 0.17.6 documentation». Accedido: 27 de abril de 2025. [En línea]. Disponible en: https://s3-us-west-2.amazonaws.com/classic.d2l.ai/chapter_recurrent-modern/lstm.html
- [59] «6.3. Preprocessing data», scikit-learn. Accedido: 27 de abril de 2025. [En línea]. Disponible en: <https://scikit-learn/stable/modules/preprocessing.html>
- [60] Universidad de las Américas y C. A. Ramos, «Los paradigmas de la investigación científica», *avpsicol*, vol. 23, n.º 1, pp. 9-17, ene. 2015, doi: 10.33539/avpsicol.2015.v23n1.167.
- [61] «15604 Search Results for “satélites Landsat y MODIS”». Accedido: 11 de abril de 2026. [En línea]. Disponible en: <https://www.nasa.gov/?search=sat%C3%A9lites+Landsat+y+MODIS>

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

ANEXOS

Anexo A.

Tabla XI. Ejemplo de consultas sql para obtención de segmentos temporales de landsat

Dataset	Rango de fecha	Consulta SQL
Landsat	D1=2010-03-01 D2=2010-03-31	<pre>CREATE table dataset_2010_3 AS SELECT R.datecapture, R.latitude, R.longitude, band1, band2, band3, band4, band5, band6, band7, Irradiation_value FROM (SELECT datecapture, latitude, longitude, band1, band2, band3, band4, band5, band6, band7 FROM public.reflectance where datecapture between '2012-01-01' and '2010-03-31') R, (SELECT datecapture, latitude, longitude, irradiation_value FROM public.irradiaceGrid450 where datecapture between '2012-01-01' and '2010-03-31') I WHERE R.datecapture=I.datecapture and R.latitude=I.latitude and R.longitude=I.longitude;</pre>

Anexo B.

Tabla XII. Modelos utilizados para la generación de las bandas espectrales

Generación de (.pkl)			
Band as	Modelo	Link Script	Tamaño
b_1	SVR (SupportVector Regression)	https://drive/MyDrive/investigacion/latlondateb1.pkl	5.6MB
b_2	KNeighborsRegressor (KNN)	/content/drive/MyDrive/investigacion/knnb1b2.pkl	105.4MB

PREDICCIÓN DE IRRADIANCIA MEDIANTE IMÁGENES SATELITALES.

b_3	KNeighborsRegressor.	/content/drive/MyDrive/investigacion/knnb1b2b3.pkl	157.1MB
b_4		/content/drive/MyDrive/investigacion/knnb1b2b3b4.pkl	208.8MB
b_5		/content/drive/MyDrive/investigacion/knnb1b2b3b4b5.pkl	260.5MB
b_7		/content/drive/MyDrive/investigacion/knnb1b2b3b4b5b6b7.pkl	312.2MB
b_6	SVR	/content/drive/MyDrive/investigacion/knnb1b2b3b4b5b6.pkl	559KB
https://drive.google.com/drive/u/1/folders/18SbiNuep8mnoeMTmqPTCdsOHKw65W5eR			

 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
---	---	---

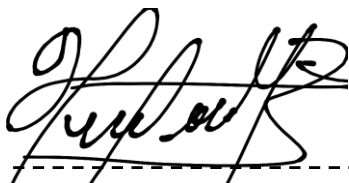
San Juan de Pasto, 6 de agosto 2026

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 comparativo de algoritmos de forecasting en la predicción de irradiancia mediante imágenes satelitales multiespectrales, presentado por las autoras Laura Sofia Gómez Rosero y Ingirdt Carolina Ordoñez López, del Programa Académico de Ingeniería al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor, 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 la paz y salvo respectivo.

Atentamente,



Héctor Andrés Mora Paz
1085251119
Ingeniería de sistemas
317 2537641
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 DE LAS AUTORAS	
Nombres y apellidos del autor: Laura Sofia Gómez Rosero	Documento de identidad: 1004215747
Correo electrónico: gomezlsg52@gmail.com	Número de contacto: 317 3824447
Nombres y apellidos del autor: Ingridt Carolina Ordoñez Lopez	Documento de identidad: 1004190930
Correo electrónico: Carolinaordonez254@gmail.com	Número de contacto: 3107307058
Nombres y apellidos del asesor: Héctor Andrés Mora Paz	Documento de identidad: 1085251119
Correo electrónico: hamora@unicesmag.edu.co	Número de contacto: 317 2537641
Título del trabajo de grado: Estudio comparativo de algoritmos de forecasting en la predicción de irradiancia mediante imágenes satelitales multiespectrales	
Facultad y Programa Académico: Ingeniería- 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

 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

presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.

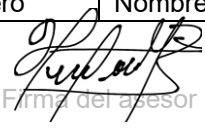
- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la 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:

- Permito(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 6 días del mes de agosto del año 2026

 Firma del autor	 Firma del autor
Nombre del autor: Laura Sofia Gómez Rosero	Nombre del autor: Ingridt Carolina Ordoñez Lopez
 Firma del asesor	
Nombre del asesor: Héctor Andrés Mora Paz	