



Desarrollo de un modelo de aprendizaje ensamblado para la detección de quema de caña de azúcar en la costa norte y centro del Perú durante el período 2017 - 2022.

Autor: Jair Francisco FLORES FARFAN

Asesor: Francisco Alejandro ALCÁNTARA BOZA

Tabla de contenidos

01. Introducción

02. Marco Teórico

03. Hipótesis y Variables

04. Área de estudio

05. Metodología

06. Resultados

07. Conclusiones

08. Discusión y Recomendaciones



01.

Introducción

"You can't fix you can't see"

- William Marshall.
CEO Planet Labs



Planteamiento del problema

01.

En la **costa norte y centro del Perú**, la tradicional práctica de quemar los residuos de la caña de azúcar (75 - 80% del total de la caña) ha generado una creciente preocupación debido a sus impactos negativos en el suelo, el medioambiente y la salud de la población circundante.



Momento exacto de la quema de un campo de caña de azúcar (izquierda) y las partes de la caña de azúcar (derecha)

RIC VASQUEZ / AP



A nivel mundial

01.

1. **Arbex et al. (2007)** relacionó directamente el aumento de las emisiones de material particulado durante los períodos de quema de caña de azúcar con un incremento del 11.6 % en las admisiones hospitalarias por asma en un hospital de la ciudad de Araraquara, Brasil.
2. **De Andrade et al. (2010)** indica un aumento de hidrocarburos aromáticos policíclicos (PAH) cuatro veces mayor en los períodos de quema en comparación a los períodos sin quema.
3. **Salas (2010, p. 42)** menciona que la quema de caña de azúcar realizada antes de la cosecha afecta a la biodiversidad del suelo, provocando la pérdida de nitrógeno, la población de microorganismos del suelo y el material orgánico disponible en el mismo.
4. **França et al. (2012)** menciona que la quema de los residuos de caña de azúcar genera un aumento significativo de contaminantes hacia la atmósfera como óxidos de carbono (CO_2 y CO), óxidos de nitrógeno (NO_x), hidrocarburos (HC), material particulado fino ($PM_{2.5}$) y grueso (PM_{10}).
5. **Silveira et al. (2013)** observaron que los hidrocarburos aromáticos policíclicos (PAH), resultantes de la quema, tienen una asociación directa con efectos genotóxicos en trabajadores de caña en la ciudad brasileña de Barretos, lo que sugiere un incremento en los casos de riesgo de cáncer en las personas afectadas por las emisiones.
6. **Pestana et al. (2017)** también asoció el aumento de los niveles de NO_2 con un incremento del 1.12 % en las admisiones hospitalarias por enfermedades cardiovasculares en la región del oeste de São Paulo.
7. **Nowell et al. (2022)** estimaron una mortalidad de 2.5 muertes por año en veinte condados del sur de Florida, en Estados Unidos, debido a la emisión de $PM_{2.5}$ producida por los incendios de caña de azúcar, las causas de las muertes estaban relacionadas con vías respiratorias, enfermedades pulmonares y cáncer de pulmón.





Proyecto de ley
N° 06243/2020-
CR, compilado de
denuncias
documentado por
el Congreso de la
República
Congreso de la
República (2020).



La Defensoría del Pueblo en La Libertad advirtió sobre la humareda constante generada por la quema de caña de azúcar en la ciudad de Trujillo, que afecta tanto a las viviendas de personas en aislamiento como a los establecimientos de salud, según múltiples quejas recibidas de la población.

2020

2019

2018

2017

2016

2015

2012

2010

Los ciudadanos de **Miguel Checa**, provincia de **Sullana**, en la región **Piura**, denunciaron ante la Municipalidad Distrital la caída de **cenizas** desde la **1 de la mañana**, afectando viviendas, calles y pistas. También reportaron un olor penetrante a humo durante las mañanas, que perjudica la salud de la población.

"Los principales problemas que atendemos acá son IRAS. Los días que queman caña se incrementan estas infecciones", comenta Laura Yarlequé Castillo, técnica en enfermería del centro de salud de Viviate

En noviembre, en la localidad de **La Huaca, Paita, Piura**, un accidente de tránsito atribuido a la humareda de la quema de caña dejó cinco muertos. La comunidad responsabilizó a la empresa **Caña Brava**.

En **Laredo, Trujillo, La Libertad** la **Municipalidad Distrital** reportó a la **OEFA** la quema de caña de azúcar realizada por la empresa agroindustrial **Vallesol SAC** en el poblado de **Barraza**. Este hecho, generó una gran cantidad de humo que invadió gran parte de la ciudad de Laredo, afectando la calidad del aire.

Población y autoridades denuncian que esta práctica realizada por **AgroAurora** afecta la salud de sus hijos en los poblados de Paita y Sullana.

La empresa **Paramonga** ha recibido denuncias de la población del asentamiento humano **Nueva Esperanza, Lima** por las emisiones de gases y partículas que generan irritación de garganta y ojos. Se estima que los afectados ascienden a **900 personas**.

Los pobladores de **Viviate, La Huaca, Paita, Piura**, denuncian las quemadas de caña realizadas por **Agrícola del Chira S.A.C** y **AgroAurora S.A.C**.

> 12 000

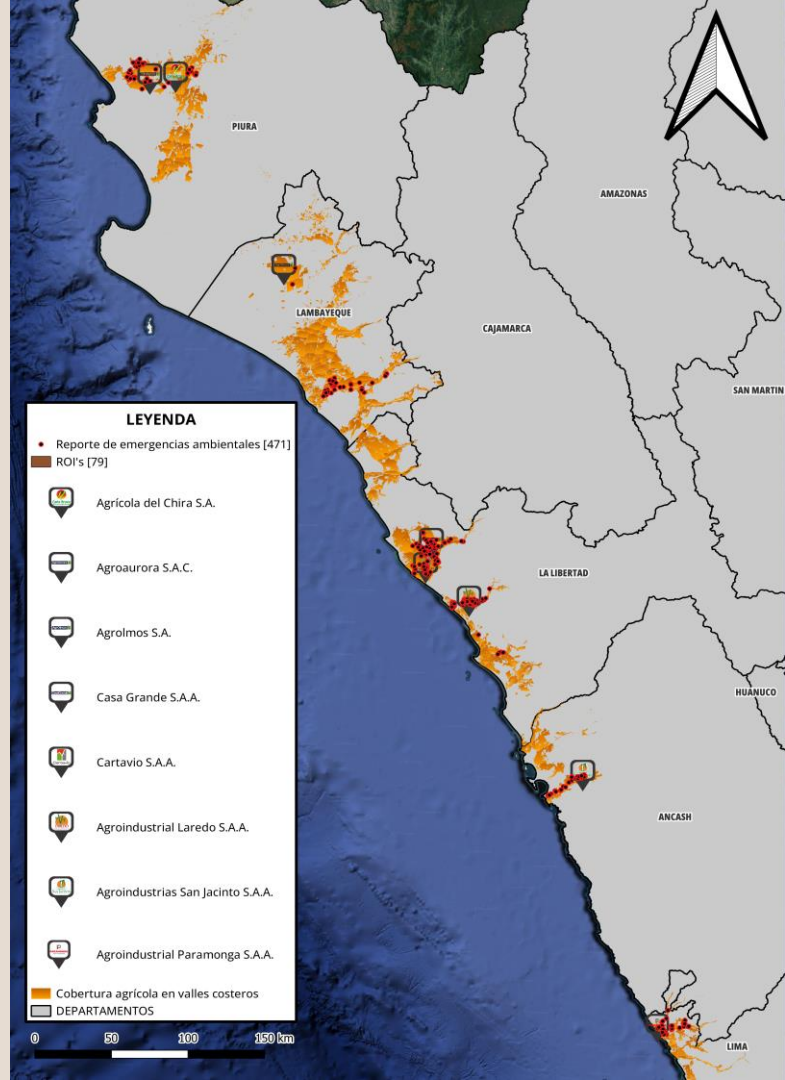
Personas afectadas por la quema de caña de
azúcar

471

Denuncias compiladas según OEFA (enero
2020- agosto 2023)

34

Distritos afectados en Perú



Legislación actual

LEY Nº 31949

LA PRESIDENTA DE LA REPÚBLICA

POR CUANTO:

EL CONGRESO DE LA REPÚBLICA;

Ha dado la Ley siguiente:

LEY QUE REGULA LA QUEMA EN PIE DE CULTIVOS DE CAÑA DE AZÚCAR Y ESTABLECE DISPOSICIONES PARA LA ADECUACIÓN DE NUEVOS MÉTODOS DE COSECHA, SALVAGUARDANDO LA SALUD PÚBLICA Y EL DESARROLLO SOSTENIBLE DEL SECTOR

Artículo 1. Objeto de la Ley

La presente ley tiene por objeto regular la quema en pie de cultivos de caña de azúcar como método de cosecha y establecer un plazo de adecuación para que los productores de caña de azúcar desarrollen e implementen procedimientos de cosecha que no deterioren la calidad del aire ni afecten la salud de la población.

Artículo 7. Vigilancia y monitoreo

7.1. El Organismo de Evaluación y Fiscalización Ambiental (OEFA) establece, difunde y facilita mecanismos de denuncia para la población que se vea afectada por la quema de caña de azúcar en las zonas donde se desarrolla esta actividad.

7.2. Sin perjuicio del trámite que corresponda a las denuncias presentadas, el Organismo de Evaluación y Fiscalización Ambiental (OEFA), en mérito a la evaluación y análisis que realice, incorpora en el Plan Anual de Fiscalización Ambiental las acciones de vigilancia y monitoreo ambiental de las zonas en las que se hayan presentado un mayor número de denuncias.

7.3. El Organismo de Evaluación y Fiscalización Ambiental (OEFA) elabora un informe anual sobre las acciones de vigilancia y monitoreo de la actividad de quema de caña de azúcar que contenga, como mínimo:

- Número de denuncias presentadas.
- Estado de las denuncias.
- Sujetos sancionados.
- Resultados de las acciones de vigilancia y monitoreo programadas.
- Estado de la calidad del aire de las zonas monitoreadas.
- Evolución de los principales indicadores.

CAÑA DE AZÚCAR: RUTA HACIA LA COSECHA EN VERDE

Determinación de distancias mínimas para la quema de caña de azúcar en las localidades de La Huaca y La Rinconada (provincia de Paita, Piura) mediante un modelamiento de dispersión de material particulado en el año 2018

Acciones regulatorias del OEFA para incluir esta metodología de distancias mínimas como parte de los IGA. Fuente: OEFA (2023).

01.



COD_ADM	DES_ADM	DEPARTAMENTO	PROVINCIA	DISTRITO	IGA	RESOLUCIÓN
ADM11681	AGROINDUSTRIAS SAN JACINTO S A A	ANCASH	SANTA	NEPEÑA	Mod. P.A.M.A	R.D.G. N° 147-14-MINAGRI-DGAAA
ADM11828	CASA GRANDE SOCIEDAD ANONIMA ABIERTA	LA LIBERTAD	ASCOPE	CASA GRANDE	P.A.M.A	R.D. N° 00490-2022-PRODUCE/DGAAMI
ADM11829	CARTAVIO SOCIEDAD ANONIMA ABIERTA	LA LIBERTAD	ASCOPE	SANTIAGO DE CAO	Act. P.A.M.A	R.D. N° 513-2022-PRODUCE/DGAAMI
ADM17036	AGROLMOS SOCIEDAD ANONIMA - AGROLMOS S.A.	LAMBAYEQUE	LAMBAYEQUE	OLMOS	Mod E.I.Asd	R.D. N° 00610-2021-PRODUCE/DGAAMI
ADM11867	AGRO INDUSTRIAL PARAMONGA S.A.A.	LIMA	BARRANCA	PARAMONGA	P.A.M.A	R.D. N° 534-2019-PRODUCE/DVMYPE-I/DGAAMI
ADM17481	AGROAURORA SOCIEDAD ANONIMA CERRADA - AGROAURORA	PIURA	PAITA	LA HUACA	E.I.A.	R.D. N° 594-2021-PRODUCE/DGAAMI
ADM19756	COOPERATIVA AGRARIA MOCHICA LIMITADA	LA LIBERTAD	TRUJILLO	TRUJILLO	D.A.A.C.	R.D.G. N° 0480-2021-MIDAGRI-DVDAFIR-DGAAA

Sólo 7 de los 21 administrados por el OEFA tienen un IGA donde se compromete con la cosecha mecanizada en verde.

Objetivos de la investigación

01.

Objetivo General

Desarrollar un **modelo de aprendizaje ensamblado** para la detección de quema de caña de azúcar.

02.

Objetivos Específicos

1. **Generar el conjunto de datos** necesario para el funcionamiento del modelo de aprendizaje ensamblado para detectar la quema de caña de azúcar.
2. **Determinar los algoritmos que mejor se ensamblan al modelo** para detectar la quema de caña de azúcar.
3. **Determinar mediante métricas de evaluación si el modelo de aprendizaje ensamblado presenta un desempeño superior** en comparación con los algoritmos individuales.



Importancia y alcance de la investigación

01.

01. Importancia

El modelo desarrollado en este estudio facilitará la vigilancia y monitoreo de los predios agrícolas, fortaleciendo la fiscalización de las empresas por parte del OEFA mediante la identificación y validación de estas áreas.

02. Alcance

- El modelo en este estudio se aplica específicamente a las regiones de la costa norte y centro del Perú, utilizando una base de datos y los pesos preentrenados para estas áreas, disponibles públicamente en Hugging Face.
- El modelo tiene un gran potencial de escalabilidad, es por ello que el código fuente está disponible en GitHub para su replicación y futura aplicación en otros contextos geográficos y tipos de cultivos, lo que permite su uso más allá de las zonas estudiadas.



Limitaciones de la investigación

01.

Limitaciones Operacionales

- Limitaciones Metodológicas
- Disponibilidad y calidad de los datos
- Generalización del modelo
- Simplificaciones y supuestos en la modelización

02.

Limitaciones Operacionales

- Tiempo y recursos humanos para el etiquetado
- Replicabilidad y escalabilidad del modelo

03.

Limitaciones en la Validación

- Falta de datos en campo detallados





02.

Marco Teórico

*"In order to seek truth, it is necessary
once during our life to doubt, as far as
possible, of all things."*

- Rene Descartes
Philosopher and mathematician



Antecedentes

02.



<u>Autores</u>	<u>Contribución</u>	<u>Metodología</u>	<u>Área de estudio</u>	<u>Resultados</u>
Lee et al. (2022)	se compararon tres técnicas representativas de este enfoque para la detección de áreas quemadas: Random Forest, LightGBM y U-Net.	Random Forest, LightGBM, U-Net	Imágenes Sentinel-2 post-incendio, dos eventos de incendio (2019-2020) en Corea del Sur	U-Net superó en todas las métricas: Precisión (0.97), Recall (0.87), F1-score (0.88), Kappa (0.88).
Farhadi et al. (2023)	Presentó un índice espectral novedoso denominado Índice de Detección de Área Quemada (BADI) que aprovecha al máximo el espectro disponible en Sentinel-2	Índice BADI	Imágenes Sentinel-2 post-incendio, eventos de incendio (mayo 2020) en Fars, Irán.	BADI alcanzó métricas Kappa (0.87) y precisión (0.92), mayor que BAIS2 y NBR.
Al-Dabbagh e Ilyas (2023)	Se demuestra la elección de la arquitectura U-Net como el más óptimo en la segmentación semántica de áreas quemadas	Se entrenaron catorce modelos de aprendizaje profundo basados en combinaciones de cinco arquitecturas (U-Net, U-Net++, Attention ResU-Net, LinkNet y DeepLabV3+	Imágenes Sentinel-2 post-incendio durante el año 2021 en cinco provincias de Turquía.	U-Net con ResNet50 logró mejor IoU y F1-score: (0.97, 0.98), seguido de U-Net++ y LinkNet.
Hu et al. (2024)	Desarrollaron un modelo de aprendizaje ensamblado para el mapeo de la progresión de áreas quemadas en tiempo casi real.	Aprendizaje ensamblado (HRNet + SVM)	Imágenes Sentinel-2 en el periodo de julio 2016, en el evento "Sand Fire" , Santa Clarita (EE.UU.)	HRNet-SVM mejoró la precisión de clasificación: índice Kappa (70.75 %) comparado con el inicial (66.77 %).

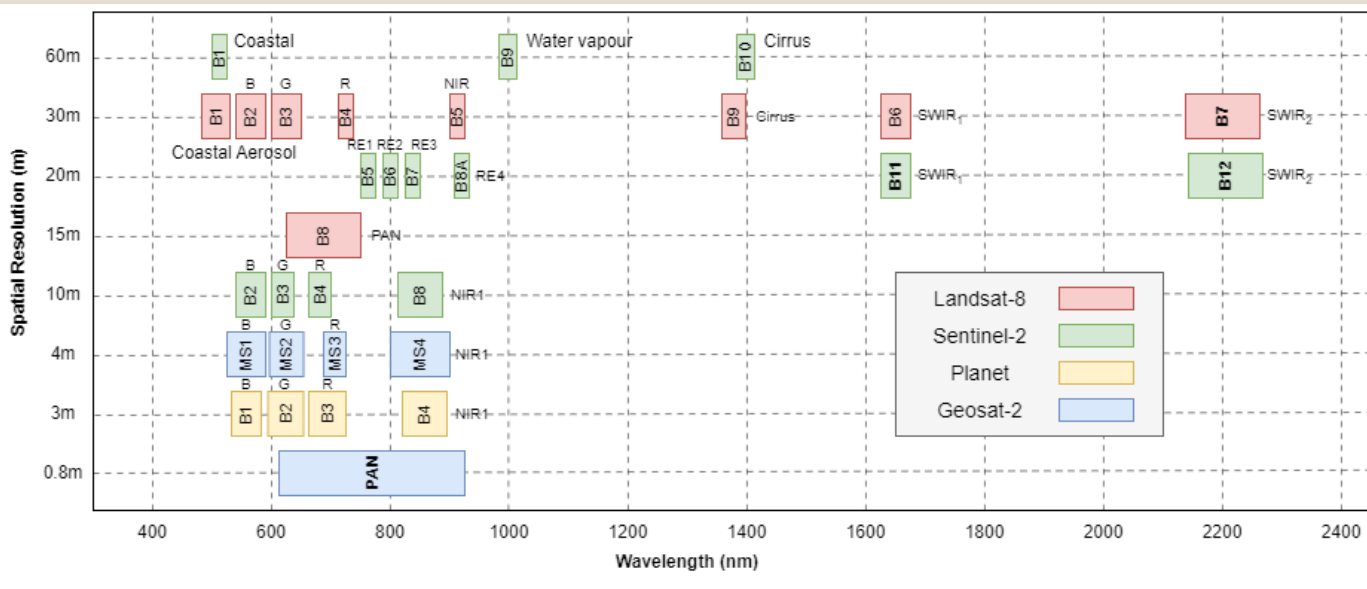


Bases Teóricas

02.

Imágenes satelitales Sentinel-2

El programa Sentinel-2 proporciona imágenes globales en doce bandas espectrales con una resolución de 10 a 60 m y un tiempo de revisita de aproximadamente cinco días con dos satélites (S2A y S2B) en condiciones libres de nubes.



Comparación de las bandas espectrales de diferentes sensores satelitales (Landsat-8, Sentinel-2, Planet, y Geosats-2)

USGS. Comparison of Sentinel-2 and Landsat.

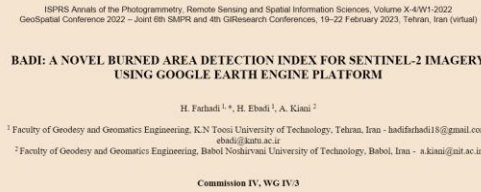


Detección de áreas quemadas

Según Knopp et al. (2020), los algoritmos empleados para la delimitación de áreas quemadas siguen dos enfoques: **aquellos basados en reglas** y los que se apoyan en técnicas de **aprendizaje automático**.

También se clasifican basados en el uso exclusivo de una escena posterior al incendio (**enfoque mono-temporal**) o la integración adicional de datos previos al incendio o series temporales (**enfoque multi-temporal**).

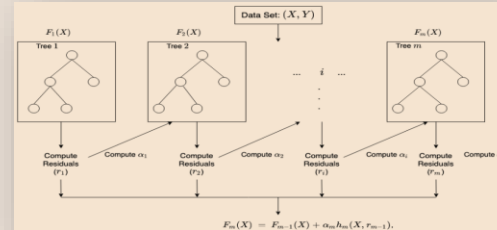
Algoritmos basados en reglas



Farhadi (2023)

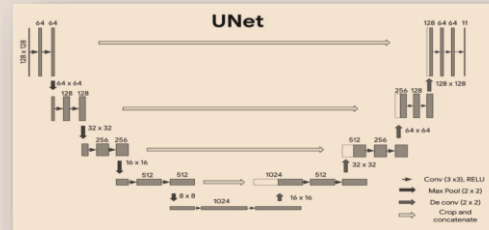


Árboles de decisión potenciado por gradiente (GBDT)



Lee et al. (2022)

Redes Neuronales Convolucionales (CNN)



Al-Dabbagh e
Ilyas (2023)



Algoritmos basados en reglas

02.

Normalized Burned Ratio

El NBR es un índice utilizado para evaluar el impacto de los incendios forestales y cambios en la vegetación, siendo utilizado como un clasificador de áreas quemada.

La vegetación saludable y no afectada por incendios tiene un valor positivo de NBR, mientras que las áreas quemadas se expresan en valores negativos

$$NBR = \frac{NIR - SWIR2}{NIR + SWIR2}$$

$$\Delta NBR = NBR_{pre-fire} - NBR_{post-fire}$$



☐ Monotemporal
☐ Bitemporal

Burned Area Detection Index

Espectralmente hace uso de las bandas: infrarrojo de onda corta (SWIR), infrarrojo cercano (NIR), borde del rojo (Red Edge) y rojo (Red) aprovechando información indirecta como la actividad fotosintética y la concentración de humedad que existe en la cobertura vegetal. Los valores positivos suelen indicar presencia de áreas quemadas.

$$F1 = \frac{(SWIR2 + SWIR1) - (NIR + \text{Red Edge } 4)}{(SWIR2 + SWIR1) + (NIR + \text{Red Edge } 4)}$$
$$F2 = \left(2 - \sqrt{\frac{\text{Red Edge } 2 \times \text{Red Edge } 3 \times (NIR + \text{Red Edge } 4)}{\text{Red} + \text{Red Edge } 1}} \right)$$

$$BADI = F1 \times F2$$

$$\Delta BADI = BADI_{pre-fire} - BADI_{post-fire}$$



Algoritmos de aprendizaje automático

02.

Regresión Logística

Se puede definir la probabilidad de que ocurra un evento dado un conjunto de variables predictoras $X_1, X_2, \dots, X_n$ utilizando la función sigmoide.

Sirve como **metaalgoritmo de stacking** para el aprendizaje ensamblado.

Funciones de pérdida

- Weighted Binary Cross Entropy

$$L(y, p) = -\frac{1}{N} \sum_{i=1}^N w_i [y_i \log(p) + (1 - y_i) \log(1 - p)]$$

Donde w_i es el peso asignado a la muestra i , y es la etiqueta verdadera de clase binaria y p es la probabilidad predicha para la clase positiva.

- Dice Loss $DSC = 1 - \frac{2 \times |X \cap Y|}{|X| + |Y|}$

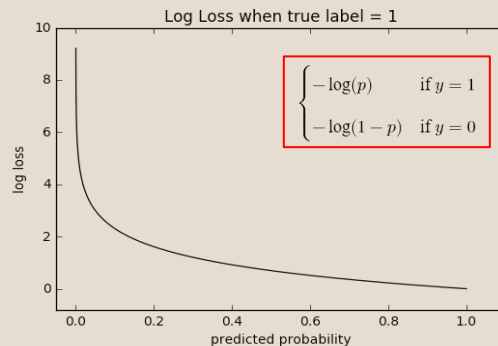
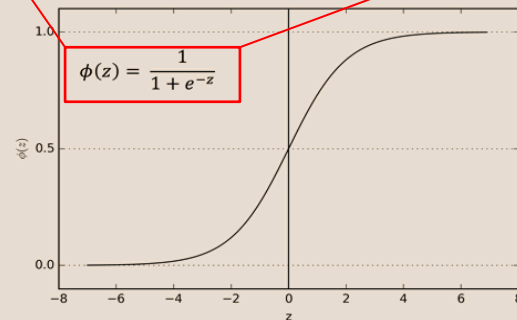
X es la máscara de segmentación predicha, Y es la máscara de segmentación verdadera

- Combined Loss $L_{WCE-Dice} = (1 - \alpha) \times L_{WCE} + \alpha \times L_{Dice}$

Donde α es el coeficiente de ponderación (entre 0 y 1) que controla la importancia relativa de cada función de pérdida.

$$\text{logit}(y) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

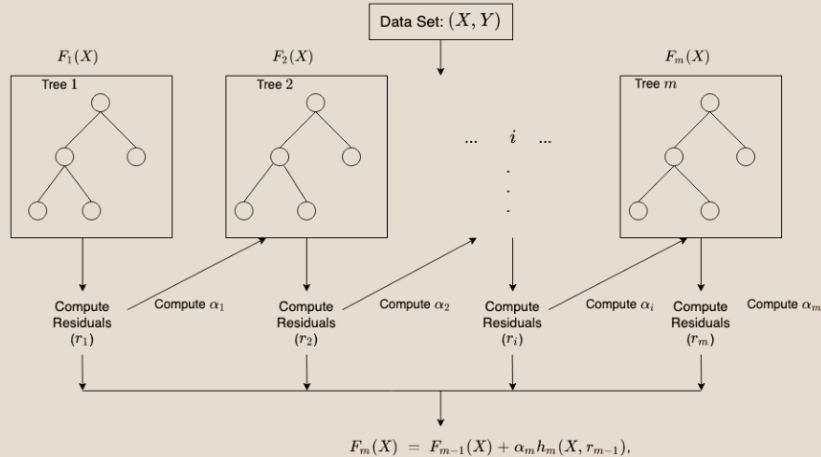
$$P(Y = 1 | X_1, X_2, \dots, X_n) = \frac{1}{1 + e^{-\text{logit}(y)}}$$



Gradient Boosting Decision Tree

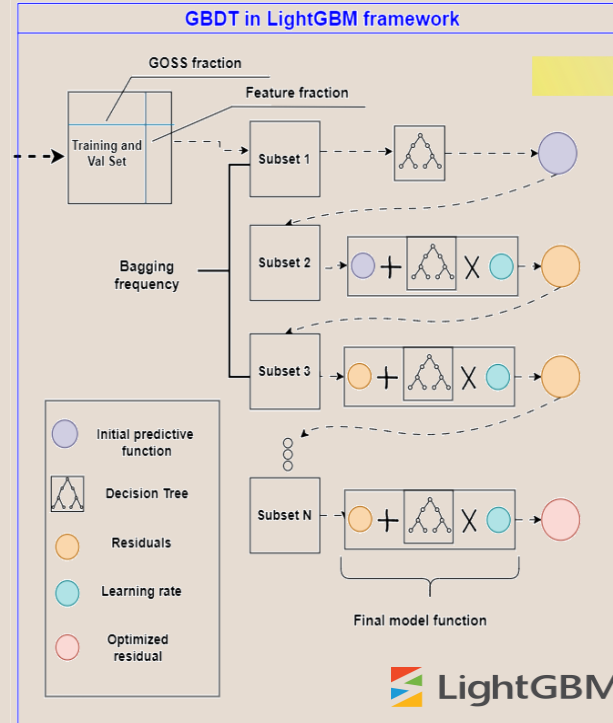
El GBDT, o árbol de decisión potenciado por gradiente, es un algoritmo de aprendizaje automático que utiliza boosting como técnica de ensamblado (Zhang et al., 2022). Este modelo entrena árboles de decisión en secuencia, ajustando en cada iteración los gradientes negativos, también conocidos como **errores residuales**.

Ke et al. (2017) desarrollaron el algoritmo LightGBM incorporando dos métodos innovadores: **Gradient-based One-Sided Sampling (GOSS)** y **Exclusive Feature Bundling (EFB)**, que reducen el número de instancias y de características respectivamente, mejorando la eficiencia y velocidad del algoritmo al manejar datos de alta dimensionalidad y gran volumen



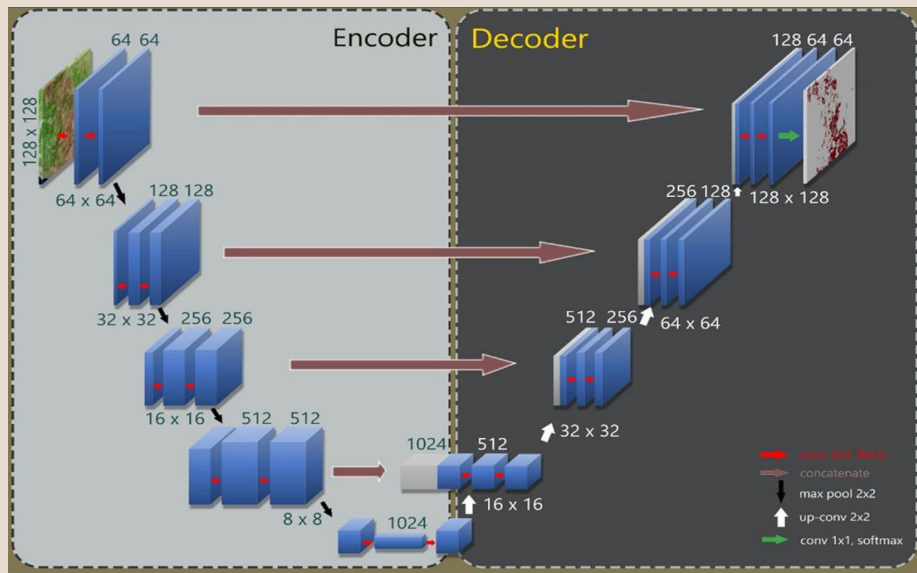
Fuente: Amazon Web Services (2024).

02.



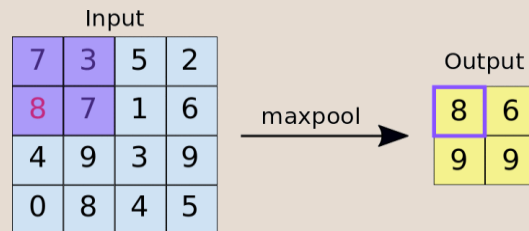
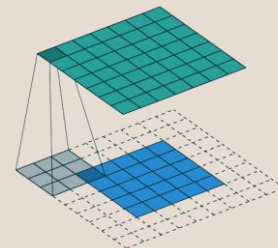
Segmentación Semántica (U-Net)

Balas et al. (2020) lo definen como un tipo de red neuronal artificial (ANN) con una arquitectura de retroalimentación profunda que le permite aprender diversas características de los objetos, especialmente características espaciales. Las capas iniciales de un modelo CNN aprenden y extraen las características de alto nivel con una abstracción más baja y las capas más profundas aprenden y extraen las características de bajo nivel con mayor abstracción.



Fuente: Al-Dabbagh e Ilyas (2023).

conv 3x3



max pool 2x2

up-conv 2x2

$$\begin{array}{|c|c|} \hline 0 & 1 \\ \hline 2 & 3 \\ \hline \end{array} + \begin{array}{|c|c|} \hline 4 & 1 \\ \hline 2 & 3 \\ \hline \end{array} + \begin{array}{|c|c|} \hline 8 & 2 \\ \hline 4 & 6 \\ \hline \end{array} + \begin{array}{|c|c|} \hline 12 & 3 \\ \hline 6 & 9 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 0 & 4 & 1 \\ \hline 8 & 16 & 6 \\ \hline 4 & 12 & 9 \\ \hline \end{array}$$

Input: $\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$, Transposed Conv (Stride 1): $\begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$, Output: $\begin{bmatrix} 0 & 4 & 1 \\ 8 & 16 & 6 \\ 4 & 12 & 9 \end{bmatrix}$

02.



Métricas de evaluación

<u>Métrica</u>	<u>Importancia</u>	<u>Interpretación</u>	<u>Fórmula</u>
Precisión	Es útil para entender la exactitud del modelo al identificar áreas quemadas	Una alta precisión indica que el modelo tiene <u>pocos falsos positivos</u> , es decir, no clasifica incorrectamente áreas no quemadas como quemadas.	$Precision = \frac{TP}{TP + FP}$
Recall	Es útil porque garantiza que todas las áreas quemadas sean identificadas.	Es crucial para asegurar que <u>no se pasen por alto zonas afectadas</u> , lo cual es fundamental en la planificación de la respuesta a emergencias.	$Recall = \frac{TP}{TP + FN}$
F1-Score	Es útil cuando es importante mantener un equilibrio entre evitar falsos positivos (precisión) y no perder áreas quemadas reales (recall).	Es particularmente relevante en situaciones <u>donde tanto las omisiones como las falsas alarmas tienen consecuencias significativas y deben minimizarse</u> .	$F_1 = \frac{2 \times TP}{2 \times TP + FN + FP}$
Intersection Over Union	Es útil para medir la superposición entre la predicción del modelo y la verdad del terreno.	Permite una evaluación precisa de cuán bien el modelo ha capturado los límites de la quema.	$IoU = \frac{TP}{TP + FP + FN}$
Índice Kappa	Permite evaluar la concordancia del modelo más allá del azar.	Es útil como una medida <u>más robusta en comparación con la simple precisión</u> , ya que considera tanto los aciertos como los errores del modelo.	$P_o = \frac{TP + TN}{TP + FP + FN + TN}$ $P_e = \frac{(TP + FP) \cdot (TP + FN) + (FN + TN) \cdot (FP + TN)}{(TP + FP + FN + TN)^2}$ $\kappa = \frac{P_o - P_e}{1 - P_e}$

02.



- ☐ **TP:** Verdaderos positivos
- ☐ **TN:** Verdaderos negativos
- ☐ **FP:** Falsos positivos
- ☐ **FN:** Falsos negativos



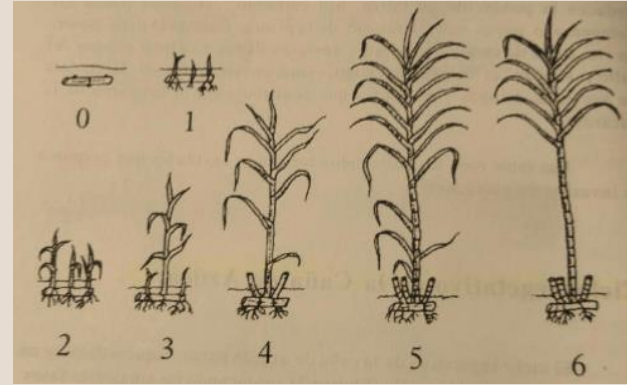
Quema de caña de azúcar

Fenología de la caña de azúcar

Según Helfgott (2016), la caña de azúcar, con nombre científico *Saccharum officinarum*, es una planta perenne de la familia de las gramíneas, originaria de Asia tropical y subtropical que se adapta bien a una variedad de suelos, siempre y cuando haya disponibilidad de agua, aunque prefiere las temperaturas tropicales y suelos de textura franca, con un pH cercano a 7.0, bien drenados y profundos.

El ciclo vegetativo de la caña de azúcar se desarrolla en las siguientes fases:

- **20 - 40 días después de la siembra:** Prebrotamiento (0) , Brotamiento (1), Enraizamiento (2), Macollaje (3)
- **18 - 24 meses después de la siembra:** Etapa inicial de crecimiento (4), Crecimiento (5), Madurez (6)



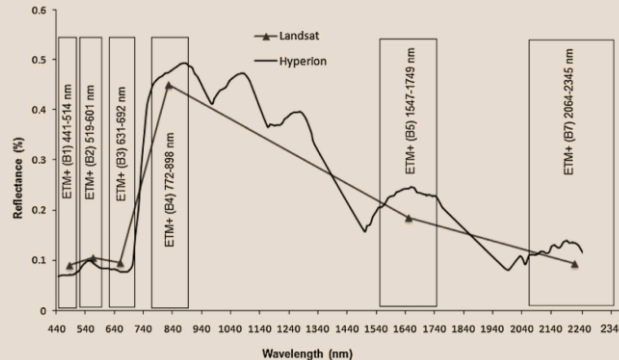
El ciclo vegetativo de la caña de azúcar

Fuente: Helfgott (2016).

Firma espectral de caña de azúcar

Los perfiles espectrales son típicos de la vegetación saludable ya que mostraron la **mayor reflectancia en el infrarrojo cercano (NIR)**, 772 a 898 nm y **valores más bajos en el infrarrojo de onda corta (SWIR)**, a 2064 a 2345 nm y **en el visible ≤ 740 nm.**

Este perfil espectral se obtuvo durante la etapa de madurez de la caña de azúcar en uno de los siete terrenos agrícolas e industriales de caña de azúcar ubicados en Juzestán, una provincia en el suroeste de Irán.



Fuente: Alcaras et al. (2022).

02.



Quema de caña

Cosecha de caña en Perú

La cosecha de caña se realiza todo el año, salvo durante la *Parada Anual* (abril/mayo), cuando se suspende el riego (*agoste*) para secar hojas verdes y acumular sacarosa (Helfgott, 2016).

La etapa de cosecha y procesamiento, ambas se llevan a cabo dentro de un lapso no mayor a 48 horas

Quema de caña

La quema, crucial antes de la cosecha, facilita el corte eliminando hojas verdes y secas. Considera accesibilidad, proximidad a cultivos, viviendas y condiciones del campo (Casa Grande S.A.A., 2018).

Técnica de quema

Se inicia a contraviento para evitar propagaciones a centros poblados y, si es necesario, se realiza una requema con lanzallamas o paja impregnada con kerosene (Agroindustrias San Jacinto S.A.A., 2014).

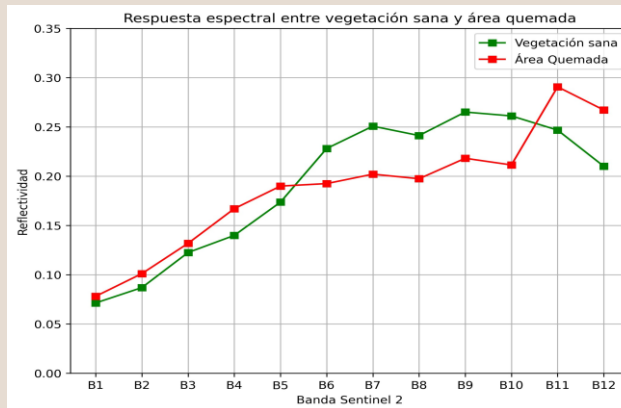
Firma espectral de áreas quemadas

Dado que la vegetación verde es muy reflectante en el NIR, la quema suele causar disminuciones más o menos significativas de la reflectancia ya que tanto la vegetación verde y seca antes del incendio tienen reflectancias mayores que las quemadas recientes, por lo que una caída en la reflectancia en el NIR es sistemático y sirve como un indicador confiable de áreas quemadas recientemente.



Quema de caña de azúcar en campos de San Jacinto

Fuente: Agroindustrias San Jacinto S.A.A. (2014)



Fuente: Elaboración propia utilizando la propia base de datos

02.





03.

Hipótesis y Variables

"A problem well stated is a problem half-solved."

- Charles Franklin Kettering.
American inventor, engineer, and
businessman



Hipótesis

01. Hipótesis General

El modelo de aprendizaje ensamblado para la detección de quema de caña de azúcar permitirá resultados con un gran porcentaje de precisión.

02. Hipótesis Específicas

- La correcta generación del conjunto de datos necesario para el funcionamiento del modelo de aprendizaje ensamblado para detectar la quema de caña de azúcar optimizará la precisión del modelo.
- Los algoritmos de GBDT y U-Net serán los que mejor se ensamblan al modelo para detectar la quema de caña de azúcar.
- El modelo de aprendizaje ensamblado presentará un desempeño superior en comparación con los algoritmos individuales.



Variables

03.



Variables	Dimensión	Indicador	Unidad de medida	Instrumento
Variable Independiente				
Desarrollo de un modelo de aprendizaje ensamblado	Modelo U-Net	Métricas de evaluación	Porcentaje	Librería Torch y PytorchLightning en Python
	Gradient Boosting Decision Tree			Librería LightGBM en Python
	Regresión Logística			Librería scikit-learn en Python
Variable Dependiente				
Detección de quema de caña de azúcar	Ubicación de eventos de quema	Coordenadas geográficas	Grados decimales	Plataforma SIG (QGIS)
	Extensión de las áreas quemadas	Área quemada	Hectáreas	Plataforma de análisis de imágenes satelitales (Python)





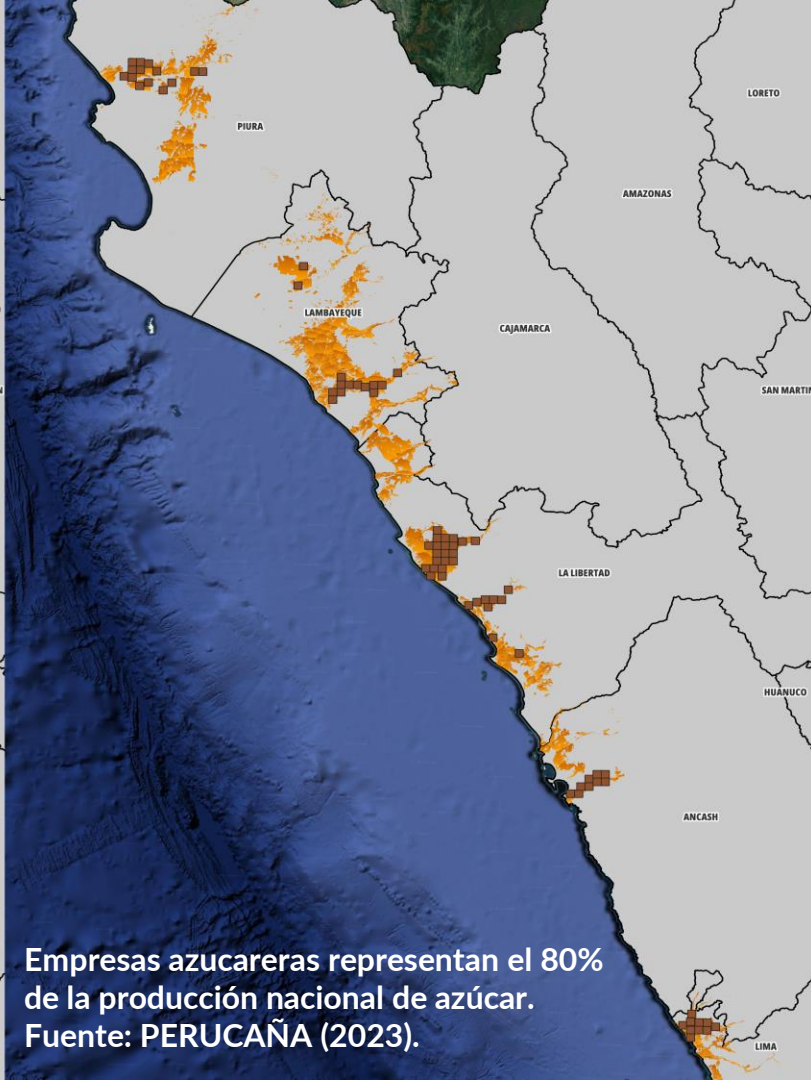
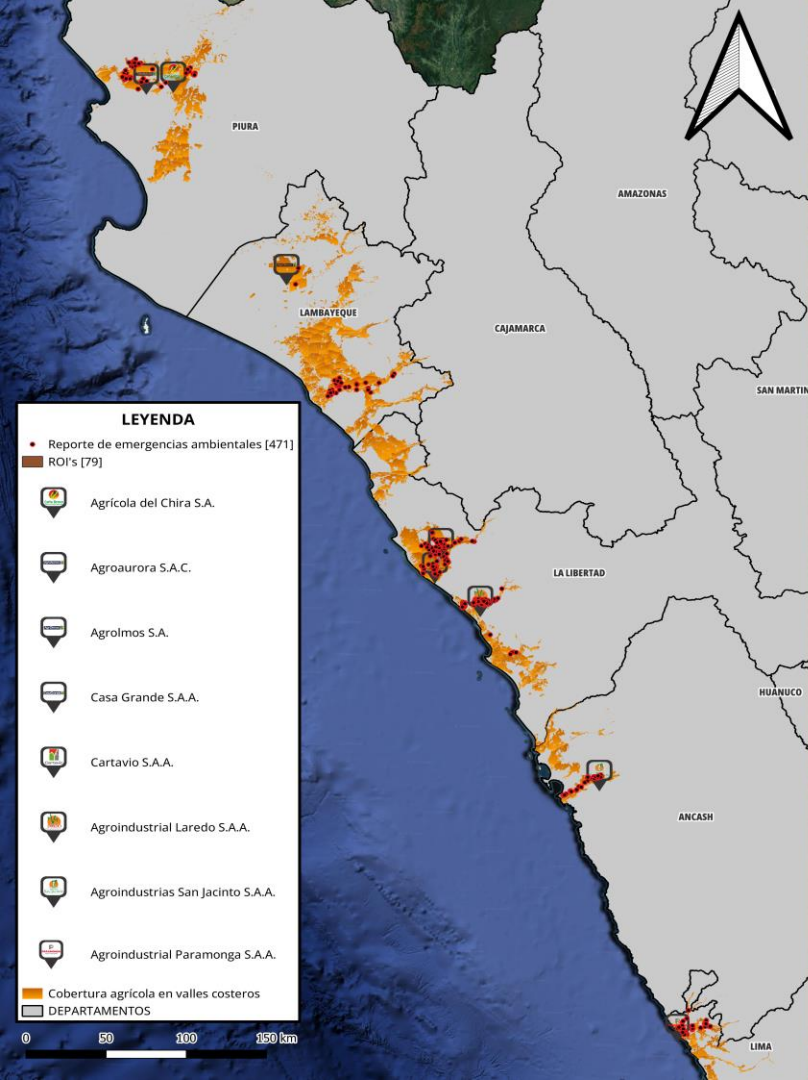
04.

Área de estudio

"The great aim of education is not
knowledge but action."

- Herbert Spencer.
British philosopher





Empresas azucareras representan el 80% de la producción nacional de azúcar.
Fuente: PERUCAÑA (2023).



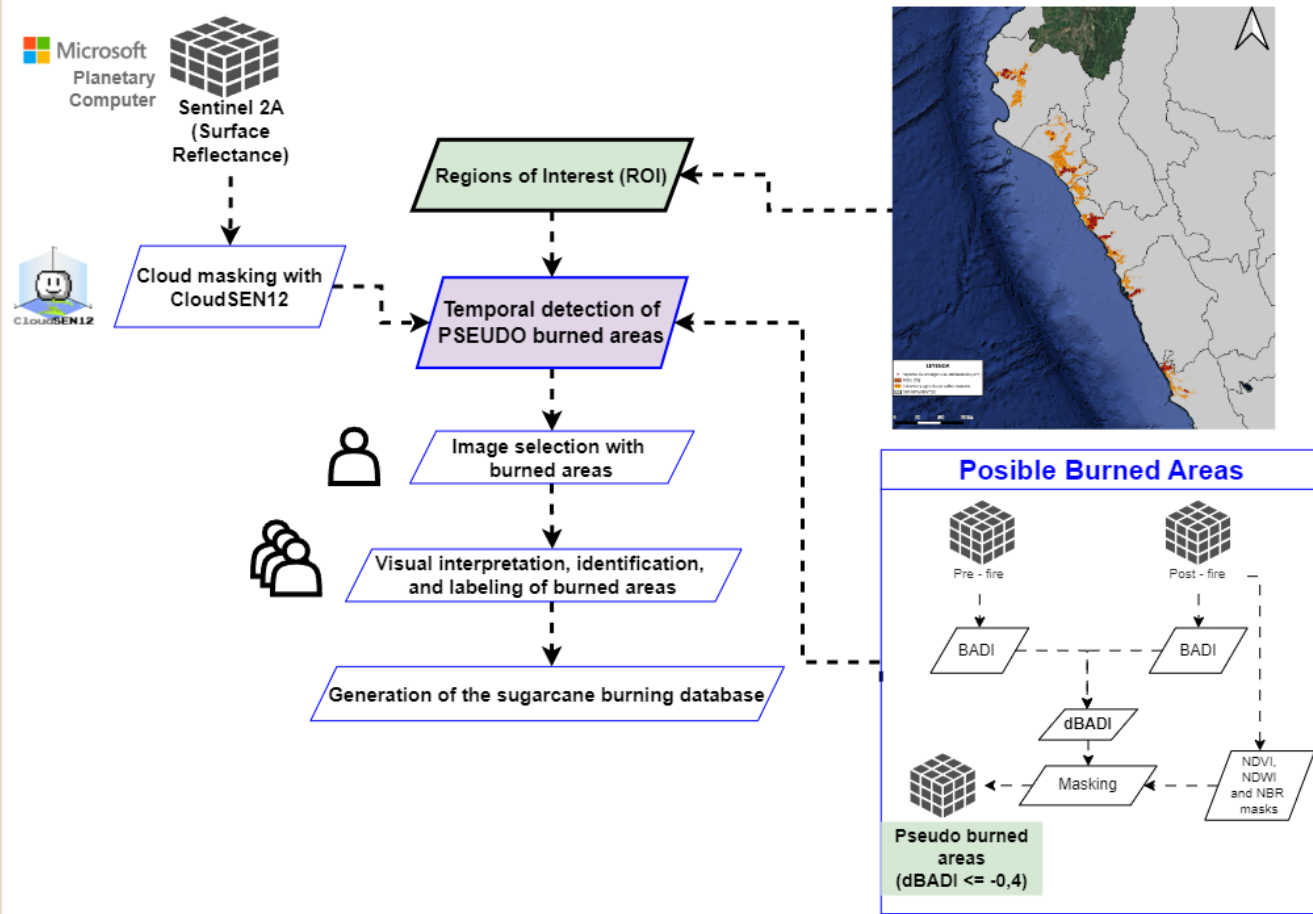
05.

Metodología

*"What I cannot create, I do not
understand"*

- Richard P. Feynman
Physicist

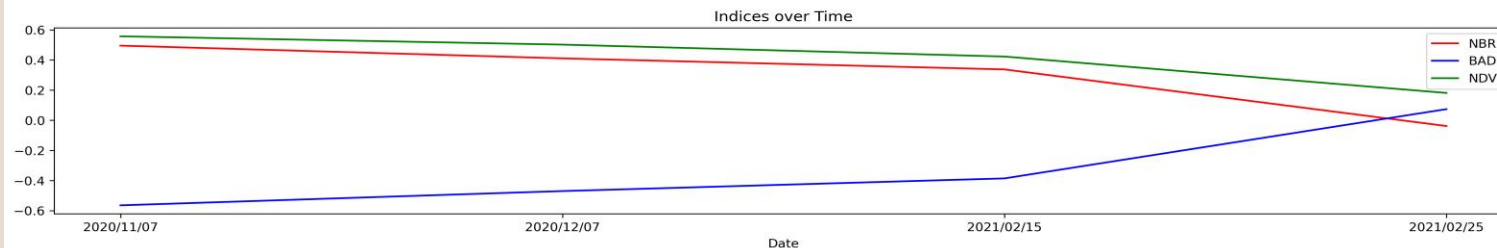
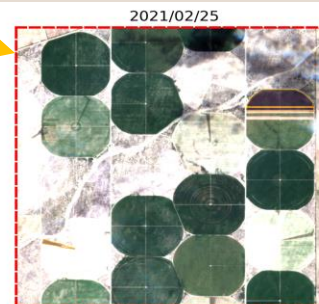
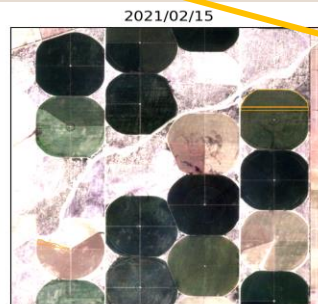
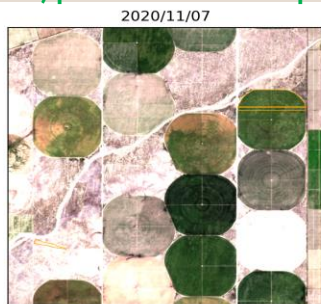
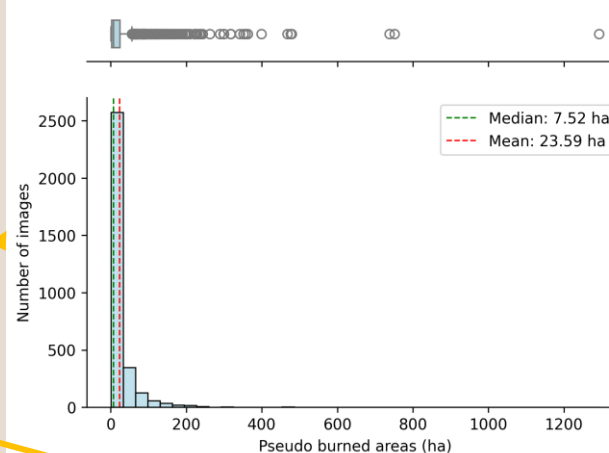




1. Cada ROI contenía aprox 400. escena, post filtro de nubes CloudSEN12: **10,149** imágenes.
2. Luego del filtro de áreas pseudo quemadas: **3,195** imágenes
3. Revisión manual de **2,000** imágenes luego del filtro del muestreo estratificado de acuerdo al histograma

Histograma de áreas pseudo quemadas, donde se realizará un muestreo estratificado utilizando los percentiles 25, 50 y 75.

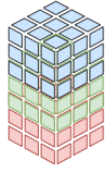
Aplicación del filtro de áreas pseudo quemadas (posible burned areas) para una serie temporal del ROI_0005_04



05.



17 BANDS

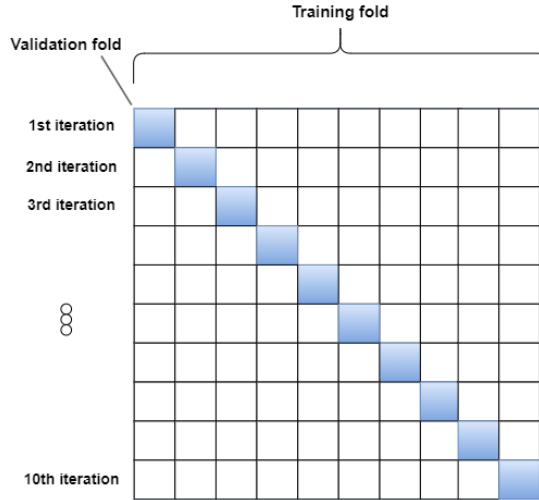


- 10 bands of Sentinel-2
- Normalized Burnt Ratio (NBR)
- Burned Area Detection Index (BADI)
- Slope (%)
- Normalized Difference Vegetation Index (NDVI)
- Normalized Difference Water Index (NDWI)
- Distance to Crops cover (m)
- Burned Mask

TO TABLE
FORMAT

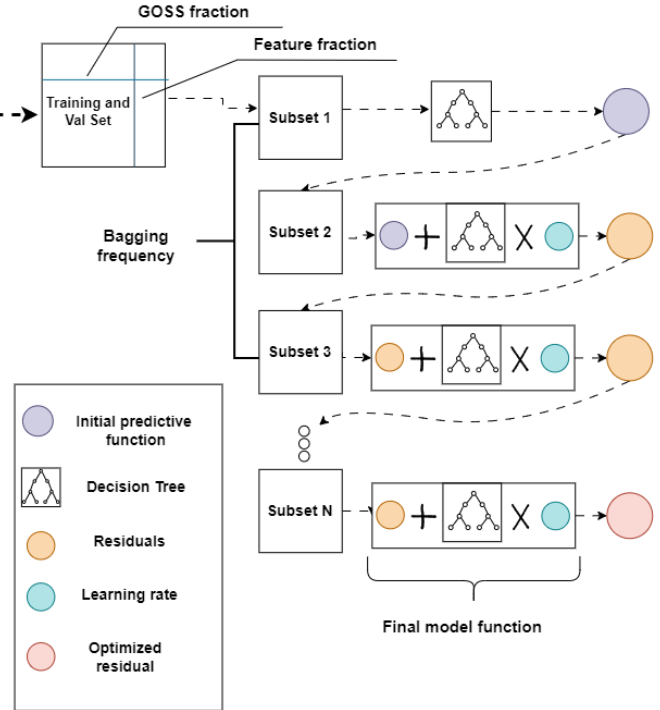
TRAIN -
VAL

TEST



In each iteration

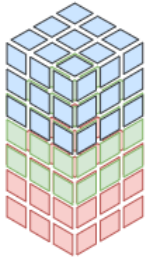
GBDT in LightGBM framework



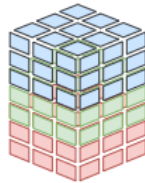
Predicción inicial con Gradient Boosting Decision Tree



Input (16 bands)



- 10 bands of Sentinel-2
- Normalized Burnt Ratio (NBR)
- Burned Area Detection Index (BADI)
- Slope (%)
- Normalized Difference Vegetation Index (NDVI)
- Normalized Difference Water Index (NDWI)
- Distance to Crops cover (m)



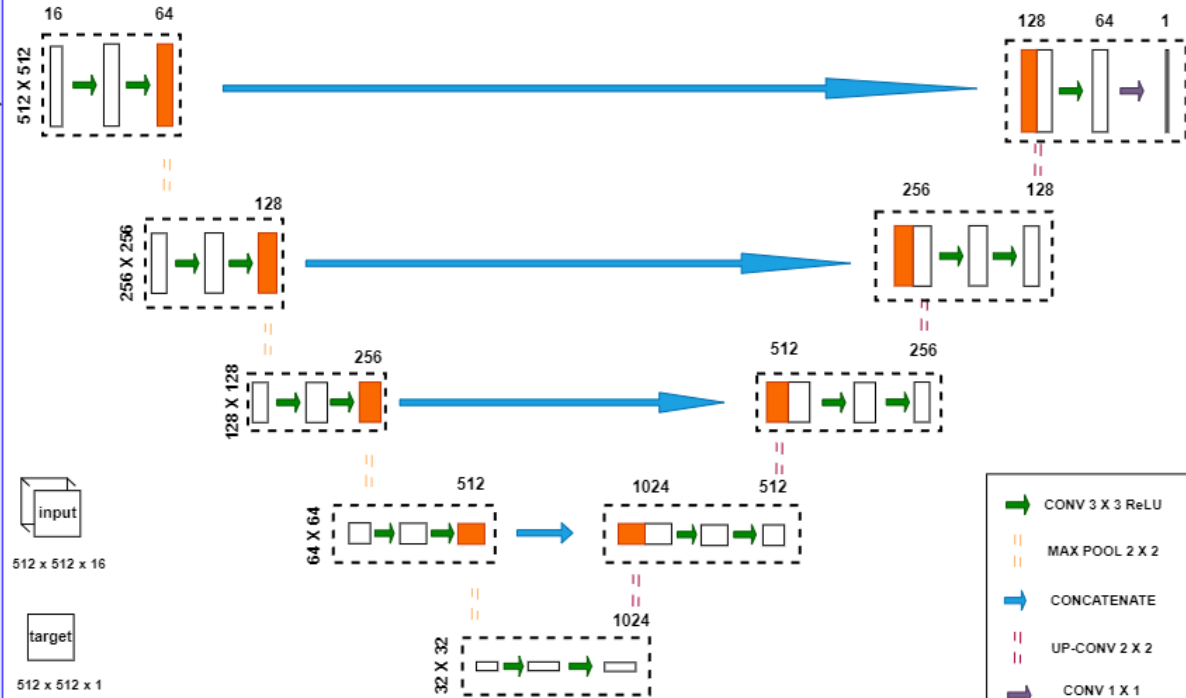
Burned Mask

TRAIN

VAL

TEST

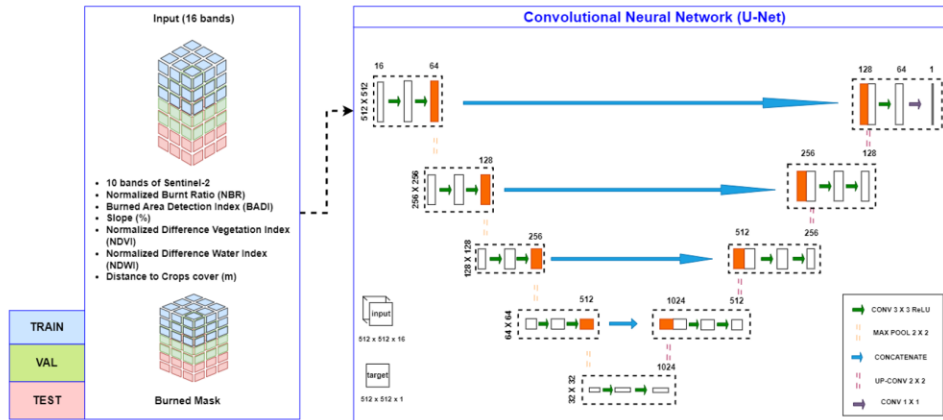
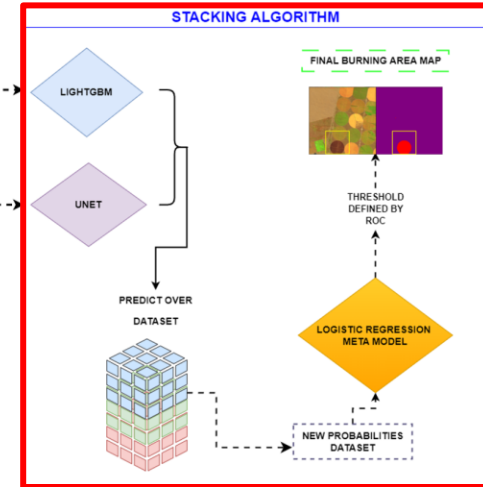
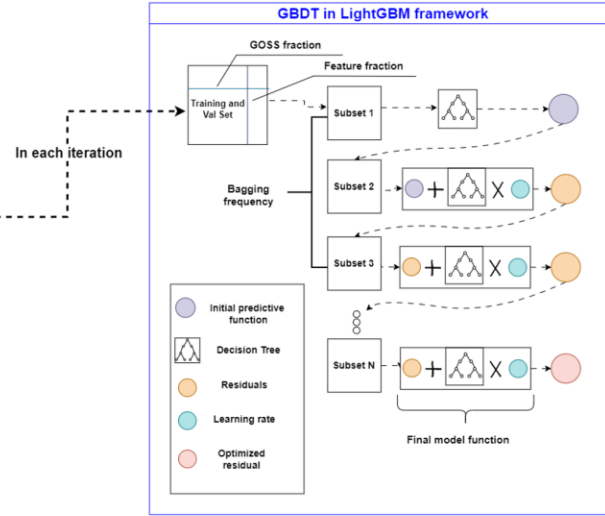
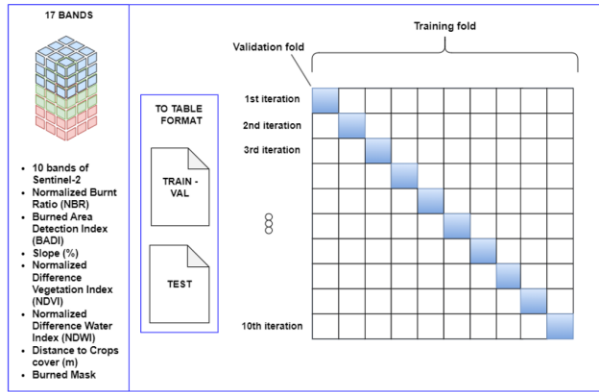
Convolutional Neural Network (U-Net)



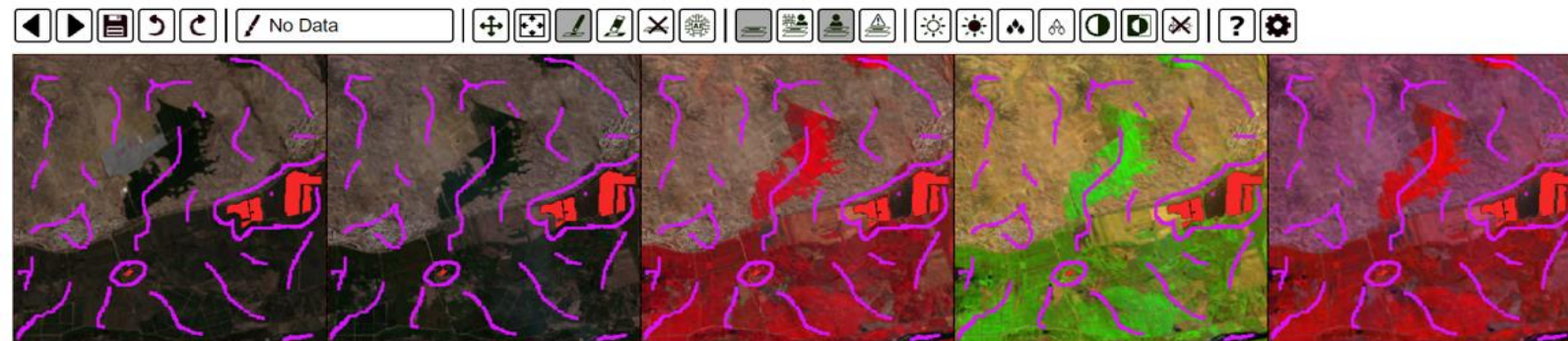
Redes Neuronales Convolucionales (CNN)



Algoritmo de Stacking usando Regresión Logística



05.



100%

0
123

ROI_0011_...11T135646
image

2
Classes

23.9k
Drawn pixels

0
AI-Score

Start the training!

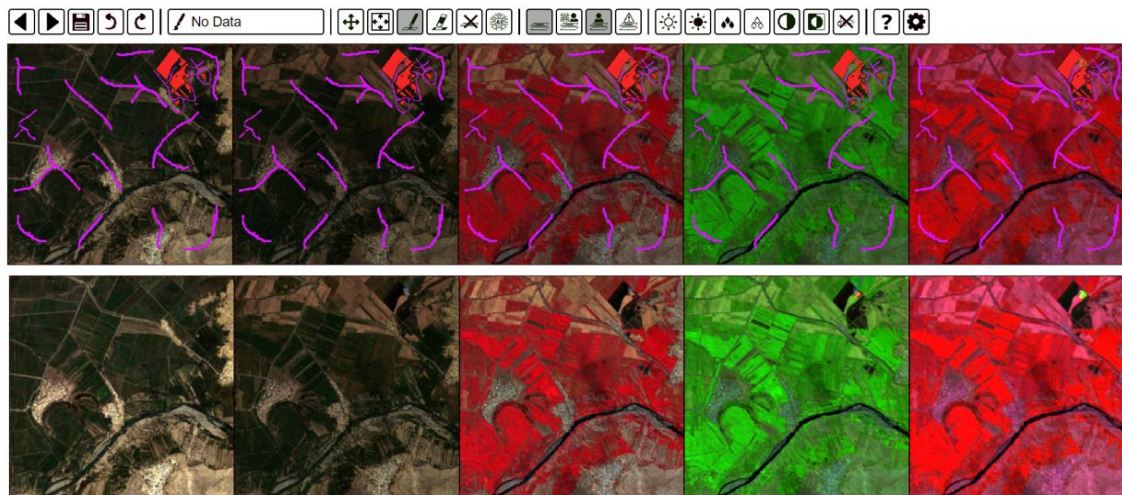
Etiquetado con IRIS

Class selection

Here is an overview about all classes:

	Drawn pixels by user	Description
No Data	0	Layer without data.
Burn Scar	4651	Burned areas.
No Burn Scar	19268	Non burned areas.

Método Scribble



Protocolo de etiquetado de quema en Imágenes Sentinel-2

	Criterio	No Quema (0)	Quema (1)
1	Visualización espectral	Los colores son típicos de otro uso de suelo.	La imagen cumple las tonalidades típicas en las cuatro visualizaciones.
2	Contraste	El área en cuestión, no se diferencia significativamente de sus alrededores.	Hay un contraste evidente con las áreas a su alrededor.
3	Focos de calor	No se detectan focos de calor en la imagen.	Se detectan focos de calor en la imagen, apreciándose el humo.
4	Ubicación y contexto espacial	Es difícil determinar si han ocurrido quemas en el área en cuestión	Está en una ubicación conocida por tener quemas frecuentes.

Natural Color (B4, B3, B2)



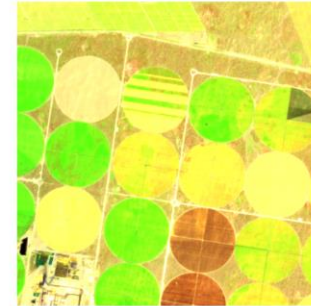
Infrared Color (B8, B4, B3)



Short-Wave Infrared (B8A, B12, B4)



Agriculture (B11, B8, B2)



Natural Color with NBR mask

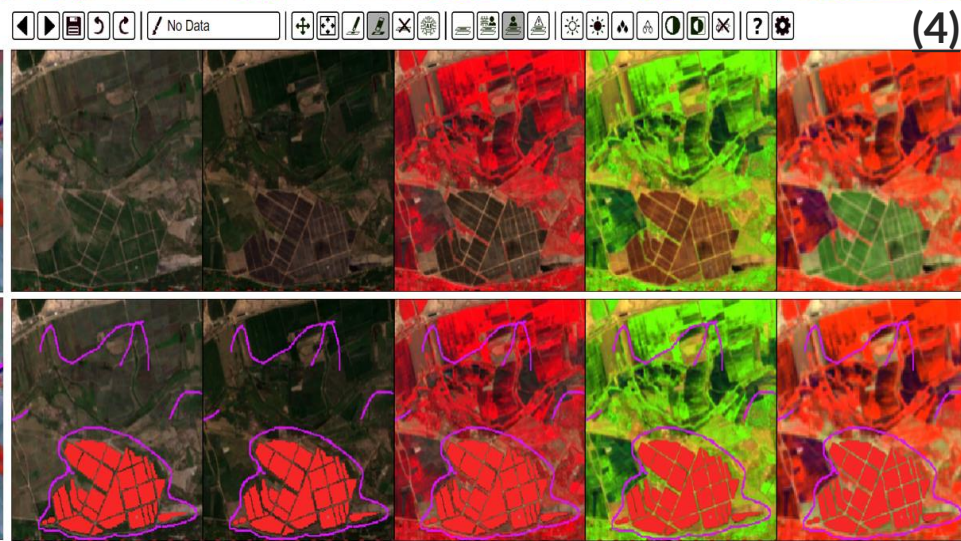
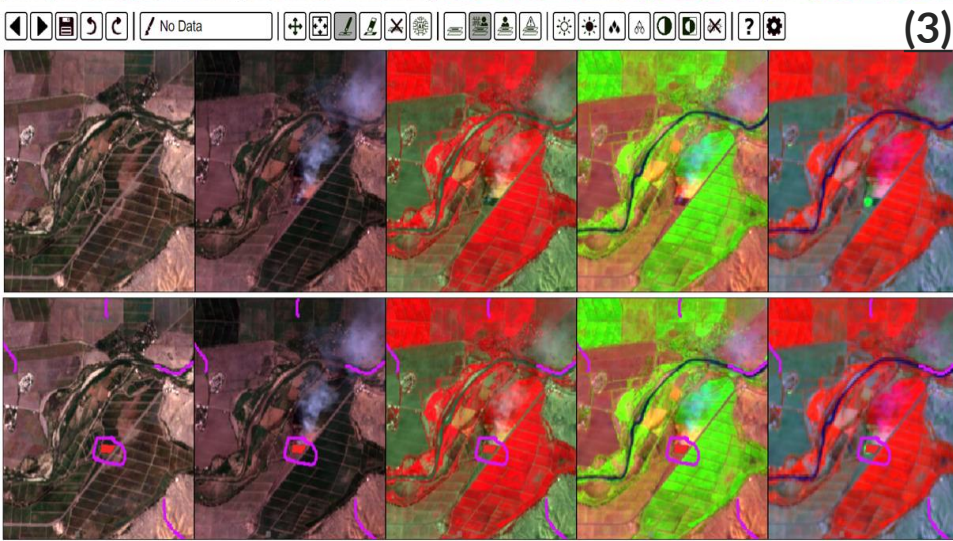
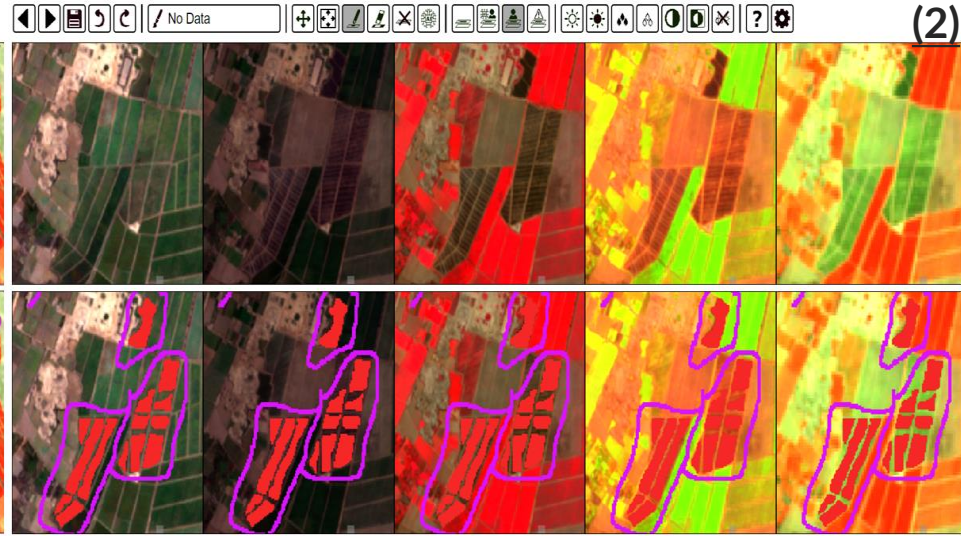
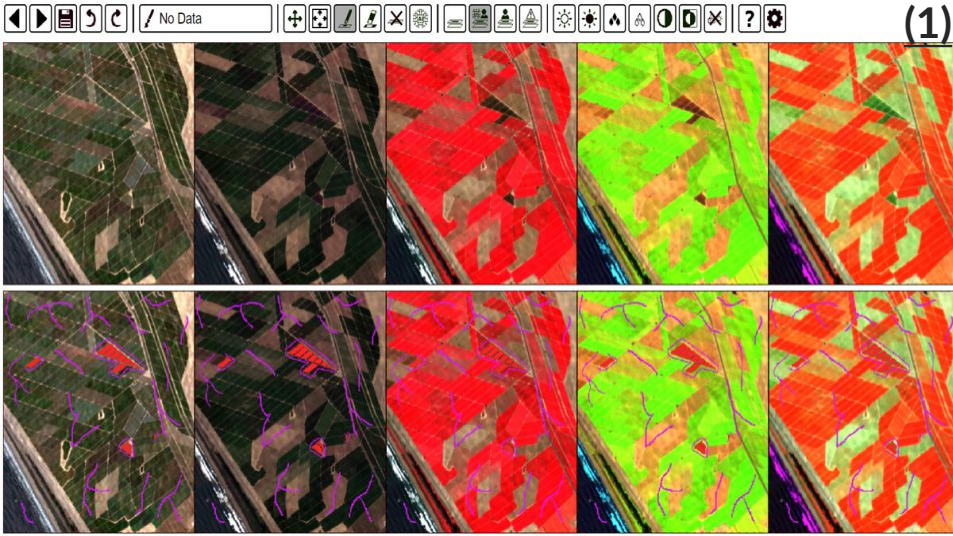


Natural Color with BADI mask



05.

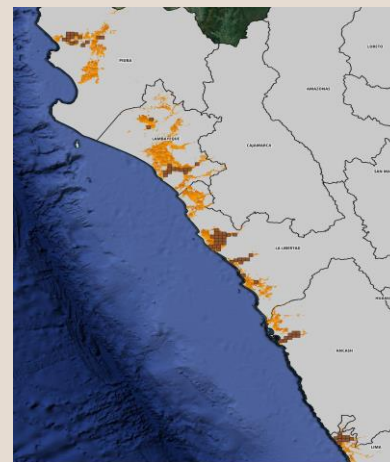




Casos de Validación OEFA

Estas imágenes de validación corresponden a áreas quemadas que son coherentes con los 471 reportes dentro del registro histórico del OEFA. Para este fin, de un total de **478 imágenes analizadas se obtuvieron 58 imágenes** que cumplían con los siguientes criterios:

- La imagen muestra una extensión de quema dentro de un rango de ± 7 días respecto a la fecha reportada al OEFA.
- La imagen muestra una extensión de quema que es visualmente detectable.
- Las coordenadas del evento de quema son espacialmente coherentes con la ubicación mostrada en la imagen.



05.

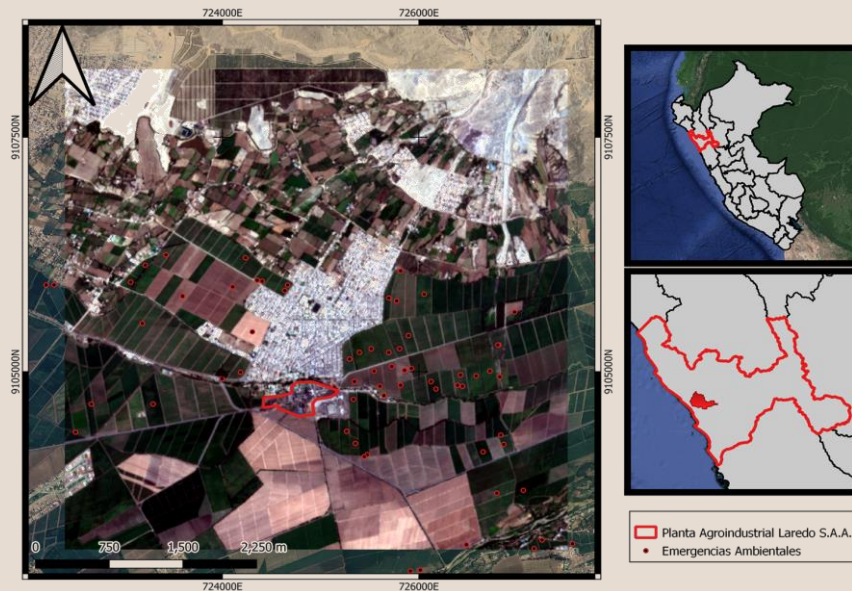


Inferencia en el área piloto Enero – Octubre 2024

El criterio parte de la selección de imágenes por la cobertura de nubosidad menor al 10 % y la detección de áreas pseudo quemadas.

De esta selección se obtuvieron 30 imágenes de Sentinel-2 L2A para el área piloto de Laredo.

Se eligieron conjunto de cinco imágenes que se encuentren dentro de un rango menor a tres meses para ser evaluada como serie de tiempo.





06.

Resultados

"Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world."

- Louis Pasteur

French chemist and microbiologist.



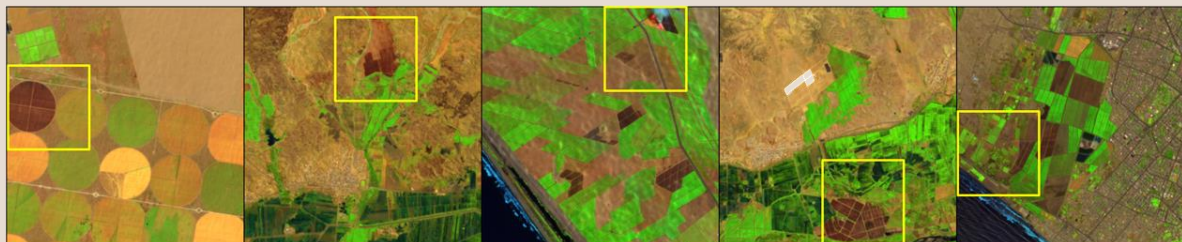
Resultados U-Net

06.

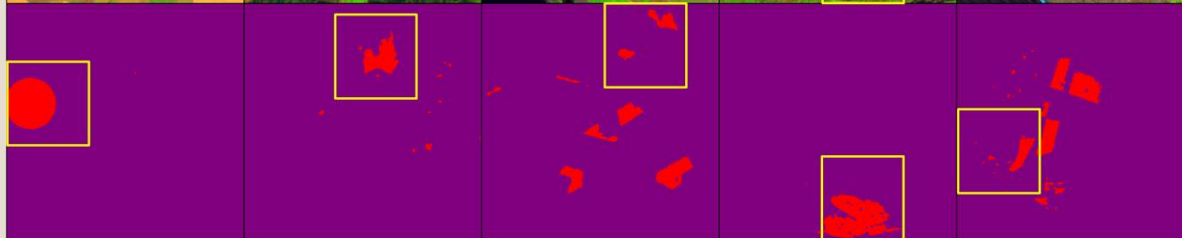
Zoomed Area



Agriculture Color



U-Net ($p \geq 0.75$)

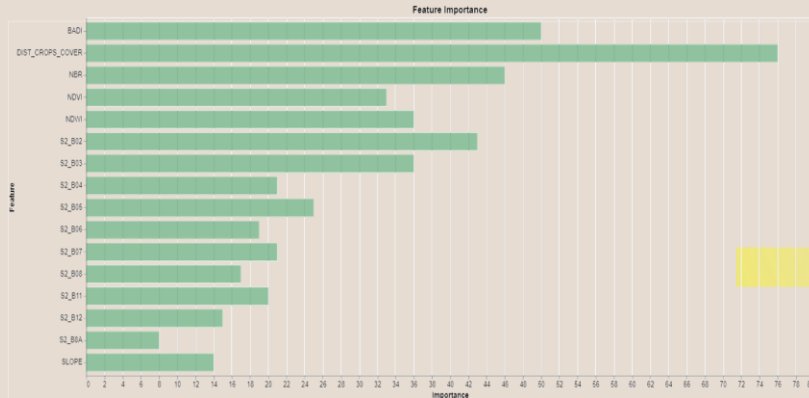


Parámetro	model	encoder_name	encoder_weights	in_channels	classes	lr_init	class_weight	loss
Valor	U-Net	Resnet50	ImagineNet	16	1	0.001	[1, 15]	WCE + Dice Loss

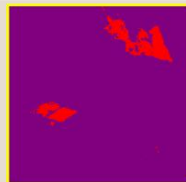


Resultados LightGBM

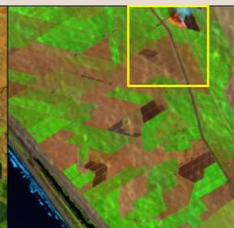
LightGBM prioriza bandas visibles, infrarrojo cercano y onda corta por su sensibilidad a quemas, junto con índices como NDVI y NBR (Quintano et al., 2011; Van Dijk et al., 2021; Lee et al., 2022). La distancia a coberturas agrícolas es clave para diferenciar áreas cultivadas de terrenos no agrícolas.



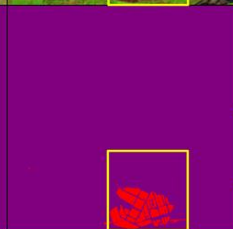
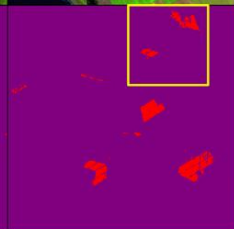
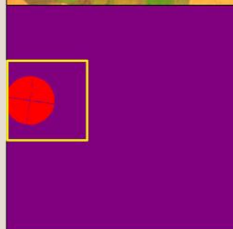
Zoomed Area



Agriculture Color



GBDT ($p \geq 0.75$)



06.



Resultados Stacking Model

06.



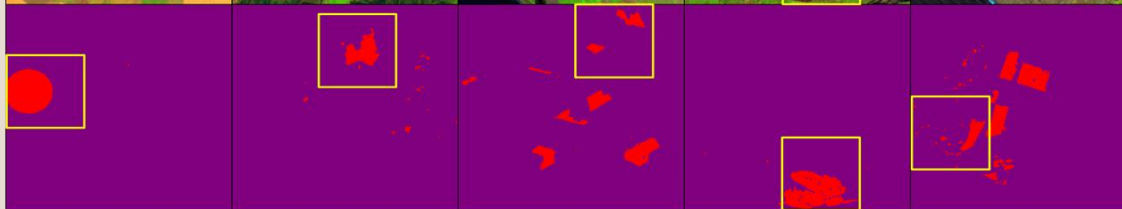
Zoomed Area



Agriculture Color



Stacking ($p \geq 0.75$)



U-Net genera **segmentaciones gruesas**, mientras que LightGBM delimita **áreas de cultivo con mayor precisión**, diferenciando píxeles de caminos (Knopp et al., 2020). Este estudio introduce un modelo de *stacking* que combina ambos enfoques, mejorando la detección (Zhang et al., 2022).

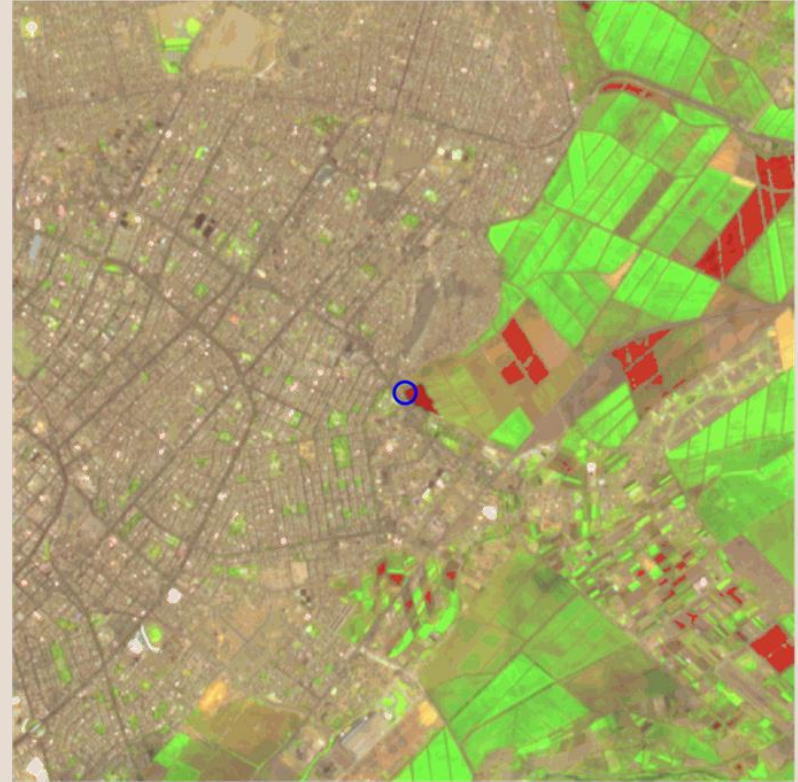
Métrica	U-Net	LightGBM	Stacking Model
F1 – Score (↑)	67.2	86.0	90.6
Recall (↑)	77.2	93.8	95.3
Precision ((↑)	60.2	79.3	86.4
IoU (↑)	51.1	75.4	82.8
Kappa (↑)	67.5	81.2	87.5

Emergency - 0016 on 2022/02/14 | Sentinel-2 on 2022/02/15
Latitude: -8.104, Longitude: -79.001

Agriculture (B11, B8, B2)



Threshold: 0.80



Método convencional

Comparación entre modelos

El modelo supera técnicas basadas en reglas como BADI, integrando distintas fuentes de información para una detección y delimitación más robusta de áreas quemadas.

Emergency - 0036 on 2022/05/29 | Sentinel-2 (2022/05/31)

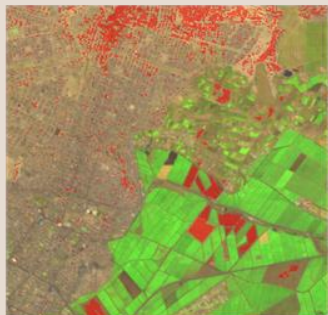
Agriculture (B11, B8, B2)

BADI (threshold=0.15)

UNET (Threshold=0.75)

GBM (Threshold=0.15)

Stacking (Threshold=0.20)



Emergency - 0495 on 2021/04/08 | Sentinel-2 (2021/04/09)

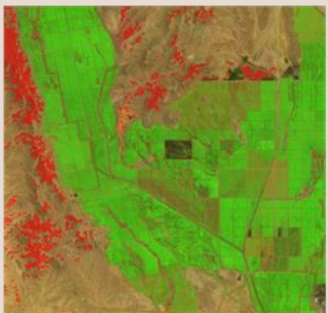
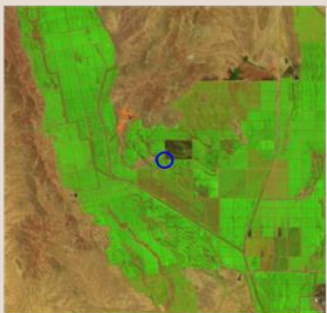
Agriculture (B11, B8, B2)

BADI (threshold=0.15)

UNET (Threshold=0.75)

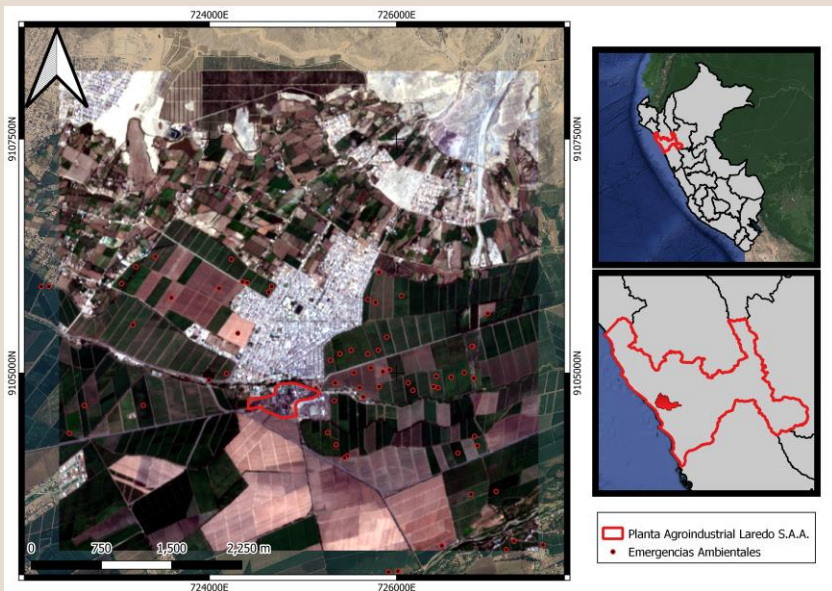
GBM (Threshold=0.15)

Stacking (Threshold=0.20)

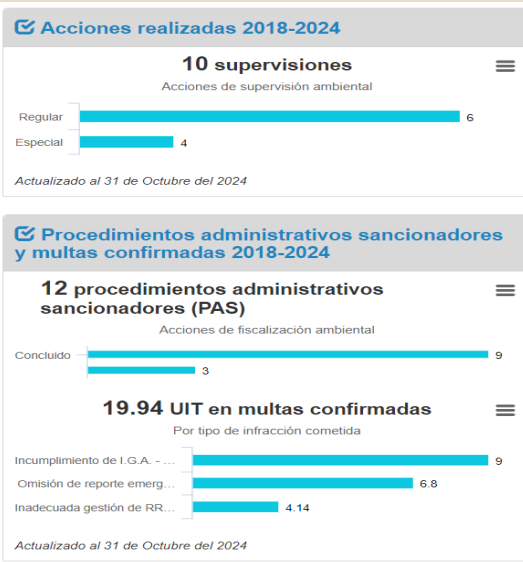
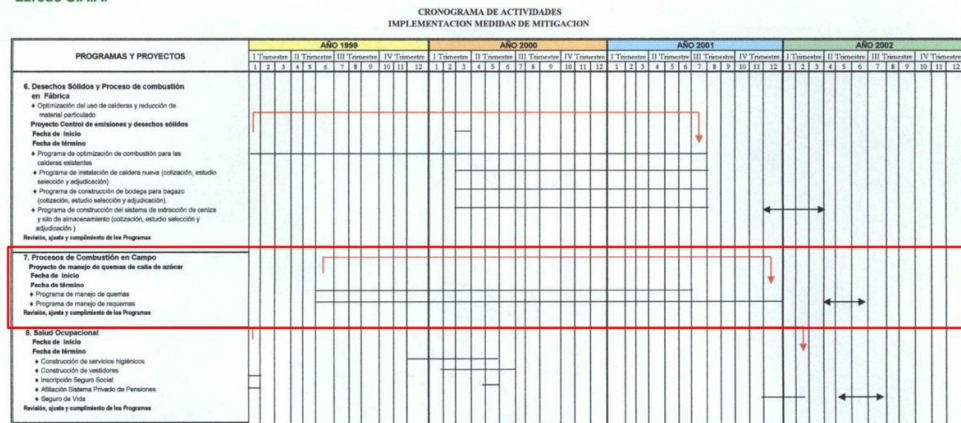


Inferencia en área piloto

Este distrito, con una población proyectada en 2022 de 44,440 habitantes según datos del Instituto Nacional de Estadística e Informática (2022), fue elegido debido a su proximidad a los fundos azucareros propiedad del ingenio azucarero Laredo S.A.A.



Empresa Agroindustrial
Laredo S.A.A.



Esta empresa carece de un instrumento de gestión ambiental (PAMA) actualizado que incluya la cosecha verde y un cronograma más detallado para el manejo de quema.

Multas ascienden a 19.94 UIT = s/. 102,691.00

Inferencia en área piloto

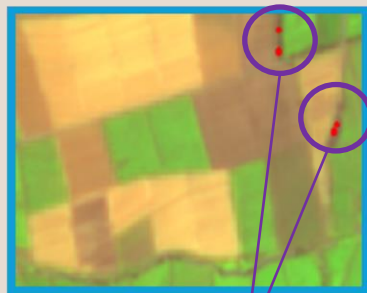
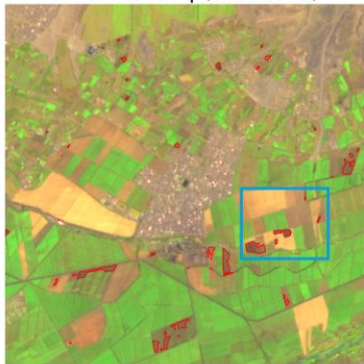
2023-10-13 | (4.67 ha)

2023-10-18 | (19.68 ha)

2023-11-07 | (14.91 ha)

2023-11-22 | (28.72 ha)

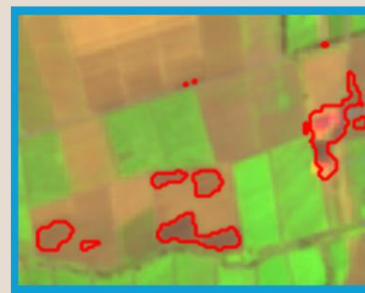
2023-11-27 | (31.32 ha)



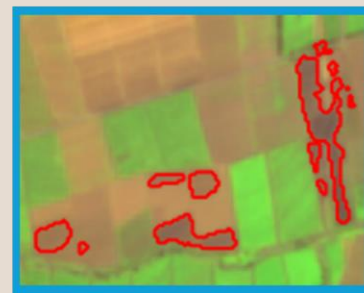
salt n' pepper



4.88 ha de áreas quemadas recientemente



Área con quema activa de 2.94 ha

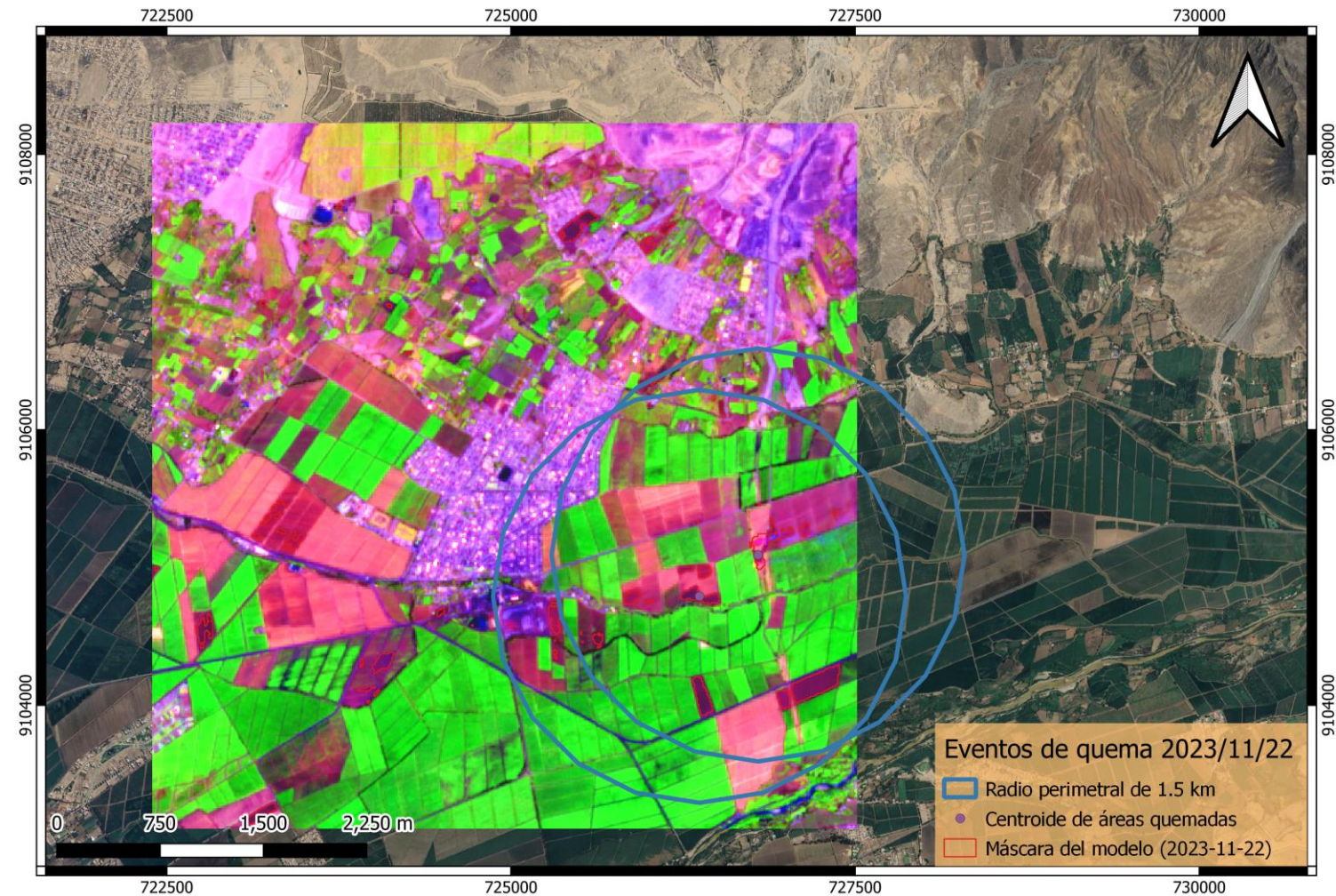


Tono más tenue y una extensión de 4.61 ha.

Try yourself in
Google Colab!!



SCAN ME



Radio perimetral de 1.5km propuesto por el OEFA para la Determinación de distancias mínimas para la quema de caña de azúcar en las localidades de La Huaca y La Rinconada.



S2 Image Date: 2022-10-30 with Latitude: -9.158 & Longitude: -78.347

Sentinel-2 Agriculture Color

BADI (Threshold ≥ 0.30)

U-Net Probability (Threshold ≥ 0.75)

GBM Probability (Threshold ≥ 0.15)

Stacking Probability (Threshold ≥ 0.20)



S2 Image Date: 2018-12-13 with Latitude: -6.0847 & Longitude: -80.0554

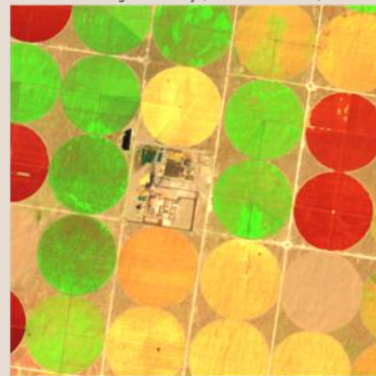
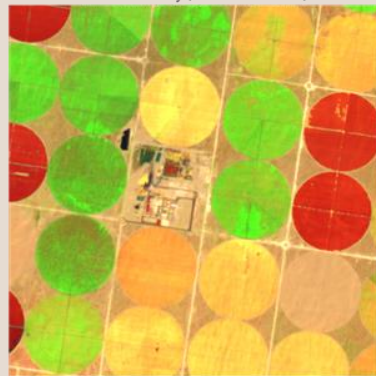
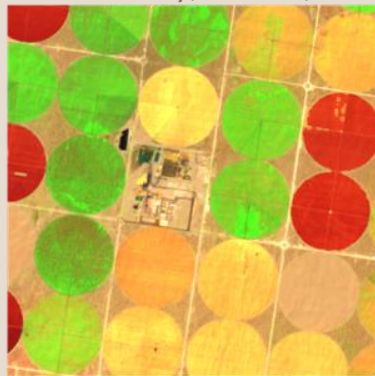
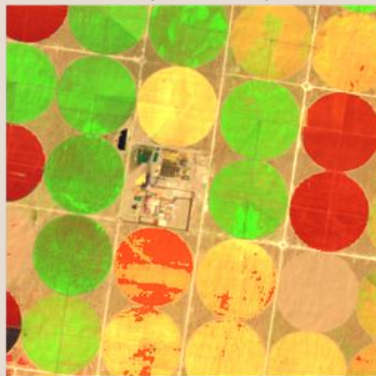
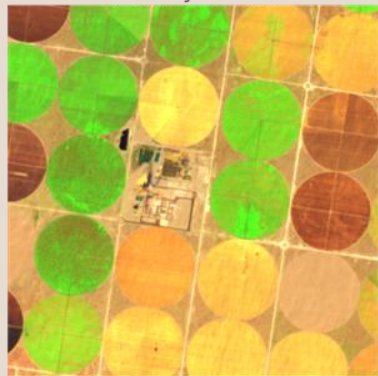
Sentinel-2 Agriculture Color

BADI (Threshold ≥ 0.30)

U-Net Probability (Threshold ≥ 0.75)

GBM Probability (Threshold ≥ 0.15)

Stacking Probability (Threshold ≥ 0.20)





07.

Conclusiones

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.”

- Sir William Bragg

British physicist and Nobel laureate.





- Este estudio confirma que es posible desarrollar un modelo de aprendizaje ensamblado para detectar la quema de caña de azúcar. Combinando aprendizaje profundo y clasificación a nivel de píxel, el modelo logró 100 % de efectividad en **58 casos de validación del OEFA, utilizando un umbral de 0.20 de la curva ROC.**
- El modelo fue entrenado con **1,054 imágenes Sentinel-2, índices de vegetación (NDVI, NDWI) y variables topográficas (pendiente, distancia a cultivos).** Las etiquetas de áreas quemadas se generaron mediante detección temporal basada en reglas y validación visual.
- Los algoritmos más efectivos para el modelo ensamblado son **U-Net**, que abstrae características espaciales, y **LightGBM**, un clasificador basado en árboles de decisión. Ambos enfoques han demostrado ser complementarios en la detección de áreas quemadas.
- **El modelo de Stacking supera a U-Net y LightGBM en las métricas evaluadas, mostrando delimitación equilibrada y aprovechando la coherencia de U-Net y la clasificación precisa de LightGBM. Presenta un desempeño superior con 82.8 % IoU, 90.6 % F1-Score y 87.5 % Kappa. El valor de F1-Score es el predilecto cuando se quiere mantener un equilibrio entre identificar correctamente las quemas y evitar las falsas alarmas (falsos positivos).**





08.

Discusión y Recomendaciones

*"Research is formalized curiosity. It is
poking and prying with a purpose."*

- Zora Neale Hurston.
Anthropologist



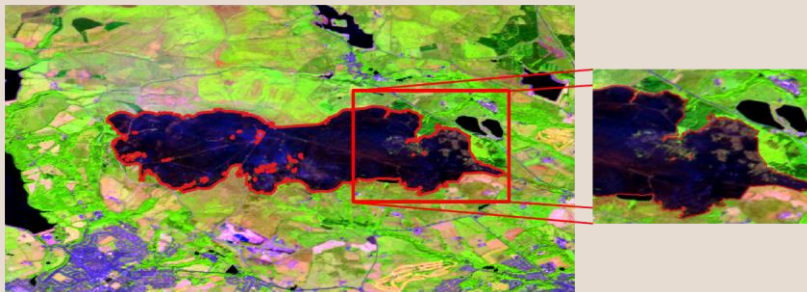
1. Generación y etiquetado del conjunto de datos

La detección de **áreas quemadas en cultivos de caña de azúcar** enfrenta dificultades por la necesidad de bases de datos confiables.

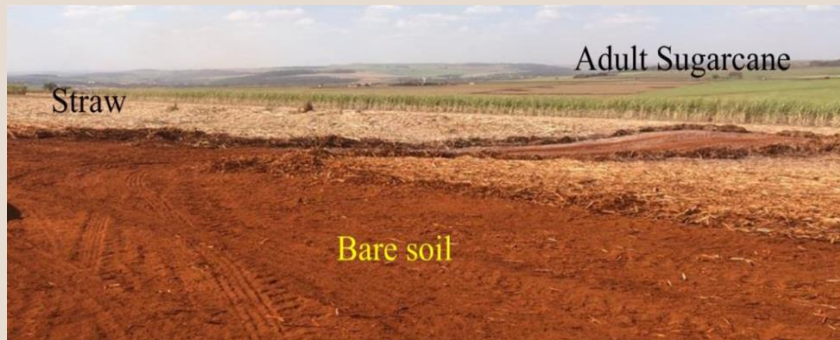
Este estudio propuso una metodología específica para etiquetar áreas quemadas, superando limitaciones de modelos globales que no logran capturar áreas menores a 100 hectáreas en contextos locales. (Arnaudo et al., 2023; Seydi et al., 2022).

El uso de "scribbles" puede generar ambigüedades en áreas no etiquetadas, afectando la segmentación, **especialmente en áreas parcialmente quemadas o con recuperación avanzada** (Knopp et al., 2020). Este problema es crítico para definir cuándo una zona deja de ser clasificada como quemada.

En el Perú, la caña de azúcar **no tiene un período único de zafra, permitiendo quemas durante todo el año** (Helfgott, 2016). Las quemas y requemas frecuentes dificultan distinguir entre suelos desnudos y áreas quemadas, ya que estas últimas pueden no mostrar vegetación en imágenes previas a la quema.



Se visualiza la etiqueta EMSR291_AOI02_01 de 714 ha de área quemada. Extraído de Arnaudo et al. (2023).



La imagen muestra tres estados: caña de azúcar madura (adult sugarcane), suelo desnudo (bare soil) y residuos de caña (straw). Extraído de Campos et al. (2022).



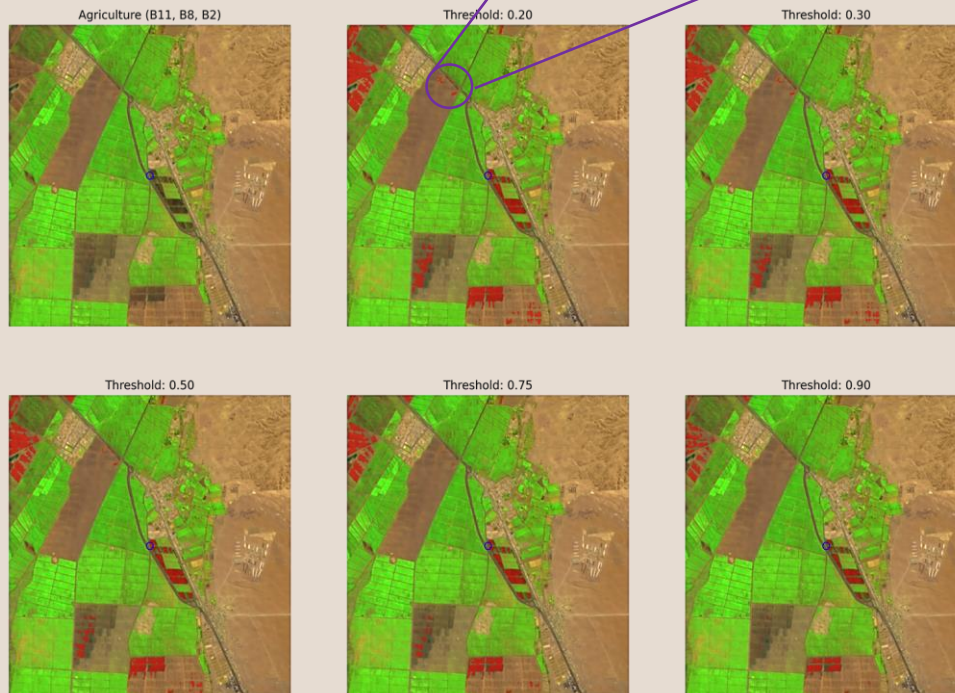
2. Rendimiento de los modelos base

El fenómeno *salt n' pepper* introduce ruido en la clasificación por píxeles, afectando la precisión, especialmente en condiciones complejas (Sdraka et al., 2024). Aunque el modelo ensamblado mitiga estos problemas parcialmente, sigue siendo sensible a la confusión espectral.

Umbrales altos mejoran la precisión reduciendo falsos positivos, mientras que umbrales bajos amplían la detección. La selección del umbral depende de las necesidades de análisis del OEFA.

El modelo ensamblado supera técnicas tradicionales basadas en reglas como BADI, integrando distintas fuentes de información para una detección y delimitación más robusta de áreas quemadas.

El modelo identificó eficazmente quemas activas y recientes (>1 ha), diferenciando áreas quemadas de zonas urbanas, cuerpos de agua y suelo desnudo, brindando fiabilidad para la toma de decisiones.



08.



Actualización de los Instrumentos de Gestión Ambiental:

- Revisar y actualizar los Instrumentos de Gestión Ambiental (IGA) de las empresas azucareras, incluyendo los Planes de Adecuación y Manejo Ambiental (PAMA) para empresas en operación y los Estudios de Impacto Ambiental (EIA) para la ampliación, modificación o formalización de sus operaciones.
- Esta actualización debe incorporar la implementación de la **cosecha mecanizada en verde** como una alternativa viable a la quema de caña de azúcar
- El modelo servirá como herramienta de vigilancia y monitoreo del OEFA para una mejor fiscalización para el cumplimiento de los compromisos ambientales estipulados en los IGAs que se creen.



Realizar IGA

Actualizar IGA

ADM11837	AGROINDUSTRIAL LAREDO S.A.A.
ADM15042	EMPRESA AZUCARERA DEL NORTE S.A.C.

ADM11768	CENTRAL AZUCARERA CHUCARAPI PAMPA BLANCA S.A.
ADM11868	EMPRESA AGROINDUSTRIAL TUMAN S.A.A.
ADM12292	EMPRESA AGROINDUSTRIAL POMALCA S.A.A.
ADM12297	EMPRESA AGROINDUSTRIAL CAYALTI S.A.A
ADM14614	INDUSTRIAL PUCALA S.A.C.
ADM11701	EMPRESA AGRARIA AZUCARERA ANDAHUASI SAA.
ADM19633	EMPRESA AGRARIA CHIQUITOY S.A.
ADM19720	EMPRESA AGRICOLA SINTUCO S.A.
ADM14396	JM UCAYALI S A C
ADM19037	LAON S.A.C.
ADM19755	VALLESOL S.A.C.



- Realizar estudios específicos sobre el ciclo de cultivo de la caña de azúcar: Implementar investigaciones detalladas utilizando imágenes satelitales para monitorear todas las etapas del proceso de cultivo de caña de azúcar, desde la preparación del terreno hasta la cosecha, para identificar y gestionar mejor las quemadas de manera remota.
- Incluir detalles en los reportes de emergencia y de recuperación de campos: Incorporar en los reportes de emergencia datos específicos sobre el área quemada estimada, junto con información detallada proporcionada por los administrados, como cronogramas de fechas planificadas para las quemadas, la extensión estimada de las áreas afectadas por las quemadas una vez realizadas, y el tiempo estimado de recuperación de las tierras hasta que estén listas para una nueva siembra. Esta información adicional permitirá una validación más precisa y temporalmente coherente de los modelos mediante el uso de imágenes satelitales, mejorando así la capacidad de monitoreo y gestión de las áreas quemadas.
- Aumentar la base datos con nuevas imágenes en diferentes condiciones: Ampliar la base de datos con imágenes de satélite en diferentes condiciones climáticas, de iluminación y diversas con presencia de superficies similares a la quema en términos de reflectancia, para mejorar la generalización del modelo y su capacidad para detectar áreas quemadas en diversas situaciones.





**Thank you
for the
attention!
Q & A**

