

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



**APLICACIÓN DE TÉCNICAS DE MACHINE LEARNING E IMÁGENES DE
RADAR PARA LA DETECCIÓN TEMPRANA DE INVASIONES EN ZONAS
DE ALTO RIESGO DE DESASTRES**

Tesis para obtener el título profesional de **INGENIERO CIVIL**

AUTOR:

Javier Alonso Jaimes Cucho

ASESOR:

Luis Angel Moya Huallpa

Lima, octubre, 2023

Informe de Similitud

Yo, Luis Angel Moya Huallpa, docente de la Facultad de Ciencias e Ingeniería de la Pontificia Universidad Católica del Perú, asesor de la tesis/el trabajo de investigación titulado

Aplicación de técnicas de machine learning e imágenes de radar para la detección temprana de invasiones en zonas de alto riesgo de desastres, del autor Jaimes Cucho, Javier Alonso

dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 20%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 15/09/2023.
- He revisado con detalle dicho reporte y la Tesis o Trabajo de Suficiencia Profesional, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha: Lima, 5 de octubre del 2023

Apellidos y nombres del asesor: Moya Huallpa, Luis Angel	
DNI: 43010087	Firma 
ORCID: 0000-0003-1764-3160	

Resumen

La presente tesis aborda la problemática de las invasiones de terrenos por grupos vulnerables. En la mayoría de casos estos grupos se asientan en zonas de alto riesgo de desastres debido a fenómenos naturales. Lo expuesto previamente se evidenció en los procesos migratorios del siglo pasado y en invasiones más recientes donde grupos de personas vulnerables se asentaron en zonas costeras periféricas a las ciudades. Estas zonas según los distintos mapas elaborados por el SIGRID y CISMID tienen mayor probabilidad de ocurrencia de desastres.

Por lo tanto, esta investigación tiene por finalidad identificar, de forma temprana y remota, la creciente tasa de asentamientos informales en zonas de alto riesgo de desastres. Para tal propósito se plantea una metodología que permita detectar estructuras y patrones de asentamientos informales. Para la detección de invasiones se emplean diversas técnicas de *machine learning* empleando imágenes satelitales de radar, de libre acceso, de media resolución (10m) y técnica de postprocesamiento para la mejora en el desempeño de la predicción.

Para la evaluación de la metodología planteada se empleó como caso de estudio la invasión en Lomo de Corvina, ocurrido en abril del 2021. Para las áreas invadidas se obtuvo valores promedio de *precision* del 39%, lo cual es indicador que los algoritmos sobrestiman las áreas invadidas debido a las distorsiones complejas y ruido en las imágenes de radar, y *recall* del 85%, lo que indica que los algoritmos identifican correctamente un alto porcentaje del área invadida. Por lo tanto, se puede emplear esta metodología para la detección temprana de áreas invadidas con características similares a las estudiadas. Se lograrán mejores resultados si las invasiones son repentinas, de gran extensión y están ubicadas en zonas poco accidentadas y sin cubierta vegetal.

Palabras clave: Teledetección, aprendizaje de máquina, asentamientos informales, gestión de riesgo, invasiones repentinas

Abstract

This thesis addresses the problem of land invasions by vulnerable groups. In most cases, these groups settle in areas at high risk of disasters due to natural phenomena. The aforementioned was evidenced in the migratory processes of the last century and in more recent invasions where groups of vulnerable people settled in peripheral coastal areas of the cities. These areas, according to the different maps prepared by SIGRID and CISMID have a higher probability of disaster occurrence.

Therefore, this research aims to identify, early and remotely, the increasing rate of informal settlements in high disaster risk areas. For this purpose, a methodology is proposed to detect structures and patterns of informal settlements. For the detection of invasions, several machine learning techniques are used using radar satellite images, freely accessible, of medium resolution (10m) and post-processing technique to improve the prediction performance.

For the evaluation of the proposed methodology, the invasion in Lomo de Corvina, which occurred in April 2021, was used as a case study. For the invaded areas, average precision values of 39% were obtained, which is an indicator that the algorithms overestimate the areas. invaded due to complex distortions and noise in the radar images, and recall of 85%, indicating that the algorithms correctly identify a high percentage of the invaded area. Therefore, this methodology can be used for early detection of invaded areas with characteristics similar to those studied. Better results will be achieved if the invasions are sudden, extensive and are located in areas with little hilly terrain and no vegetation cover.

Keywords: Remote sensing, machine learning, informal settlements, risk management, sudden invasions

Agradecimientos

A mi asesor, el Dr. Luis Moya, por su orientación, disponibilidad, motivación, apoyo incondicional y sobre todo por brindarme la confianza y oportunidad de desarrollar el presente proyecto.

A mi familia por su apoyo y soporte incondicional, quienes me ayudaron en todo momento y contribuyeron en gran medida a ser quien soy y estar donde estoy.

A la PUCP por brindarme la oportunidad de realizar mis estudios universitarios y conocer increíbles personas.

A los profesores de la sección de Estudios Generales Ciencias, Ciencias e Ingeniería y maestría en ingeniería civil de la PUCP por las enseñanzas transmitidas durante estos años de estudio.

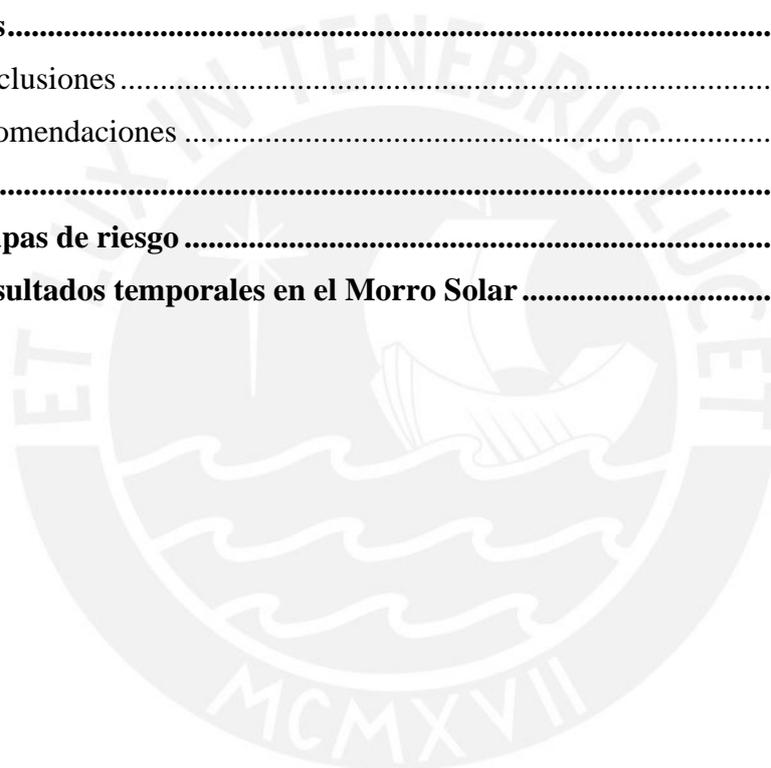
A CONCYTEC/Programa PROCIENCIA por financiar esta investigación a través del proyecto “Inventario de ocupaciones informales de terrenos en Lima Metropolitana mediante imágenes satelitales de radar” (Proyecto No. PE501077930-2022).

Índice

Resumen	i
Abstract	ii
Agradecimientos	iii
Índice	iv
Lista de figuras	vii
Lista de tablas	ix
Capítulo 1	
Generalidades	1
1.1 Introducción	1
1.2 Objetivos	3
Objetivo general	3
Objetivos específicos	3
1.3 Pregunta de investigación	3
1.4 Problemática y justificación.....	4
Capítulo 2	
Marco conceptual	7
2.1 Evaluación del peligro sísmico en el Perú	7
2.2 Políticas de gestión de riesgos en el Perú	9
2.3 Crecimiento urbano en Lima Metropolitana.....	11
2.4 Machine learning	12
Decision Tree.....	13
Gaussian Naïve Bayes	¡Error! Marcador no definido.
Random Forest.....	¡Error! Marcador no definido.
Support Vector Machine.....	14
Stochastic Gradient Descent.....	14
K-Neighbors	14
Logistic Regression	14
Neuronal Network	14
Elliptic Envelop	15

Local Outlier Factor	15
2.5 Detección de valores atípicos.....	15
2.6 Matriz de confusión	17
2.7 Métricas de evaluación de resultados	18
2.8 Teledetección (remote sensing)	19
Teledetección para observaciones terrestres.....	19
2.9 Imágenes de radar y SAR	20
2.10 Proyecto Sentinel	26
2.11 Operadores morfológicos.....	27
Capítulo 3	
Estado del arte	30
3.1 Detección de ASENTAMIENTOS INFORMALES	30
3.2 Metodología general para la clasificación digital de imágenes	34
3.3 Metodología actual para la clasificación de imágenes multiespectrales.....	36
Capítulo 4	
Metodología.....	37
4.1 Introducción	37
4.2 Herramientas, métodos y procedimientos a usar	38
4.3 Obtención y procesamiento de imágenes satelitales SAR.....	40
Apply-Orbit-File.....	41
Radiometric Calibration	41
Speckle-Filter	42
Terrain-Correction.....	42
Linear to from dB	43
Resampling y subset de imágenes empleando un sistema de grillas.....	44
4.4 Detección de invasiones.....	44
4.5 Limitaciones.....	45
Capítulo 5	
Evaluación experimental.....	46
5.1 Caso de estudio: invasiones en el Morro Solar, Chorrillos, y Lomo de Corvina, Villa El Salvador del 2021	46
5.2 Descripción de los datos	47
5.3 Resultados cuantitativos.....	52
Árbol de decisión.....	53
Bayes Ingenuo Gaussiano.....	55

Bosques Aleatorios	57
Máquina de Soporte Vectorial	59
Descenso del Gradiente Estocástico	61
K-Vecinos Cercanos	63
Regresión Logística	65
Envolverte Elíptica	67
Factor Local Atípico	69
Redes Neuronales	71
5.4 Discusión de resultados	73
Capítulo 6	
Conclusiones	75
6.1 Conclusiones	75
6.2 Recomendaciones	76
Referencias	77
Anexo 1: Mapas de riesgo	83
Anexo 2: Resultados temporales en el Morro Solar	87



Lista de figuras

Figura 1: Número de artículos por año relacionados con investigaciones en teledetección.	2
Figura 2: Factores de vulnerabilidad	9
Figura 3: Factores de peligro	9
Figura 4: Esquema de un árbol de decisiones.....	13
Figura 5: Tipos de valores atípicos.....	16
Figura 6: Matriz de confusión.....	17
Figura 7: Detección de inundaciones en Puerto Maldonado.....	19
Figura 8: Espectro electromagnético. Nota: Traducido de Khorram et. al. (2016) [42]	20
Figura 9: Flujo de trabajo para la obtención de imágenes digitales.....	21
Figura 10: Mecanismo de reflexión en distintas superficies.	22
Figura 11: Ventana atmosférica.	22
Figura 12: Atenuación de la radiación espectral electromagnética en la atmósfera.	23
Figura 13: Polarización idéntica.	24
Figura 14: Penetración de diferentes longitudes de onda en diferentes cubiertas.	24
Figura 15: Efectos geométricos en una imagen SAR	25
Figura 16: Ejemplo de aplicación del operador de Erosión.....	28
Figura 17: Ejemplo de aplicación del operador de Dilatación	28
Figura 18: Ejemplo de aplicación del operador de Apertura	29
Figura 19: Ejemplo del operador de Cierre	29
Figura 20: Número de artículos por año relacionado con investigación en asentamientos informales empleando la teledetección.	30
Figura 21: Variables comúnmente analizadas en teledetección.	31
Figura 22: Imágenes ópticas de la constelación Sentinel-1 sobre Lima en el año 2022	33
Figura 23: Variación promedio del índice de retrodispersión en superficie con diferente rugosidad. Fuente: [44]	37
Figura 24: Flujo de trabajo para el preprocesamiento de imágenes SAR en el software SNAP.....	41
Figura 25: Izquierda: Imagen con órbita estándar. Derecha: con órbita precisa	41
Figura 26: Izquierda: imagen Sentinel-1 en valores digitales. Derecha: Imagen luego de aplicar la calibración radiométrica.....	42
Figura 27: Imagen Sentinel-1 antes (izquierda) y después (derecha) de aplicar el filtro Speckle	42
Figura 28: Imagen Sentinel-1 antes (izquierda) y después (derecha) de la aplicación del proceso Terrain correction. El eje vertical en la imagen de la izquierda representa la dirección de vuelo del satélite y el eje horizontal es la dirección perpendicular a la dirección de vuelo. El eje vertical y horizontal de la imagen derecha corresponden a las coordenadas latitud y longitud del sistema WGS84.....	43
Figura 29: Imagen Sentinel-1 antes (izquierda) y después (derecha) de la aplicación del escalamiento logarítmico	43
Figura 30:(a) Grilla definida en la ciudad de Lima. (b) Cuadrículas de estudio	44
Figura 31: Áreas de estudio. (a) Localización de las áreas de estudio y tipo de zonas. Fuente: CISMID. (b) Invasión en Lomo de Corvina. Fuente: El Comercio. (c) Invasión del Morro Solar. Fuente: El Comercio	46
Figura 32: Distribución de la retrodispersión en el área de estudio.	48
Figura 33: Distribución de la retrodispersión en una zona urbana.....	48
Figura 34: Distribución de la retrodispersión en una zona no urbana.....	49
Figura 35: (a) Distribución de la retrodispersión del área invadida. (b) Polígono verde que representa el área invadida en Lomo de Corvina.	50
Figura 36: Distribución de la retrodispersión antes de la invasión (03/12/20) y durante la invasión (14/04/21).	50

Figura 37: Variación de la intensidad promedio de la retrodispersión en el área invadida.	51
Figura 38: (a) Distribución de la retrodispersión en la zona del Morro Solar. (b) Polígono definido dentro del área invadida en el Morro Solar.	51
Figura 39: (a) Área predicha por el algoritmo de Árbol de decisión como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	53
Figura 40: Resultados temporales mostrados por el algoritmo de Árbol de decisión.	54
Figura 41: (a) Área predicha por el algoritmo de Bayes ingenuo Gaussiano como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	55
Figura 42: Resultados temporales mostrados por el algoritmo de Bayes ingenuo Gaussiano.....	56
Figura 43: (a) Área predicha por el algoritmo de Bosques aleatorios como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	57
Figura 44: Resultados temporales mostrados por el algoritmo de Bosques aleatorios	58
Figura 45: (a) Área predicha por el algoritmo de Máquina de soporte vectorial como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	59
Figura 46: Resultados temporales mostrados por el algoritmo de Máquina de soporte vectorial	60
Figura 47: (a) Área predicha por el algoritmo de Descenso del gradiente estocástico como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina.....	61
Figura 48: Resultados temporales mostrados por el algoritmo de Descenso del gradiente estocástico....	62
Figura 49: (a) Área predicha por el algoritmo de K-vecinos cercanos como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	63
Figura 50: Resultados temporales mostrados por el algoritmo de K-vecinos cercanos	64
Figura 51: (a) Área predicha por el algoritmo de Regresión logística como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	65
Figura 52: Resultados temporales mostrados por el algoritmo de Regresión logística	66
Figura 53: (a) Área predicha por el algoritmo de Envolverte elíptica como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	67
Figura 54: Resultados temporales mostrados por el algoritmo de Envolverte elíptica	68
Figura 55: (a) Área predicha por el algoritmo de Factor atípico local como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	69
Figura 56: Resultados temporales mostrados por el algoritmo de Factor atípico local	70
Figura 57: (a) Área predicha por el algoritmo de Redes neuronales como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina	71
Figura 58: Resultados temporales mostrados por el algoritmo de Redes neuronales	72
Figura 59: Lomo de Corvina. Fuente: INDECI [70].....	74
Figura 60: Sectores más crítico a monitorear según el tipo de zona	76
Figura 61: Mapa de riesgos por sismo en Lima Metropolitana y Callao. Fuente: INDECI-DIPRE.....	83
Figura 62: Mapa de inundación por tsunami en Lima metropolitana y Callao. Fuente: INDECI-DIPRE.....	84
Figura 63: Zonas susceptibles a movimientos de masas. Fuente: INDECI-DIPRE.....	85
Figura 64: Zonas probables a licuación de suelos. Fuente: INDECI-DIPRE.....	86
Figura 65: Resultados temporales mostrados por el Árbol de decisión.....	87
Figura 66: Resultados temporales mostrados por Naive Bayes Gaussiano.....	88
Figura 67: Resultados temporales mostrados por los Bosques aleatorios	89
Figura 68: Resultados temporales mostrados por la Máquina de Soporte Vectorial.....	90
Figura 69: Resultados temporales mostrados por el Descenso del gradiente estocástico	91
Figura 70: Resultados temporales mostrados por K-Vecinos Cercanos.....	92
Figura 71: Resultados temporales mostrados por la regresión logísticas	93
Figura 72: Resultados temporales mostrados por la Envolverte Elíptica.....	94
Figura 73: Resultados temporales mostrados por el Factor Atípico Local.....	95
Figura 74: Resultados temporales mostrados por las redes neuronales.....	96

Lista de tablas

<i>Tabla 1: Detalles de la población potencialmente expuesta a diferentes peligros.</i>	5
<i>Tabla 3: Principales eventos sísmicos en Perú desde 1970</i>	8
<i>Tabla 4: Modelos no supervisados con sus respectivos análogos supervisados</i>	16
<i>Tabla 5: Distribución de valores típicos de retrodispersión.</i>	21
<i>Tabla 6: Ancho de banda, longitud de onda y frecuencia</i>	26
<i>Tabla 7: Productos disponibles ofrecidos por Sentinel 1</i>	26
<i>Tabla 8: Productos disponibles ofrecidos por Sentinel 2</i>	27
<i>Tabla 9: Comparación de algunos métodos de clasificación supervisado</i>	32
<i>Tabla 10: Metodologías empleadas en algunos algoritmos supervisados. Fuente: [43]</i>	35
<i>Tabla 11: Limitaciones generales al clasificar imágenes.</i>	35
<i>Tabla 2: Resultados esperados, herramientas a usar y métodos a utilizar en la investigación.</i>	38
<i>Tabla 12: Modos de adquisición y aplicación. Fuente: [35]</i>	40
<i>Tabla 13: Fecha y características de las imágenes empleadas</i>	47
<i>Tabla 14: Fechas empleadas para el entrenamiento de los algoritmos.</i>	52
<i>Tabla 15: Evaluación de los resultados predichos por el algoritmo de árbol de decisión</i>	53
<i>Tabla 16: Evaluación de los resultados predichos por el algoritmo de Bayes ingenuo Gaussiano</i>	55
<i>Tabla 17: Evaluación de los resultados predichos por el algoritmo de Bosques aleatorios</i>	57
<i>Tabla 18: Evaluación de los resultados predichos por el algoritmo de Máquina de soporte vectorial</i>	59
<i>Tabla 19: Evaluación de los resultados predichos por el algoritmo de Descenso del gradiente estocástico</i>	61
<i>Tabla 20: Evaluación de los resultados predichos por el algoritmo de K-vecinos cercanos</i>	63
<i>Tabla 21: Evaluación de los resultados predichos por el algoritmo de Regresión logística</i>	65
<i>Tabla 22: Evaluación de los resultados predichos por el algoritmo de Envolvente elíptica</i>	67
<i>Tabla 23: Evaluación de los resultados predichos por el algoritmo de Factor atípico local</i>	69
<i>Tabla 24: Evaluación de los resultados predichos por el algoritmo de Redes neuronales</i>	71
<i>Tabla 25: Resumen de resultados</i>	73

Capítulo 1

Generalidades

1.1 Introducción

Según la ONU, Latinoamérica es la segunda región más urbanizada del mundo con aproximadamente un 81% de la población viviendo en zonas urbanizadas [1]. El crecimiento se produjo a causa de la rápida urbanización y excedente de la mano de obra producto de la migración rural del siglo XX. No obstante, como consecuencia de este fenómeno y desigualdades sociales, se superó la oferta de viviendas e inquilinato a precio accesible lo que ocasionó asentamientos informales en zonas periféricas de las ciudades.

Los asentamientos informales se pueden definir como zonas sin planificación urbana, precariedad, con desorden morfológico, con viviendas que no cumplen con las normas de planificación y construcción vigente y sin títulos de propiedad [2]. Son una expresión de la pobreza y desigualdad de la región que se ve reflejado en su ubicación. Principalmente están localizadas en zonas de poco valor del suelo [3] debido a encontrarse en zonas sin acceso a servicios básicos y alto peligro de desastres como tsunamis y sismos en la zona costera del Perú.

Con referencia a este último aspecto es importante destacar que la presencia de estas poblaciones vulnerables en estas zonas incrementa el riesgo, pues según la filosofía de diseño Sismorresistente, el riesgo se puede entender como el producto del peligro y vulnerabilidad [4]. Como señala Tavera, los pasados escenarios sísmicos se repetirán, siendo actualmente más crítico debido al incremento de la población y al acelerado desarrollo desordenado de las ciudades que ocasiona el incremento de la vulnerabilidad y por lo tanto del riesgo de pérdidas humanas y materiales ante un eventual evento sísmico [5]. Por tal motivo es importante identificar a las invasiones de manera temprana para que

las respectivas autoridades locales tomen medidas pertinentes como la reubicación o mejoramiento de las condiciones del sitio invadido, antes que se asienten de forma definitiva y construyan viviendas no ingenieriles que pongan en riesgo sus vidas ante la ocurrencia de un desastre.

El principal desafío para la detección de asentamientos informales es la escasez de datos estadísticos y censales [6]. No obstante, existen datos satelitales de media resolución, disponibles gratuitamente, que permiten obtener indicadores consistentes, continuos y estandarizados para cuantificar cambios de la superficie terrestre [7]. Por lo tanto, el papel de la teledetección, reconocido en el artículo 76 de la agenda 2030 [8], es fundamental para el monitoreo de estos asentamientos con el fin de mejorar la calidad de vida de los habitantes y encaminar los esfuerzos globales para los objetivos de desarrollo sostenible (ODS) propuestos por la ONU para el 2030.

La teledetección aplicada en observaciones de la tierra permite monitorear la superficie terrestre y por lo tanto detectar cambios [9]. Con la disponibilidad de nuevos satélites y estaciones aéreas en los últimos 50 años, ha sido posible el desarrollo de múltiples aplicaciones en distintos campos de las ciencias, como lo demuestra el incremento exponencial de investigaciones relacionadas con la teledetección en la Figura 1. Algunas de estas aplicaciones son: obtención de modelos de elevación digital [10], clasificación de usos de suelos [11], monitoreo de la salud vegetal [12], ubicación de barcos y derrames de petróleo [13], evaluación de daños ocurridos durante un desastre [14] entre otros. En cuanto a detección de invasiones, la mayoría de los estudios llevados en esta última década han empleado imágenes ópticas comerciales de alta resolución para su identificación [15].



Figura 1: Número de artículos por año relacionados con investigaciones en teledetección.
Nota: La gráfica fue elaborada a partir del registro de publicaciones realizadas en Scopus empleado como palabra clave Teledetección [16]

1.2 Objetivos

Objetivo general

Proponer una metodología para la detección temprana de invasiones en zonas de alto riesgo de desastres empleando técnicas de *machine learning* e imágenes de radar de media resolución de libre acceso.

Objetivos específicos

- Procesar y filtrar las imágenes de radar con la finalidad de emplearlas en la calibración de los algoritmos de *machine learning*
- Desarrollar la arquitectura de un programa para la detección de invasiones en zonas de alto riesgo de desastres
- Evaluar de forma experimental los resultados obtenidos empleando el caso de estudio de la invasión en Lomo de Corvina

1.3 Pregunta de investigación

La falta de capacidad operativa para la aplicación de políticas de gestión de riesgo [17] y problemas socioeconómicos [18] ocasionan que grupos de personas vulnerables se asienten en territorios de alto riesgo de desastres. Por tal motivo, es importante contar con una herramienta de monitoreo e identificación de invasiones en sectores peligrosos para evitar futuras pérdidas materiales y humanas. Por lo tanto, se formula la siguiente pregunta: ¿será posible identificar invasiones en un sector mediante un análisis temporal de imágenes de radar empleando técnicas de *machine learning*?

1.4 Problemática y justificación

El estudio del crecimiento urbano informal es un tema de investigación de urgencia nacional debido a que involucra a una cantidad importante de personas y la cifra se encuentra en crecimiento. De acuerdo con Muñoz y Rodríguez [19], en zonas de alto riesgo de desastres, como las laderas de los cerros, viven 2.8 millones de personas. Estiman que, para el 2025, habría 225 000 viviendas adicionales en estos sectores.

En el periodo del 2003 al 2021, las emergencias en el Perú han dejado un saldo de 3056 personas fallecidas, alrededor de 13 272 personas lesionadas, 389 personas desaparecidas, 2 063 043 damnificados, 282 092 viviendas destruidas y han ocasionado pérdidas económicas de más de 4 mil millones dólares [20]. Según INDECI [21] las emergencias que han ocasionado mayor número de damnificados han sido por las inundaciones, lluvias torrenciales y terremotos; destacando el terremoto de Pisco del 2007, considerado como megadesastre de nivel 5, con 481 701 damnificados y más de 96 mil viviendas destruidas.

INDECI prevé un escenario sísmico de 8.8 Mw en Lima, sumado a las precarias viviendas ubicadas en zonas de alto riesgo indicadas en el ANEXO 1, implicaría grandes daños materiales y pérdidas humanas. Según el estudio realizado por INDECI en el 2017 [22] se estimó el fallecimiento de más de 100 000 víctimas en Lima Metropolitana y más de 10 millones de afectados, según la Tabla 1, en este futuro escenario sísmico. No obstante, debido al incremento de la vulnerabilidad, por el asentamiento de la población en zonas expuesta a peligros, el riesgo de daños está en constante incremento.

Se puede apreciar en la Tabla 1 que las poblaciones más expuestas ante la ocurrencia de desastres son las que se ubican en zonas costeras. Por ejemplo, en el caso de Lima, que es la ciudad más poblada del Perú y ubicada en una zona de alta actividad sísmica, ante la ocurrencia de un sismo presentaría más de la mitad de la población afectada en el país. Esto se debe principalmente a la directa correlación entre las clases sociales menos favorecidas, viviendas no ingenieriles y ubicación en zonas de mayor peligro de desastres en la ciudad como refiere Matsuoka en su estudio del 2014 [23].

Tabla 1: Detalles de la población potencialmente expuesta a diferentes peligros.

Población potencialmente expuesta ante la ocurrencia de desastres											
Departamento	Geodinámica Interna			Volcanes	Geodinámica externa		Hidrometeorológicos/Oceanográficos				Antropológicos
	Sismo por subducción	Sismo por def. cortical	Tsunami		Movimientos de masas	Peligro de origen glaciar	Inundaciones	Heladas	Friaje	Sequía	Incendios forestales
Amazonas		280724			290333		219160	251092	127933	422629	80217
Ancash	1083519	298478	213098		321325	271368	736583	585192		611723	503284
Apurímac	623	263726			275258	201256	116400	405759		458830	390895
Arequipa	1344632	1000308	74748	1229318	1276795	149288	1288885	1156070		89906	433643
Ayacucho	72824	302005		12980	195995	346955	342532	553012	53157	671258	205228
Cajamarca	47562	341647			651938		580124	1241232	88676	1507185	1001465
Cusco		1052658		115664	886854	1012895	691844	1056620	148442	1234565	1044565
Huancavelica	134346	193424			209194	185303	72820	347334	305	491940	222579
Huánuco	495489	200055			308930	64722	289641	530857	189964	860537	259191
Ica	850764	499491	98965		272436		795752	56635		331	253206
Junín	872543	811259			231607	112714	1037663	889454	356569	1292366	201562
La Libertad	1691942	114275	174415		578572	35767	1412365	423362		285656	615727
Lambayeque	519		90271		31887		1162009	32304		106044	301184
Lima	10471812	680097	2456463		4494093	45567	9792145	430921		134400	484627
Loreto		103818			37976		882482		883509	1024732	3730
Madre de Dios		112044			9760		140895		141070		926
Moquegua	170230	151388	70928	25658	103516		50259	100210		28424	22500
Pasco	167544	140469			137036	153098	158275	165975	88090	275489	78581
Piura	1749893	328121	229883		331596		1627165	238466		1058741	681009
Puno	11	448143		14074	565619	140214	962585	1132335	30712	1402081	479468
San Martín		593341		17930	235911		692672	11992	800182	840790	8829
Tacna	328310	322241	2320		58881		314194	22040		18818	13967
Tumbes	224638	36297	36863		52351		216178			237685	49790
Ucayali		271494			28317		491863		496459	80628	1595
Totales	19707201	8545503	3447954	1415624	11586180	2719147	24074498	9630862	3415068	13134708	7337768

Nota: Adaptado de INDECI-DIPRE [22]

A partir de lo expuesto, es posible entender la gravedad del problema. Por ello, se busca detectar de forma temprana las invasiones en zonas de alto riesgo para que estén a disposición del Gobierno y se puedan tomar medidas pertinentes para cumplir con los objetivos planteados en el plan de prevención de riesgo y reducción del desastre de Lima Metropolitana [24]. Algunos de estos objetivos son reducir las condiciones de riesgo existentes y evitar generar riesgos de desastres en la población.

Para cumplir con tales objetivos se pueden emplear las imágenes de RADAR, las cuales son de libre acceso en distintas plataformas como ASF Data Search Vertex [25] y Copernicus [26]. Estas plataformas brindan una diversidad de productos en un amplio rango de frecuencias con media-baja resolución. Brindan productos de forma periódica y en tiempo real de la superficie terrestre. En el caso de Copernicus, brinda más de 8 petabits de información anual de las constelaciones Sentinel, los cuales están montados con sensores activos y ópticos de hasta una resolución espacial de 10 m.

Uno de los productos brindados por Copernicus es el GRD (*Ground Resolution Detection*). Está formado por un conjunto periódico de imágenes SAR que poseen información de la intensidad de la reflectancia de las ondas electromagnéticas, la cual es conocida como *backscatter* o retrodispersión. Con la información de la retrodispersión, obtenido de las imágenes de radar, se pueden calibrar algoritmos que detecten cambios anómalos repentinos en una zona no urbana y de alto peligro de desastre, lo cual sería indicador de una invasión.



Capítulo 2

Marco conceptual

2.1 Evaluación del peligro sísmico en el Perú

El mapa de sismicidad mundial revela que más del 80% de la energía acumulada en su interior se libera en la zona conocida como: “cinturón de fuego del pacífico”. Por tal motivo, el Perú está ubicado en una zona de alto peligro sísmico además de peligro a tsunamis, licuación y deslizamiento de tierra y rocas. La zona costera es la de mayor peligro sísmico por encontrarse más cerca de la superficie de fricción interplacal (Placas de Nazca y Sudamericana) que darían origen a terremotos con magnitud mayor a ocho grados acompañados de tsunamis [5].

Según Tavera (2014), si un sismo se repite en determinada zona, el suelo tendrá un comportamiento similar. Por lo tanto, los niveles de sacudimiento del suelo que se presentaría en un escenario futuro se pueden estimar en base a mapas de intensidades máximas. Estos niveles de sacudimiento están influenciados por la composición geológica del suelo. Los suelos menos compactos son los más propensos a amplificar las ondas sísmicas y por ende ocasionar mayor daño estructural y afectar la integridad de las personas.

La distribución espacial de las áreas de ruptura permite identificar la presencia de lagunas sísmicas, las cuales son definidas como zonas de ocurrencia de grandes sismos que no se repiten, incrementando su posibilidad de ocurrencia. Se puede identificar a la zona central del país como una laguna sísmica pues desde el terremoto de 1746, de intensidad XI en la escala de Mercalli, no se ha presentado otro de magnitud similar (8.8 Mw), por lo que esta región no ha liberado el total de su energía acumulada incrementando la posibilidad de un futuro evento de magnitud similar. En la Tabla 3 se

realiza una breve presentación de algunos de los principales eventos sísmicos ocurridos desde 1970 hasta la actualidad.

Tabla 2: Principales eventos sísmicos en Perú desde 1970

Año	Descripción	Magnitud
1970	Terremoto y Alud-aluvión en Ancash	7.9 Mw
1974	Terremoto de Lima	8.4 Mw
1986	Terremoto de Cusco	5.8 MI
1990	Terremoto en el nororiente peruano	5.8 MI
1991	Terremoto en el nororiente peruano	6.9 Mw
1996	Terremoto en Nazca	7.7 Mw
1999	Terremoto en Chuschi, Ayacucho	4.5 Mw
2001	Terremoto en el sur del país	8.4 Mw
2003	Sismo en Capacmarca, Cusco	4.5 MI
2005	Sismo en Moquegua	5.4 MI
2005	Terremoto de Lamas	7.5 Mw
2005	Sismo de Ayacucho	5.8 MI
2007	Terremoto de Pisco	8.0 Mw
2008	Sismo de Pasco	5.3 Mw
2013	Terremoto de Acarí	7.0 Mw
2018	Terremoto de Arequipa	7.1 Mw
2019	Sismo de Lagunas	8.0 Mw
2021	Sismo en Loreto	7.5 Mw
2021	Sismo en Piura	6.2 Mw

Nota: Adaptado de INDECI [20]

Según INDECI, un terremoto de 8.8 Mw, similar a lo ocurrido en 1746, podría generar un tsunami con olas de 24 metros de altura, licuación de suelos, derrumbes, volcamiento de suelo y roca, 623 882 viviendas inhabitables y 110313 fallecidos [22]. Las poblaciones más afectadas serán las que son más vulnerables y se encuentren en zonas de mayor peligro de desastre. En las Figura 2 y Figura 3 se muestran algunos factores de vulnerabilidad y peligro en Lima Metropolitana [22] y en el anexo 1 presenta algunos mapas de peligro y riesgo ante un terremoto de magnitud 8.8 Mw.

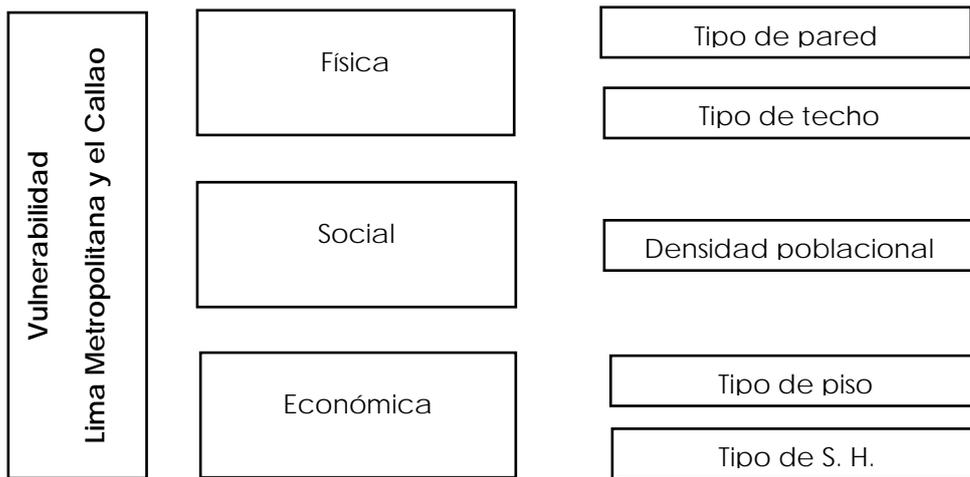


Figura 2: Factores de vulnerabilidad

Nota: Adaptado de INDECI-DIPRE [22]

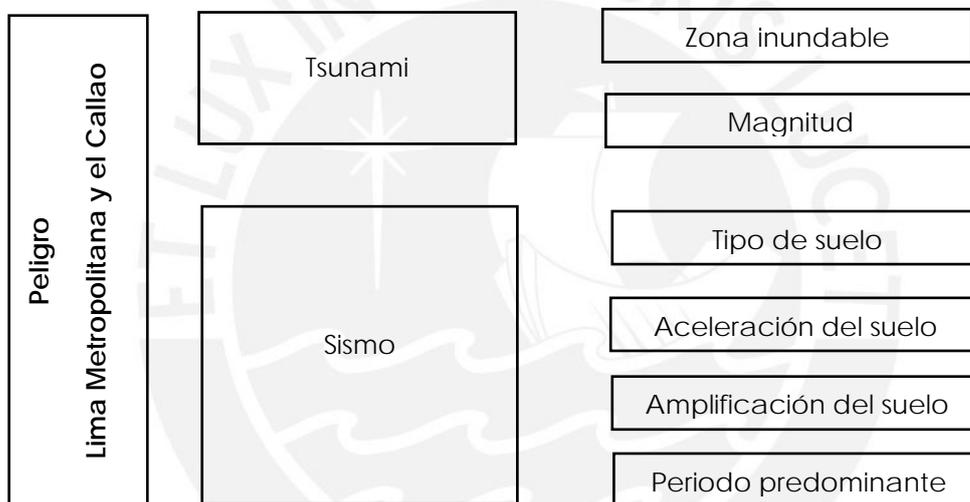


Figura 3: Factores de peligro

Nota: Adaptado de INDECI-DIPRE [22]

2.2 Políticas de gestión de riesgos en el Perú

El Sistema Nacional de Gestión del Riesgo de Desastres (SINAGERD) define a la Política Nacional de Gestión del Riesgo de Desastres como el grupo de medidas destinadas a reducir el riesgo, evitar la generación de nuevos riesgos, efectuar una preparación, atención, rehabilitación y reconstrucción ante desastres, así como minimizar los efectos adversos sobre las población, economía y ambiente.

En ese sentido, la Política Nacional de Gestión del Riesgo de Desastres al 2050 [27] plantea los siguientes 6 objetivos prioritarios los cuales son:

- 1) *Mejorar la comprensión del riesgo para la mejor toma de decisiones a nivel de la población y entidades del estado*
- 2) *Mejorar las condiciones de ocupación y uso considerando el riesgo de desastres en el territorio*
- 3) *Mejorar la implementación articulada de la gestión del riesgo de desastres en el territorio*
- 4) *Fortalecer la incorporación de la gestión de riesgo de desastres en la inversión pública y privada*
- 5) *Asegurar la atención de la población ante la ocurrencia de emergencias y desastres*
- 6) *Mejorar la recuperación de la población y sus medios de vida afectados por emergencias y desastres*

El SINAGERD señala que las principales causas de la alta vulnerabilidad de la población y sus medios de vida se deben a:

- a) Ocupación y uso inadecuado del territorio sin considerar sus condiciones de riesgo (exclusión social, pobreza, incremento acelerado de la población urbana y ausencia de una adecuada planificación que tome en cuenta la gestión del riesgo del desastre para la gestión del territorio).
- b) Falta de comprensión del riesgo del desastre en todas sus dimensiones
- c) Falta de implementación y planificación de estrategias integrales de gestión de riesgos de desastres bajo un enfoque sistemático.
- d) Deficiente inclusión del análisis del riesgo y medidas de prevención y reducción del riesgo en las inversiones públicas y privadas debido a un deficiente conocimiento del costo beneficio de incorporar la gestión del riesgo en las inversiones.
- e) Falta de eficacia y oportunidad para la respuesta oportuna y recuperación.

2.3 Crecimiento urbano en Lima Metropolitana

El crecimiento urbano acelerado en la etapa contemporánea de Lima metropolitana se puede clasificar en cuatro etapas diferenciadas según Torres, Perleche y Aiquipa [28]. Se presenta algunas de las características de las etapas:

Inicio de la expansión urbana de Lima y el Callao 1900-1960

Se inicia en 1868 con la demolición de las murallas de la ciudad. Los primeros espacios autoproducidos surgen en 1920 en zonas circundantes al centro de la ciudad y puerto del Callao. Según [29] los primeros espacios urbanos autoproducidos ocurrieron en Armatambo (1924) a causa del hacinamiento de los ranchos de Villa de Chorrillos; Puerto Nuevo (1928), crecimiento industrial y laboral en el puerto del Callao; Mendocita (1931), venta y/o alquiler de parcelas de fundo y Leticia (1933) por la inundación de Cantagallo.

Barriadas rodean la ciudad (1940-53)

La autoproducción como problema estructural se intensificó en 1940 luego del terremoto del 24 de mayo. Se crearon 3 espacios urbanos importantes: San Cosme (1946), El Agustino (1947) y 27 de noviembre en 1949. El fenómeno de urbanización popular fue denominado Barriada y reconocida por parte del presidente Manuel Odría. Este reconocimiento permitió que se crearan asociaciones con fines de Lucro iniciándose un mercado ilegal de lotes sobrantes

Barriadas rompen el cerco (1954-60)

Durante la década de los 50 las invasiones ocuparon terrenos más lejos del centro en: Arenales del Sur, Ciudad de Dios y Comas. Con la finalidad de reducir la presencia de urbanizadores piratas en 1961 se promulgo la ley 13517 ley de barrios urbanos marginales que reconocía la urbanización autoproducida. El estado debía encargarse de pagar al supuesto propietario por la tierra ocupada y continuar con el saneamiento físico legal. No obstante, debido a los altos costos de las habilitaciones urbanas el estado solo se encargó de entregar títulos de propiedad.

Producción del espacio urbano (1961-2020)

COFOPRI (Comisión de Formalización de la Propiedad Informal) brindo titularidad de vivienda en zonas no aptas para vivienda. Desde entonces las municipalidades asumen la responsabilidad de la asesoría técnica para el levantamiento del riesgo y luego COFOPRI procede con la formalización.

En el 2° gobierno de Alan García se promulgo la ley 28923 ley que estableció el régimen temporal extraordinario de formalización y titulación de predios urbanos mediante el cual COFOPRI se convirtió en el organismo de propiedad informal. Debido a este cambio se registró la mayor corrupción pues beneficio a inmobiliarias que compraron las tierras de las comunidades campesinas y traficantes de terrenos.

En 2007 el terremoto de Pisco dejo muchas pérdidas de vidas y graves daños de la infraestructura y la vivienda. Debido a esto se promulgo un decreto supremo para que se puedan formalizar posesiones informales con el fin de que puedan acceder a los bonos de techo propio. Sin embargo, Vera [30] indica que el 90% de las invasiones en Lima y la costa norte son producto del tráfico de terrenos. Los Traficantes crearon organizaciones sociales ficticias para el proceso de titulación ante COFOPRI con lo cual se beneficiaron de estos bonos.

2.4 Machine learning

Se refiere a los cambios en los sistemas que realizan tareas, como reconocimiento, planificación control, predicción, entre otros, con el fin de mejorar dichos sistemas. Las principales razones mencionadas por Nilsson [31] son: ayudan a entender el proceso de aprendizaje de personas y animales, permite encontrar relaciones y correlaciones ocultas entre una gran cantidad de información, permiten mejora los diseños de sistemas existentes y los hacen capaces de adaptarse a los cambios en el tiempo por lo que reduce la necesidad de su rediseño

Entre los tipos de machine learning se encuentran los tipos supervisados y semi-supervisados, algunos de los cuales se detallan a continuación.

Aprendizaje supervisado: Requiere de datos etiquetados, es decir que se tiene conocimiento de lo que es lo que el modelo debe aprender. Los modelos supervisados empleados en este proyecto fueron:

Decision Tree

Es un algoritmo empleado en problemas de clasificación. Consiste en una estructura similar a un diagrama de flujo condicional que asimila un conjunto de nodos, ramas y hojas igual que un árbol. Las ramas conectan nodos relacionados con una condición y la hoja representa el grupo o valor predicho para los parámetros de entrada. Se presenta una representación de esta estructura en la Figura 4.

La construcción de este algoritmo se basa en mediciones de impureza definidos principalmente por el índice de Gini y la entropía. La generación de árboles de decisión se basa en algoritmos como ID3, CART y CHAID principalmente.

Gaussian Naïve Bayes

Es un algoritmo probabilístico paramétrico basado en el teorema de Bayes. Por tal motivo asume independencia entre las características de entrada. El modelo puede ser fácilmente entrenado encontrando la media y desviación estándar de los datos de entrenamiento de cada clase.

Random Forest

Los bosques aleatorios son una composición de diferentes arboles de decisiones. Su principal ventaja es que reduce el sobreajuste al que se expone un solo árbol de decisión. Su principal desventaja es su alta dilación en la predicción de datos a partir del modelo entrenado.

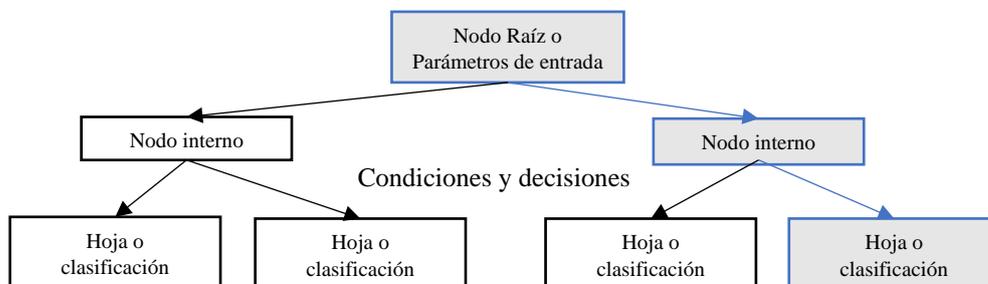


Figura 4: Esquema de un árbol de decisiones.

Support Vector Machine

Modelo de clasificación basada en la creación de hiperplanos para la separación de clases a partir del mayor margen resultante entre las clases de la data de entrenamiento.

Stochastic Gradient Descent

Es un método de optimización aleatorio empleado en la reducción del error cuadrático medio entre los valores reales y predichos. Consiste en emplear el algoritmo del descenso de gradiente en un grupo diferentes de datos de entrenamiento en cada iteración. Las principales ventajas son que reducen el coste computacional de cada iteración y reducen el sobreajuste del modelo.

K-Neighbors

Método que tiene por finalidad predecir o clasificar cierta instancia a partir de la distancia más cercana a una clase. Este algoritmo esta basado en instancias por lo que no genera un modelo explícitamente, memoriza toda la data de entrenamiento para predecir una instancia. Por lo tanto, tiene un alto coste computacional en caso se empleen muchos datos de entrenamiento.

Logistic Regression

Método multivariante que se basa en realizar una combinación lineal de las variables independientes y aplicarle una trasformación no lineal, como la aplicación de una función sigmoide, para la predicción de la variable dependiente.

Neuronal Network

Modelo inspirado en el comportamiento de las neuronas y el cerebro. Esta formado por un conjunto de capas de entrada, ocultas y de salida. Cada capa posé un conjunto de neuronas que reciben información de la capa anterior por medio de composición entre una combinación lineal de los valores en las neuronas en la capa anterior y una función no lineal de transformación. La calibración de los parámetros se basa en la regla de la cadena, aplicada en la derivación, y en la propagación hacia atrás, conocido como *back propagation*.

Aprendizaje semi-supervisado: Modelos empleados cuando no se disponen completamente los datos etiquetados. Los modelos semi-supervisados empleados en este proyecto se basaron en modelos de detección de valores atípicos los cuales fueron:

Elliptic Envelop

Método paramétrico basado en la asunción de distribución gaussiana de la data. Es un método basado en la detección de valores atípicos a partir de la matriz de covarianza

Local Outlier Factor

Método que mide la desviación local de la densidad de una muestra con respecto a sus vecinos más cercanos. Se denomina local debido a que la identificación de la anomalía depende de que tan alejado esta el objeto respecto a los objetos cercanos.

2.5 Detección de valores atípicos

Un valor atípico se considera a aquel valor que tienen un comportamiento distinto del resto de la data que se puede modelar de forma estándar. En la literatura del aprendizaje automático se le entiende como una anomalía, discordancia, desviación o anomalía. El poder estimar que comportamiento es lo suficientemente atípico para considerar a un grupo de datos como anomalías es subjetivo [32]. Principalmente dependerá de la cantidad de datos disponibles y el juicio de expertos.

Como se puede apreciar en la Figura 5, los valores atípicos se pueden clasificar como ruido y anomalías. Lo que se busca en las diversas aplicaciones de detección de cambios es atenuar en la medida de lo posible el gran contenido del ruido. El ruido se considera como una forma débil de los valores atípicos, conocido como *weak outliers*, presente en la data. Se debe tratar de atenuar su efecto para enfocarse en la detección de anomalías (*strong outliers*).

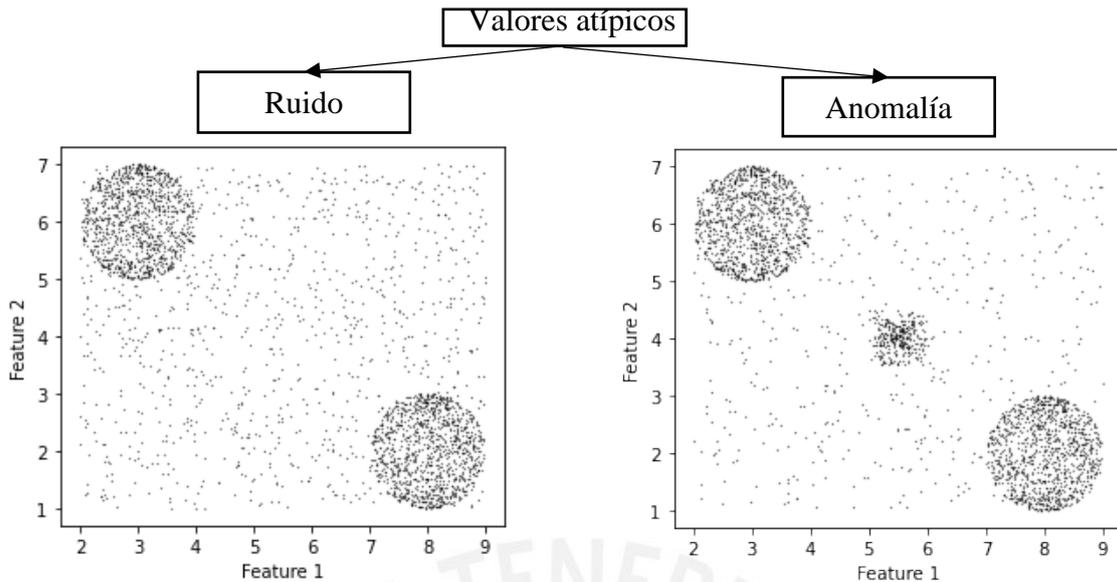


Figura 5: Tipos de valores atípicos. En la gráfica de la izquierda se aprecia 2 grupos claramente delimitados. En la gráfica de la izquierda se observa la aparición de un tercer pequeño grupo en la zona central del gráfico

El principal problema para la detección de cambios es que la mayoría de los datos disponibles en la realidad son de eventos recurrentes o convencionales, lo que supone la falta de valores atípicos para el aprendizaje de su comportamiento. A pesar que esta limitación convierte la detección de anomalías en una tarea más desafiante, existen enfoques semisupervisados y no supervisados (revisar la Tabla 4 en la que se presentan algunos de estos modelos con sus analogías supervisadas) que hacen uso de diversos algoritmos con los cuales se pretende modelar el comportamiento típico de la data y detectar a aquellos datos que se desvían de forma “significativa” de los patrones del modelo. Para poder evaluar estas desviaciones se adoptan distintas métricas en base a la proximidad de los valores reales y estimados.

Tabla 3: Modelos no supervisados con sus respectivos análogos supervisados

Modelos no supervisados	Modelos supervisados análogos
k-NN, LOF, LOCI	k-nearest neighbor
PCA	Linear Regression
Expectation-maximization	Naive Bayes
Mahalanobis method, Clustering	Rocchio
Isolation trees, Isolation Forest	Decision trees, Random Forest
FP-Outlier	Rule-based
One Class Support Vector Machine	Support Vector Machine
Replicator neuronal networks	Neuronal Networks
PCA Matrix Factorization	Matrix factorization

Una de las principales desventajas de los enfoques previamente mencionados es que la mayoría de sus métodos son de generalización explícita por lo que usan el mismo conjunto de datos para el entrenamiento y evaluación. Esto puede conllevar el sobreajuste de los modelos, además, se presentará errores en la detección debido a la suposición que todos los datos de entrenamiento pertenecen a la clase convencional de la data. Sin embargo, se puede reducir estos errores dividiendo de forma aleatoria a la data.

La elección del algoritmo o modelo es crucial pues de ello dependerá la precisión y exactitud de nuestros resultados. Esta elección depende del entendimiento de la naturaleza de los datos estudiados, con la finalidad de evitar un mal ajuste del modelo debido a una suposición errónea de la distribución de la data. Además, depende de la cantidad de datos disponible para el aprendizaje del algoritmo, que puede ocasionar problemas de *overfitting* y *underfitting* de los modelos seleccionados.

2.6 Matriz de confusión

La matriz de confusión es un esquema que muestra una relación de la cantidad de instancias clasificadas por cada clase. Si se considera un problema de clasificación con dos clases C_0 y C_1 , la matriz de confusión sería la que se ilustra en la Figura 6. Los valores en la diagonal representan el número de instancias correctamente clasificadas por cada clase, mientras que el resto de los valores representan instancias clasificadas erróneamente. Por ejemplo, el valor n_{01} indica la cantidad de instancias que eran de tipo C_1 pero que fueron clasificadas erróneamente como de tipo C_0 .

Matriz de Confusión		Resultados reales		Total
		C_0	C_1	
Resultados predichos	C_0	n_{00}	n_{01}	$nr_{clase\ 0}$
	C_1	n_{10}	n_{11}	$nr_{clase\ 1}$
Total		$nr_{clase\ 0}$	$nr_{clase\ 1}$	n

Figura 6: Matriz de confusión. El primer índice de los elementos de la matriz hace referencia a la clase predicha del elemento y el segundo índice a la clase real.

2.7 Métricas de evaluación de resultados

A partir de los resultados obtenidos de la matriz de confusión, se pueden emplear las siguientes métricas para evaluar el desempeño obtenido por distintos algoritmos.

Overall Accuracy: Métrica empleada para determinar la fracción de correctas predicciones respecto al total de predicciones. Analíticamente se define como:

$$Accuracy(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} f(y_i, \hat{y}_i) \Rightarrow f(y_i, \hat{y}_i) = \begin{cases} 1 & \text{sí } y_i = \hat{y}_i \\ 0 & \text{sí } y_i \neq \hat{y}_i \end{cases}$$

n es el total de predicciones

Precision: Métrica empleada para determinar la fracción de correctas predicciones respecto al total de predicciones de una clase en específica. Analíticamente se define como:

$$Precision(y, \hat{y}) = \frac{1}{n_{clase\ j}} \sum_{i=0}^{n_{clase\ j}-1} f(y_i, \hat{y}_i) \Rightarrow f(y_i, \hat{y}_i) = \begin{cases} 1 & \text{sí } y_i = \hat{y}_i \\ 0 & \text{sí } y_i \neq \hat{y}_i \end{cases}$$

n_{clase j} es el total de predicciones de la clase j

Recall: Métrica empleada para determinar la fracción de correctas predicciones respecto al total de elementos de una clase en específica. Analíticamente se define como:

$$Recall(y, \hat{y}) = \frac{1}{n_{clase\ j}} \sum_{i=0}^{n_{clase\ j}-1} f(y_i, \hat{y}_i) \Rightarrow f(y_i, \hat{y}_i) = \begin{cases} 1 & \text{sí } y_i = \hat{y}_i \\ 0 & \text{sí } y_i \neq \hat{y}_i \end{cases}$$

n_{clase j} es el total de elementos de la clase

F1: Es conocida como una puntuación equilibrada debido a que consiste en la media armónica de los valores de *precision* y *recall*. Analíticamente se define como:

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

Kappa: Índice que indica la similaridad de distintos resultados o predicciones.

2.8 TELEDETECCIÓN (REMOTE SENSING)

La teledetección se puede definir según Bakker et. al. [33] y Khorram et. al. [34] como la adquisición de información geoespacial a distancia empleando sensores (satelitales o aerotransportados) que captan señales electromagnéticas emanadas por parte de la superficie terrestre a determinado tamaño de pixel o resolución. Su finalidad es a partir de mediciones indirectas cuantitativas [35] detectar y caracterizar objetos a distancia [36]. Posee diversas aplicaciones en ingeniería, arquitectura, arqueología, medicina, control de calidad industrial, robótica, mapeo extraterrestre, etc. En la Figura 7 se presenta una composición de colores basado en dos imágenes de radar tomadas en distintas fechas, donde las tonalidades en color cian denotan la detección de algunas de las áreas inundadas en Puerto Maldonado en el 2021.

Existen dos tipos de sensores de acuerdo a la fuente de energía empleada los cuales son:

Sensores pasivos: Las ondas electromagnéticas que perciben proceden de una fuente de energía natural como el sol. Por tanto, su principal desventaja es que depende de la intensidad percibida de la fuente emisora.

Sensores activos: Denominados sistemas radar, emiten y reciben su propia energía.

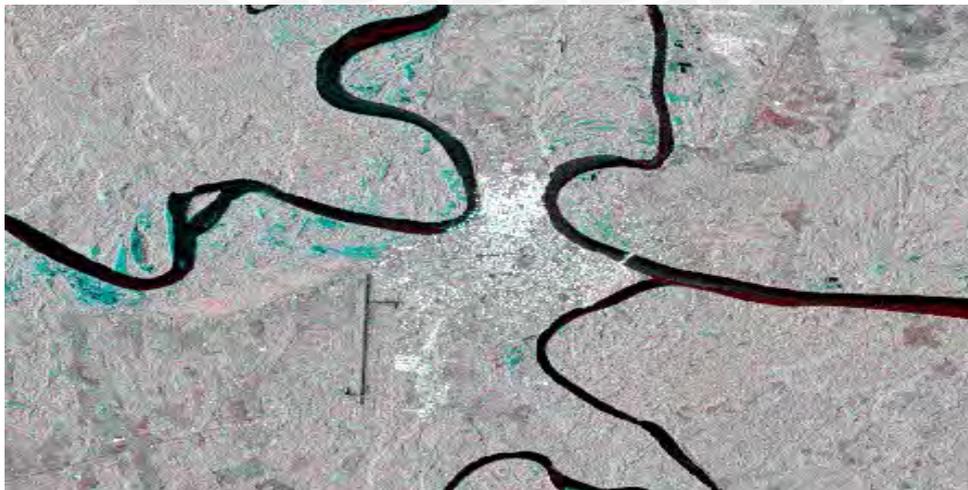


Figura 7: Detección de inundaciones en Puerto Maldonado. Los pixeles en color cian representan las áreas inundadas en Puerto Maldonado en el 2021

Teledetección para observaciones terrestres

Permite recopilar información las propiedades físicas, químicas, biológicas y geométricas de nuestro planeta; Nos ayuda a evaluar el estado y monitorear los cambios

de la naturaleza y antropológicos. Por ejemplo, permite monitorear el crecimiento urbano [37] [38] [39] [40], salud de los cultivos, determinar regiones de mayor peligro sísmico [41], edificaciones colapsadas [42] entre otros.

Los sensores empleados poseen 2 tipos de orbitas las cuales son:

Geoestacionarias: Se mueven con la misma velocidad de rotación y traslación terrestre para una toma continua de datos de una zona en particular del planeta.

Cuasi-Polares: Se mueven de norte a sur. Permiten tomar datos de toda la superficie terrestre periódicamente debido a la rotación terrestre. Algunos de las constelaciones que describen estas orbitas son: Sentinel 1-6, Radarsat, entre otros. En el caso de los sensores pasivos se debe garantizar que las condiciones de luz sean similares para que los niveles de radiancia captado en un lugar en distintos tiempos sean comparables.

2.9 Imágenes de radar y SAR

El *radar* (Radio Detection and Ranging) es un sistema de sensores activos que operan en un mayor rango de longitudes de onda (1 cm-1 m) que los sistemas visibles y sistemas ópticos multiespectrales como se muestra en la Figura 8. La magnitud que miden estos sensores es la radiancia. La radiancia se define como la cantidad de energía electromagnética medida desde un área en una dirección específica.

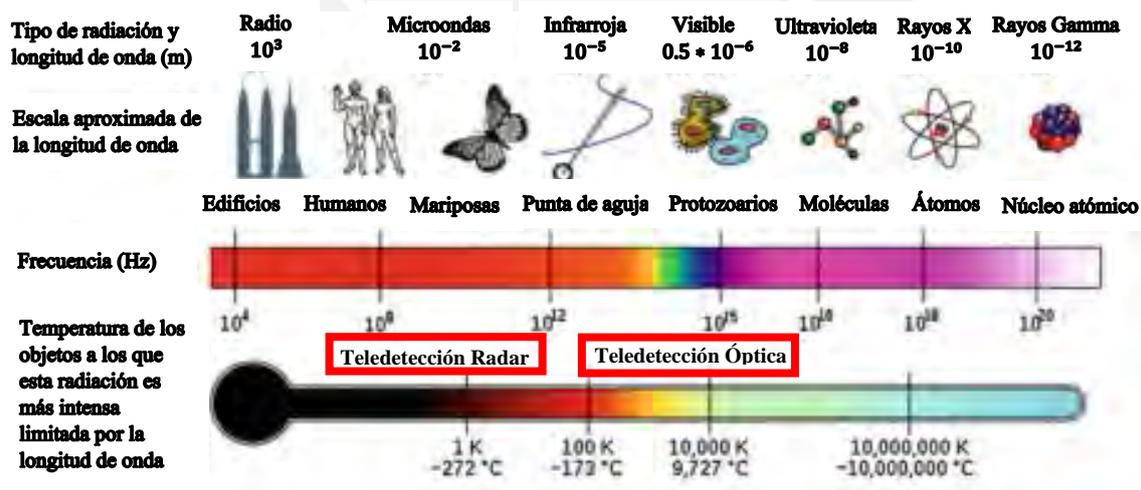


Figura 8: Espectro electromagnético. Nota: Adaptado de Khorram et. al. [43]

El sensor tiene 3 funciones primarias: transmisión de señales microondas, recepción de la energía retrodispersada y medición del tiempo de demora de retorno de la señal. El

proceso para la formación de imágenes se presenta en la Figura 9. Las imágenes están compuestas por píxeles y su nivel de brillo se cuantiza de forma discreta.

El nivel de brillo se caracteriza a partir del coeficiente de retrodispesión medido por píxel. Este coeficiente mide la capacidad reflectiva de los objetos, depende de las propiedades geométricas y dieléctricas. La medida de retrodispersión se encuentra en un rango promedio de -25db a 1 db como se evidencia en la Tabla 5.

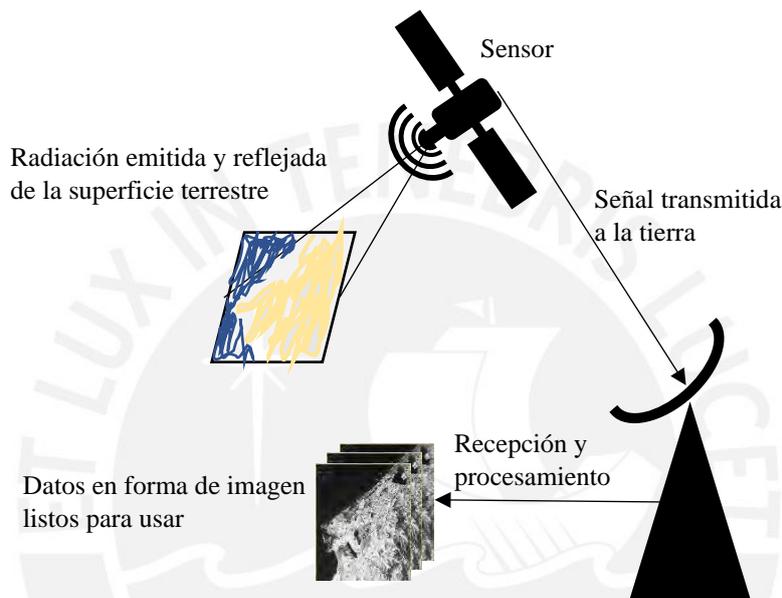


Figura 9: Flujo de trabajo para la obtención de imágenes digitales

Nota: Adaptado de Richards [44]

Tabla 4: Distribución de valores típicos de retrodispersión.

Valores de retrodispersión	Superficies típicas
Mayores a -5db	Objetos antropogénicos (ambientes urbanos) Pendientes orientadas hacia el sensor
Entre -10db y 0db	Superficies muy rugosas Superficies rugosas Vegetación densa (bosques)
Entre -20db y -10db	Niveles medios de vegetación Cultivos Superficies moderadamente rugosas
Menores a -20db	Superficies lisas Aguas en calma Carreteras Suelos muy secos (arenas)

Nota: Adaptado de IGN [36]

La medida de retrodispersión esta fuertemente influenciada por la rugosidad de la superficie determinada como se aprecia en la Figura 10. Superficies lisas reflejan poca o ninguna energía emitida de vuelta al sensor por lo que se observan en las imágenes de Radar con tonalidades oscuras mientras que superficies más rugosas se observan más brillantes.

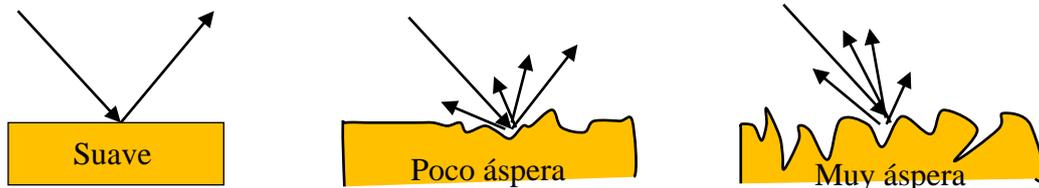


Figura 10: Mecanismo de reflexión en distintas superficies.

Nota: Adaptado de Richards [45]

El sistema SAR (Synthetic Aperture Radar) es un sistema capaz de emitir su propia señal microondas a la tierra y procesarla. Trabaja en el rango de longitudes de onda de los sensores activos, lo que garantiza la observación óptima de cada fenómeno dado que se encuentra fuera del rango de la ventana atmosférica (Figura 11), es decir que se encuentra en las zonas de menor absorción y dispersión de la atmosfera. Esto se evidencia con la tendencia de incremento de la atenuación para mayores frecuencias como se muestra en la Figura 12.

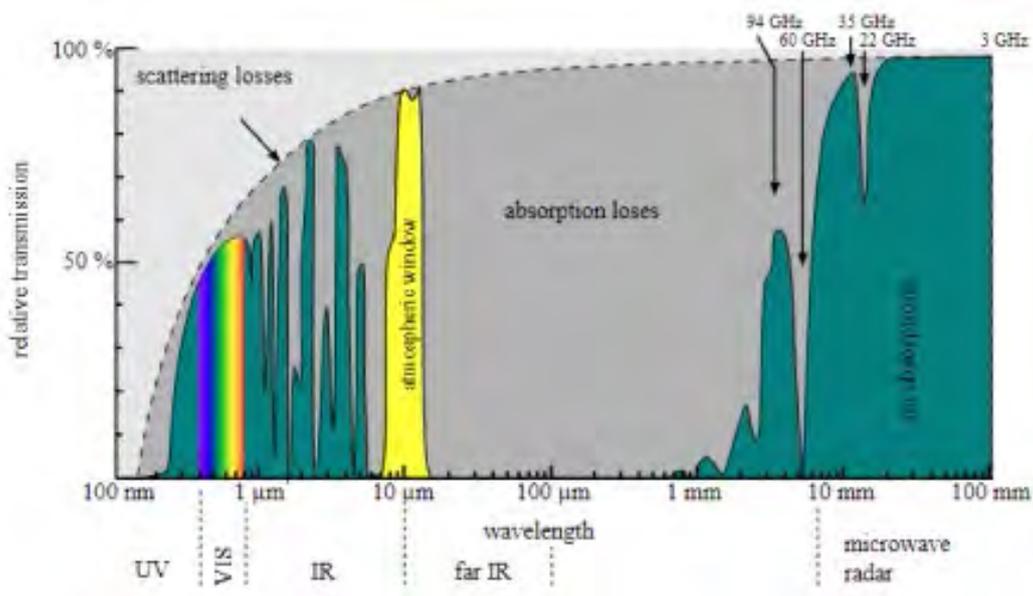


Figura 11: Ventana atmosférica.

Nota: Tomado de IGN [36]

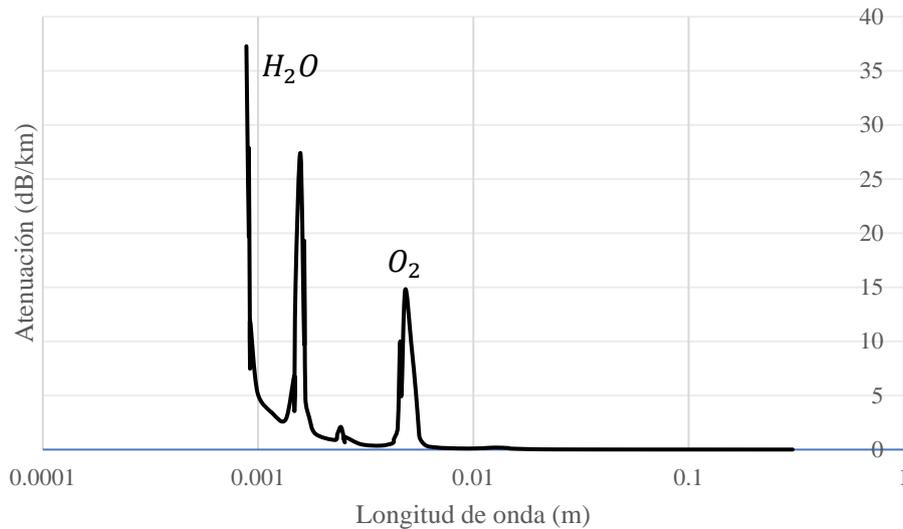


Figura 12: Atenuación de la radiación espectral electromagnética en la atmósfera.

Nota: Adaptado de Richards [45]

El sistema SAR mediante el efecto Doppler y creación de una antena virtual permite formar imágenes, con anchos de barrido de hasta 400km, mediante información en los píxeles. Los píxeles indican el coeficiente de retrodispersión dentro de esa área del terreno. El valor almacenado en el pixel será mayor mientras mayor sea la señal de retorno. Los parámetros que afectan a la señal retrodispersada son:

Polarización: Plano de propagación del campo eléctrico de las señales de propagación y retorno. Existente las siguientes posibilidades:

- HH: transmisión horizontal y recepción horizontal
- HV: transmisión horizontal y recepción vertical
- VH: transmisión vertical y recepción horizontal
- VV: transmisión vertical y recepción vertical

La polarización idéntica (HH y VV) es originado por la presencia de una superficie de reflejo casi especular. Por otro lado, la polarización cruzada (HV Y VH) está asociada a diferentes reflexiones debido principalmente a la rugosidad del terreno. En la Figura 13 se presenta una representación de la polarización idéntica.

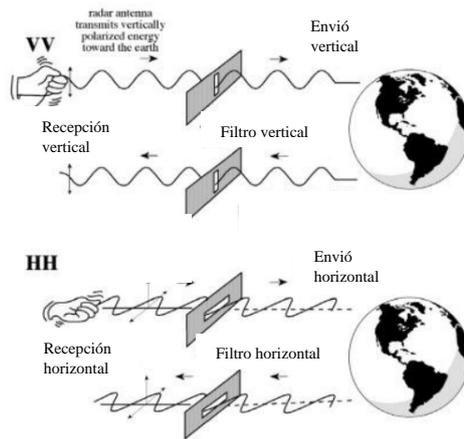


Figura 13: Polarización idéntica.

Nota: Adaptado de IGN [36]

Longitud de onda: Mientras sea mayor tendrá mayor penetración en el medio. En la Figura 14 se muestra la penetración de la longitud de onda en distintos medios. Se aprecia que el comportamiento de penetración es mayor a una mayor longitud de onda para cualquiera de esos medios.

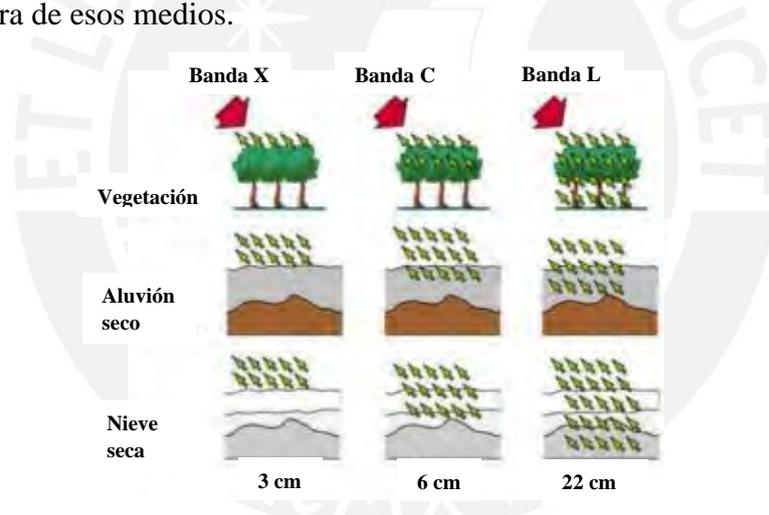


Figura 14: Penetración de diferentes longitudes de onda en diferentes cubiertas.

Nota: Adaptado de ESA [46]

Angulo de incidencia: Provoca distorsiones en la resolución del azimut y alcance debido a las observaciones laterales.

Mecanismos de retrodispersión: Son producto de la interacción de las ondas microondas con la superficie de estudio. El mecanismo a destacar es el efecto de esquina reflectora, el cual ocasiona una amplificación de la señal retornada al radar dada la presencia de superficies espejo como las paredes de los edificios.

En la geometría de adquisición de las imágenes tiene una gran influencia el sentido del satélite que captura la imagen en su órbita (ascendente o descendente) y la configuración del relieve debido a que se producen distorsiones geométricas que desfiguran las formas debido a los ecos que llegan a la antena de forma inversa a su posición respecto a su trayectoria.

Las principales desventajas de los sistemas SAR son: la interpretación de las imágenes es más complicada que las obtenidas por sensores ópticos, presencia de ruido por moteado y la forma de adquisición de la retrodispersión. Con respecto a este último aspecto se producen una serie de efecto geométricos que resultan en un desplazamiento por el relieve. Estos efectos son los siguientes:

Shadowing: Como se aprecia en la Figura 15 con la letra S, consiste en zonas no iluminadas (4-5) por el sensor por lo que no se registran valores de retrodispersión.

Layover: Como se aprecia en la Figura 15 con la letra L, ocurre cuando la energía reflejada en la zona superior de un elemento (4) se percibe antes que la que proviene de la zona inferior (3). Por lo tanto, la parte superior del elemento se verá desplazada y superpuesta por encima de la base en la imagen.

Foreshortening: Como se aprecia en la Figura 15 con la letra F, consiste en la aparente compresión de los elementos que están orientados hacia el sensor y alargamiento si están orientados de forma perpendicular. Implica que dependiendo de la pendiente de un elemento se puede obtener una apariencia más brillante.



Figura 15: Efectos geométricos en una imagen SAR. La letra “S” denota el *Shadowing*, la letra “L” denota el *Layover* y la letra “F” denota el *Foreshortening*

2.10 Proyecto Sentinel

Proyecto multisatelital desarrollado por la Agencia Espacial Europea (ESA por sus siglas en ingles), en el marco del programa Copernicus. Su finalidad es monitorizar cualquier cambio en la superficie terrestre, océanos, atmosfera y gestionar desastres empleando instrumentos radares o multispectrales. Las misiones de las constelaciones Sentinel son:

Los Radares gemelos Sentinel 1-A Y 1-B son empleados para respuesta ante emergencias. Están desfazados 180° por los que el periodo de monitoreo de toda la superficie es de 6 días. Trabaja en la banda C por lo que las ondas electromagnéticas emitidas tienen longitudes de onda alrededor de 4cm como se aprecia en la Tabla 6 , por lo que son capaces de atravesar las nubes. Algunas de las características de los productos que ofrecen son mostrados en la Tabla 7. Actualmente, solo el satelite Sentinel-1A se encuentra operativo.

Tabla 5: Ancho de banda, longitud de onda y frecuencia

Banda	Longitud de onda (cm)	Frecuencia (GHz)
Ka	0.75-1.20	40-25
K	1.20-1.67	25-18
Ku	1.70-2.50	17.6-12.0
X	2.50-4.00	12-7.5
C	4.00-8.00	7.50-3.75
S	8.00-15.0	3.75-2.00
L	15.0-30.0	2-1
P	60.0-120	0.50-0.25

Tabla 6: Productos disponibles ofrecidos por Sentinel 1

Producto	Descripción
Level-0 SAR	Datos en bruto
Level 1 GRD	Datos con información de intensidad. La distribución es luego de 1 hora en caso de emergencias
Level 1 SLC	Comprende imágenes complejas con amplitud y fase
Level 2 OCN	Diseñado para recabar parámetros geofísicos del océano

La constelación Sentinel 2 es un sistema de teledetección óptica que permite el monitoreo de la superficie terrestre con 13 bandas espectrales con resolución de hasta 10m. Son empleados para el control de desastres naturales. Algunas de las características de los productos que ofrecen son mostrados en la Tabla 8.

Tabla 7: Productos disponibles ofrecidos por Sentinel 2

Producto	Descripción
Level 0	Datos comprimidos en bruto
Level 1A	Datos en bruto sin comprimir
Level 1B	Imágenes en valores de radiancia corregidos radiométricamente
Level 1C	Ortoimágenes en proyección UTM
Level 2A	Evolución del nivel 1C en el que se aplican correcciones atmosféricas para proporcionar las ortoimágenes con niveles digitales expresados en valores de reflectancia reales .

Las constelaciones Sentinel 3 y 6 son empleados para el monitoreo del agua y océanos. Sus principales objetivos son: medición de la temperatura de los océanos y la superficie terrestre, calcular la altura significativa de las olas, monitorear la superficie glaciar, monitorear la calidad del agua y pronosticar el clima marino.

Las constelaciones Sentinel 4 y 5 son empleados para el monitoreo de la atmósfera. Su principal función es medir la calidad del aire en base a la medición indirecta de la concentración de gases como el dióxido de carbono en la troposfera.

2.11 Operadores morfológicos

Se pueden definir como operadores de ventana o convolución. Son operaciones que modifican un pixel en función de un grupo de pixeles próximos o vecinos al pixel a transformar. Se puede definir como:

$$r(i, j) = T_{m,n} \left\{ \begin{array}{l} \phi_{a \in [i-\frac{m}{2}; i+\frac{m}{2}]}(a, b) \\ b \in [j-\frac{n}{2}; j+\frac{n}{2}] \end{array} \right\} ; a \wedge b \in Z$$

La función $\phi(a, b)$ se define como el valor original de intensidad del pixel en la posición (a; b) de la imagen. La función de transformación $T_{m,n}$ representa a la función de transformación de intensidad que involucra a los m pixeles próximos a i en el primer eje principal y los n pixeles próximos en el segundo eje principal. Si la operación de transformación se puede representar como combinación lineal de los pixeles en el intervalo de la vecindad analizada se le denomina *kernel*, esta operación se puede definir de la siguiente forma:

$$r(i, j) = \sum_m \sum_n \phi(m, n) * t(m, n)$$

Algunas de los operadores morfológicos se pueden emplear cuando se está interesado en objetos específicos definidos por un grupo de pixeles conectados que mediante una forma representen un elemento como casa, lagos, campos, etc. Se presentan a continuación algunos operadores en particular:

Erosión: Operador empleado para reducir el tamaño de un objeto. Este operador se define como:

$$E = O \ominus S = \{i, j | S_{i,j} \subseteq O\}$$

Donde: **O** es el objeto y **S** es la vecindad de pixeles en un pixel en específico. En la Figura 16 se presenta un ejemplo de la aplicación de este operador.

Dilatación: Operador empleado para reducir el tamaño de un objeto. Este operador se define como:

$$D = O \oplus S = \{i, j | S_{i,j} \cap O \neq \emptyset\}$$

Donde: **O** es el objeto y **S** es la vecindad de pixeles en un pixel en específico. En la Figura 17 se presenta un ejemplo de la aplicación de este operador.

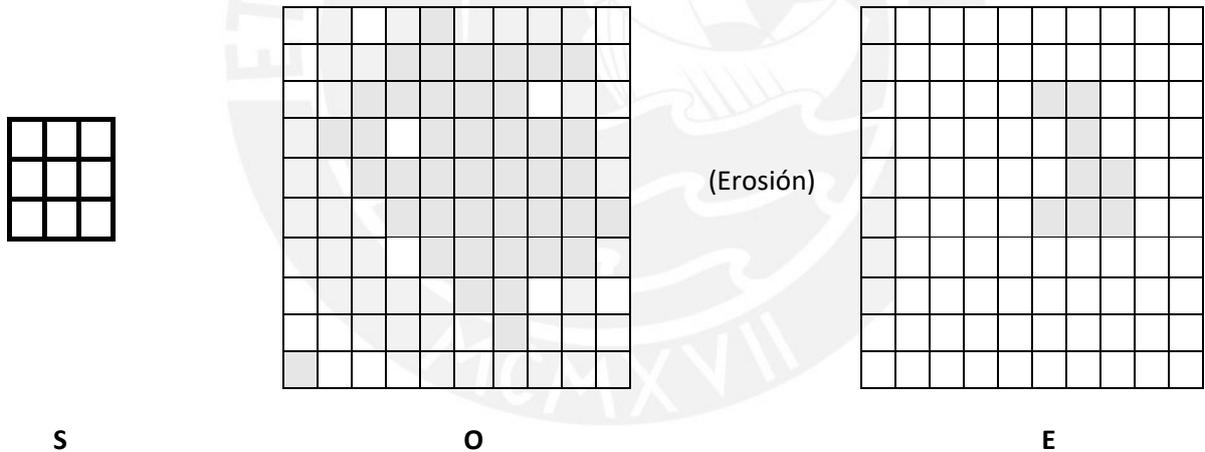


Figura 16: Ejemplo de aplicación del operador de Erosión

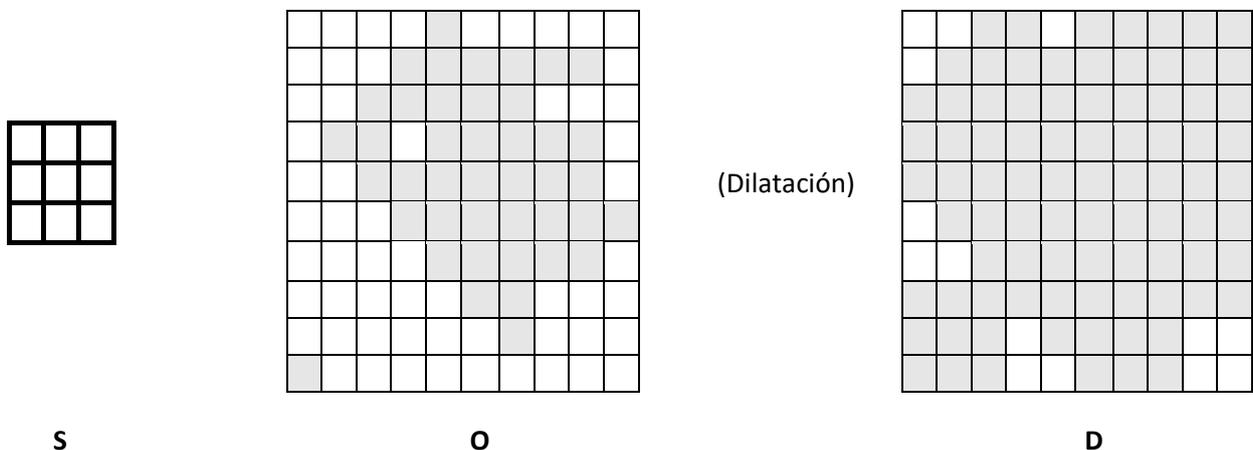


Figura 17: Ejemplo de aplicación del operador de Dilatación

Apertura: Composición del operador de erosión y dilatación. Esta composición se define como:

$$P = O \circ S = (O \ominus S) \oplus S$$

Donde: **O** es el objeto y **S** es la vecindad de pixeles en un pixel en específico. En la Figura 18 se presenta un ejemplo de la aplicación de este operador.

Cierre: Composición del operador de dilatación y erosión. Esta composición se define como:

$$C = O \bullet S = (O \oplus S) \ominus S$$

Donde: **O** es el objeto y **S** es la vecindad de pixeles en un pixel en específico. En la Figura 19 se presenta un ejemplo de la aplicación de este operador.

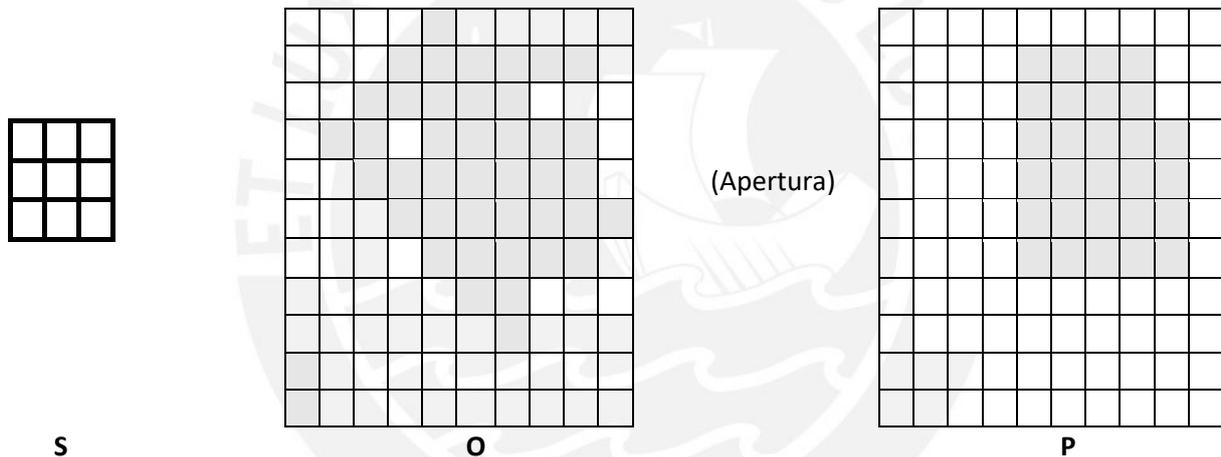


Figura 18: Ejemplo de aplicación del operador de Apertura

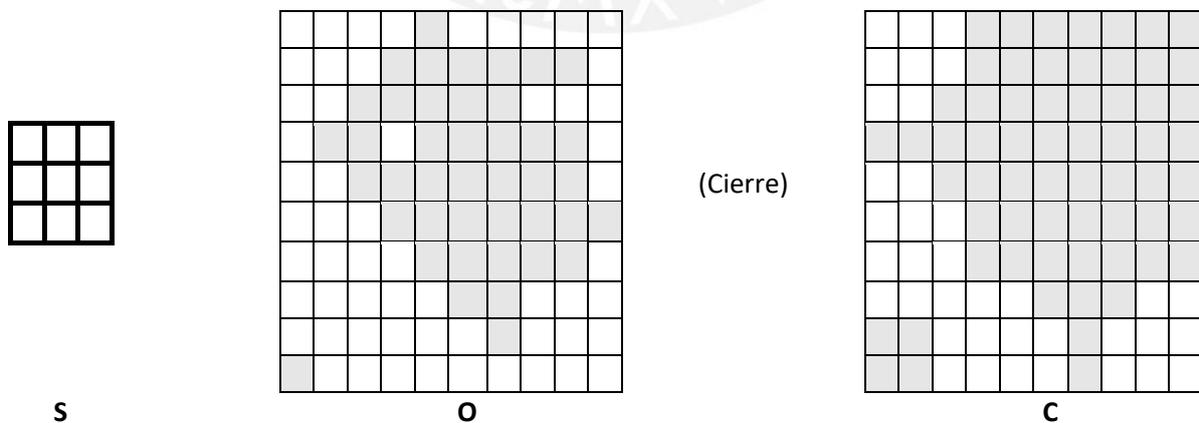


Figura 19: Ejemplo del operador de Cierre

Capítulo 3

Estado del arte

3.1 DETECCIÓN DE ASENTAMIENTOS INFORMALES

Según Kuffer en [15], la literatura sobre mapeo de asentamientos informales o barrios marginales aumentó en este nuevo milenio gracias a las imágenes de alta resolución obtenidas de los satélites. Este hecho se evidencia en la Figura 20, en la cual más del 60% de publicaciones, de un total de 110 publicaciones, referidas a mapeo de asentamientos con imágenes satelitales se han realizado en la última década. Este incremento en el uso de la teledetección para mapeo de asentamientos informales permitiría un monitoreo continuo necesario en favor de los sectores más vulnerables, que según la organización UN-habitat [2] consistiría en aproximadamente un cuarto de la población global.

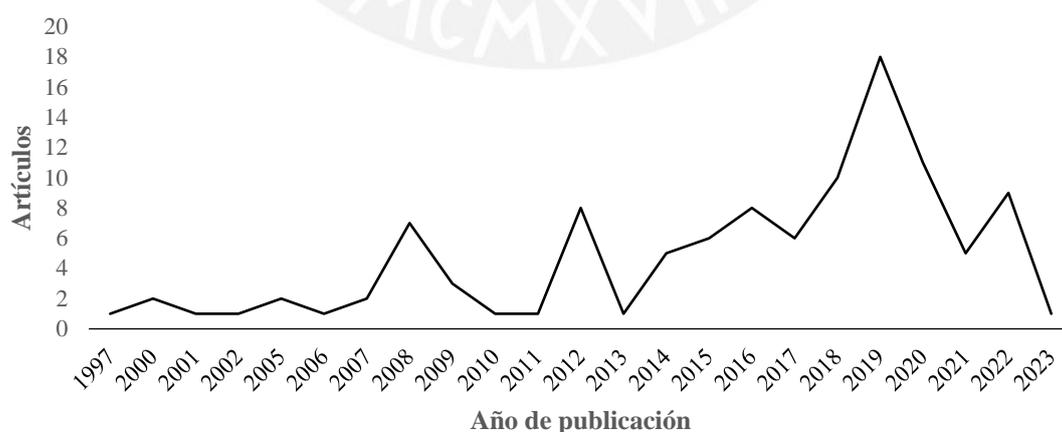


Figura 20: Número de artículos por año relacionados a las palabras clave "teledetección" y "asentamientos humanos" registrados en Scopus [16]

Diversos estudios, como los realizados por Taubenböck y Frankz [47], Owen y Wong [48] y Kohli, et al. [49], tienen como objetivo producir información sobre la geografía y dinámica de los asentamientos marginales debido a su conocimiento limitado y dinámica compleja. A pesar de estar disponibles en el mercado sensores comerciales con resoluciones de 0.5m, la diversidad y morfología compleja de estos asentamientos dificulta la extracción de información por métodos estándares como el análisis de imágenes basado en objetos (OBIA por sus siglas en inglés). Además, como se señala en [50] estos métodos carecen de transferibilidad y escalabilidad.

Los principales enfoques empleados para el mapeo de estos sectores son basados en áreas u objetos [51] [52], texturas [53], análisis temporal (TSA: *temporal series analysis*) [54] y el uso de algoritmos de *machine learning* [55] [56] [53] [57] [58] [59] con imágenes ópticas de alta resolución. Este último presenta una alta precisión al trabajar con extensas regiones de terreno y permite trabajar con grandes conjuntos de indicadores, según se evidencia en el trabajo de Ella et al. [53]. También se han empleado enfoques combinados empleando imágenes SAR de alta resolución como señala Ban [60]. En general estas metodologías analizan algunas de las variables presentadas en la Figura 21.

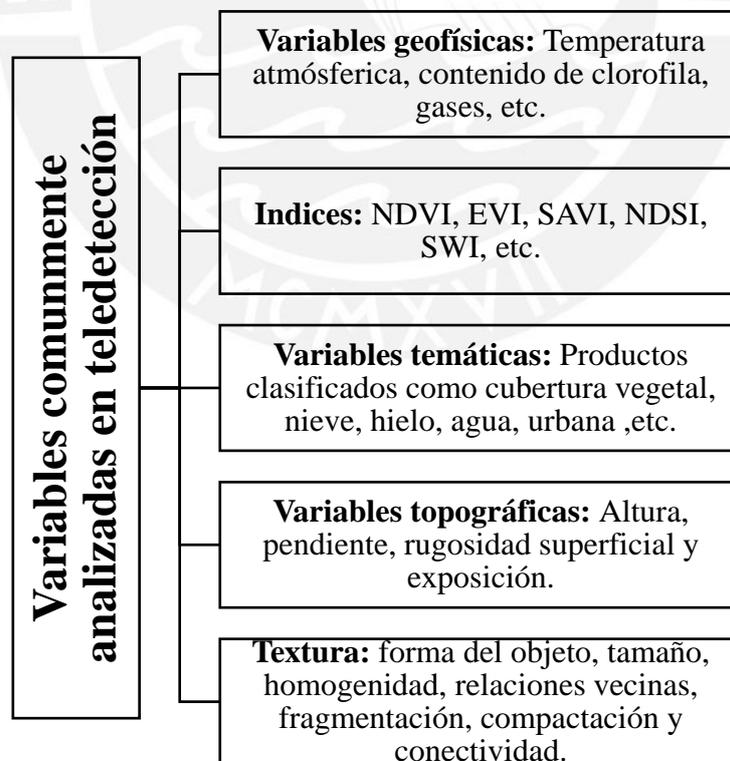


Figura 21: Variables comúnmente analizadas en teledetección.

Nota: Adaptado de Kuenzer et. al. [61]

Según el estudio de Kuffer [15] y Ambugadu y Hosni [62] los métodos más empleados son el OBIA, métodos basados en texturas y clasificación de imágenes de forma visual y con técnicas de *machine learning*. Dentro de las técnicas de *machine learning* más utilizadas se encuentran los algoritmos de aprendizaje automático supervisado como redes neuronales, máximo Likelihood Gausiano y máquina de soporte vectorial (SVM por sus siglas en inglés). Se presenta en Tabla 9 una comparación de estos métodos. Enfoques recientes empleados por Fallatah et. al. en [51] y [63], combinan las metodologías previamente descritas, mejorando los resultados en la detección de asentamientos informales.

Tabla 8: Comparación de algunos métodos de clasificación supervisado

Comparación	Máximo likelihood Gausiano	Máquina de soporte vectorial (SVM)	Redes Neuronales
Número de datos de entrenamiento	No necesita de mucha data, no obstante, se recomienda una cantidad considerable para evitar errores por la alta dimensionalidad de la data	Se recomienda una gran cantidad de data de entrenamiento para evitar el underfitting.	Usualmente se requiere de un alto número de datos de entrenamiento para que los pesos sean encontrados.
Tiempo de entrenamiento	Es de corta duración, depende de la estimación de las estadísticas.	Usualmente de larga duración, depende cuadráticamente del número de datos de entrenamiento.	Puede ser muy alto si se requiere un número alto de épocas para lograr la convergencia.
Selección de características	Esencial para data de alta dimensionalidad y evitar una pobre generalización. Consume mayor tiempo.	No es esencial, pero puede acelerar el entrenamiento	No es esencial, pero puede acelerar el entrenamiento
Tiempo de clasificación	Generalmente no alto. Es una función cuadrática de la dimensionalidad.	Depende linealmente de la dimensionalidad.	Generalmente no muy alto.

Nota: Traducido de Richards [44]

Los métodos más empleados para el mapeo de asentamientos marginales emplearon imágenes ópticas comerciales [64] [65], que son susceptibles a las condiciones climáticas debido a sus cortas longitudes de onda. Para el caso particular de la ciudad de Lima, según la página web *weather spark*, solo presenta un cielo con un porcentaje menor al 25% de nubosidad entre los meses de diciembre a marzo. Por lo tanto, se tendría una falta de información para el monitoreo el resto del año si se emplean imágenes ópticas, como se evidencia en la Figura 22.

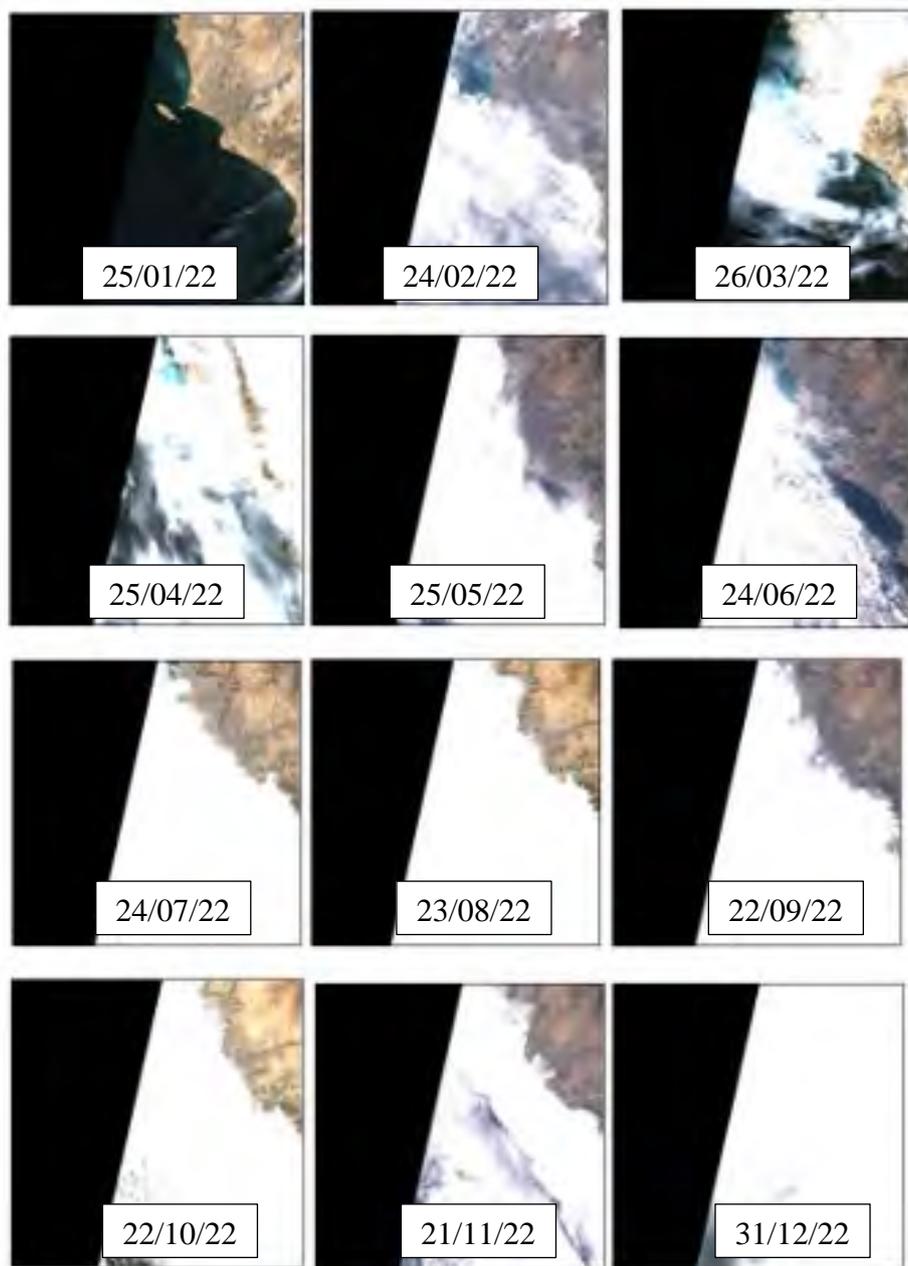


Figura 22: Imágenes ópticas de la constelación Setinel-1 sobre Lima en el año 2022

Manson y Frase [66] especificaron 3 principales características para monitorear barrios marginales de bajo costo: Adquisición y procesamiento de la data de acceso libre, como las disponibles en las plataformas de Copernicus y ASF (*Alaska Satellite Facility*), procesamiento semiautomático que entregue resultados rápidos y fiables y uso de softwares estándares de sencillo uso para operadores poco calificados.

Una metodología de monitoreo con imágenes de libre acceso de la constelación Sentinel 1 fue empleado por Moya et al. [67]. El método consistió en un análisis temporal y de *outliers* para la identificación de las pasadas invasiones en Loma de Corvina y Morro

Solar del 2021. Las características de estas invasiones fueron que ocurrieron de forma repentina por un grupo de más de 3000 personas en cada una. Para la detección de estas invasiones, se empleó de forma pionera imágenes de radar de media resolución de la constelación Sentinel-1, las cuales son de acceso abierto en la plataforma de Copernicus.

El código desarrollado por Moya et al. [67] para la identificación de las áreas vulneradas se basó en que en principio el *backscatter* o retrodispersión del haz electromagnético en la superficie terrestre se incrementa radicalmente en una topografía más escarpada, por lo que en una invasión con las características presentadas en Loma de Corvina y el Morro Solar, con asentamientos improvisados y de gran extensión, tendería a incrementar el índice de retrodispersión en las zonas invadidas.

Los resultados empleando esta metodología lograron detectar el 84% del área identificada por inspección visual empleando imágenes de drones. Estos resultados se pueden considerar satisfactorios debido a que a pesar de las complejas distorsiones de las imágenes de radar tipo SAR empleadas logró detectar de forma temporal este tipo de invasiones a partir de variaciones repentinas en el índice de retrodispersión de las imágenes de radar de libre acceso.

3.2 Metodología general para la clasificación digital de imágenes

A partir de la revisión de la literatura respecto a clasificación de imágenes, se puede identificar que la mayoría de metodologías empleadas comparten estas características:

- Analizar si se realizará un análisis digital o tratamiento digital
- Determinar si se emplearan imágenes satelitales o adicionalmente fotografías aéreas
- Definir la escala de trabajo para adaptar los métodos a esta
- Seleccionar el sensor y fecha de adquisición de las imágenes
- Aplicar las correcciones oportunas en las imágenes
- Definir etiquetas o índices en caso se empleen
- Clasificar, validar los resultados y obtener el inventario de los resultados. En la Tabla 10 se presenta alguna de las metodologías aplicadas.

Tabla 9: Metodologías empleadas en algunos algoritmos supervisados

Enfoque	Metodología
Máximo likelihood Gaussiano	<ol style="list-style-type: none"> 1. Obtener datos de entrenamiento etiquetados 2. Aplicar la reducción dimensional 3. Resolver la multimodalidad 4. Calcular estadísticas (media y covarianza) 5. Aplicar una función de discriminación 6. Generar el mapa
Máquina de soporte vectorial (SVM)	<ol style="list-style-type: none"> 1. Obtener datos de entrenamiento etiquetados 2. Definir la estrategia multiclase a usar 3. Definir la función <i>kernel</i> a usar 4. Posiblemente seleccionar características 5. Calibrar los parámetros de regularización y parámetros de la función <i>kernel</i> 6. Generar el mapa
Redes Neuronales	<ol style="list-style-type: none"> 1. Obtener datos de entrenamiento etiquetados 2. Seleccionar la codificación de las capas de entrada y salida 3. Elegir el número de capas ocultas y número de neuronas por capa 4. Entrenar la red de forma iterativa 5. Refinar la red eliminando o añadiendo nodos 6. Generar el mapa

Nota: Traducido de Richards [44]

Algunas de las limitaciones al momento de emplear esta metodología genérica son presentadas en la Tabla 11.

Tabla 10: Limitaciones generales al clasificar imágenes.

Desafío relacionado con:	Posibles limitaciones
Política de datos	<ul style="list-style-type: none"> Limitado número de adquisiciones Acceso restringido a los datos Composición de productos
Sensor	<ul style="list-style-type: none"> Deriva orbital Calibración errónea del sensor Degradación del sensor Errores de geolocalización Errores en la reflectancia Calibración lineal de los sensores a largo plazo Cambios lineales en los anchos de banda a largo plazo Cambios lineales en la resolución espacial a largo plazo
Localización	<ul style="list-style-type: none"> Cobertura de nubes Vapor de agua y aerosoles en la atmosfera Noche polar Terreno extremo Efecto de los rayos solares Efectos anisotrópicos Hundimiento repentino

Procesamiento	Descarga y almacenamiento
	Poder computacional
	Habilidades de programación
	Conocimiento de matemáticas
	Reprocesamiento mensual o anual
	Preservación a largo plazo

Nota: Adaptado de Kuenzer et. al. [7]

3.3 Metodología actual para la clasificación de imágenes multiespectrales

Se basa en identificar diversos tipos de cubierta y entrenamiento de algoritmos de clasificación para la identificación del uso del suelo como se mencionan en [68] . La cobertura hace referencia a la categorización de la superficie terrestre en base a sus propiedades biofísicas como zonas urbanas, cultivos, etc. Las cubiertas se pueden caracterizar en la mayoría de los casos por su comportamiento espectral.

Las principales ventajas de esta metodología son que se pueden adquirir imágenes extensas con rapidez sin la necesidad de acceder físicamente a la zona de estudio y, por lo general, la interpretación de las imágenes es rápida y menos costosa que las de trabajo en campo e inventario manual. Sin embargo, sus desventajas son que algunos tipos de cubiertas no son distinguibles por las imágenes satelitales, puede resultar poco económico en zonas muy pequeñas dado que requeriría imágenes satelitales comerciales de alta resolución y la sombra y presencia de nubes pueden afectar los resultados de la clasificación.

Capítulo 4

Metodología

4.1 Introducción

El sistema radar de apertura sintética (SAR) permite obtener imágenes a partir del índice de retrodispersión. Este índice depende principalmente de la rugosidad de la superficie de estudio como se evidencia en la Figura 23. En base a este hecho, se puede detectar cambios de uso del lugar, como las repentinas invasiones en zonas no urbanas, a partir de un estudio temporal del cambio de patrones de retrodispersión empleando algoritmos de clasificación.

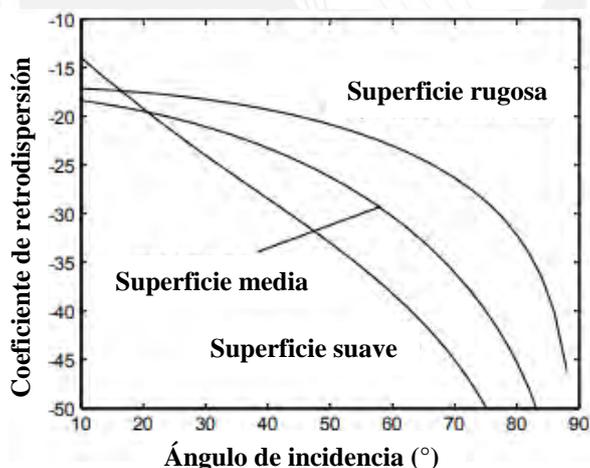


Figura 23: Variación promedio del índice de retrodispersión en superficie con diferente rugosidad.

Nota: Adaptado de Richards [45]

La metodología consiste en obtener un conjunto de imágenes temporales del área de interés no urbano en zonas de alto peligro de desastre. Se empleará el producto GRD pues para los fines de este sistema de detección solo bastará conocer la intensidad de la retrodispersión del área de estudio. Como ventaja adicional, los productos GRD se almacenan en archivos de menor tamaño en comparación de los otros formatos.

Las imágenes adquiridas serán procesadas para reducir el ruido en la data, alinear las imágenes y recortar el tamaño de las imágenes para trabajar solo en el área de interés. Para este fin se empleará el software SNAP el cual posee un entorno amigable e intuitivo para el procesamiento de las imágenes. Estas imágenes se guardarán en un formato .data y se alinearán usando el lenguaje de programación Python.

Después se procederá a ejecutar el algoritmo de detección con las imágenes procesadas para el monitoreo y detección temprana de las invasiones. La idea clave del algoritmo es que en las zonas de alto peligro no urbana es menos probable que circulen vehículos, personas y se instalen edificaciones en el lugar. Por lo tanto, el índice de retrodispersión debe mantenerse bajo debido a que no se produciría el efecto de esquina reflectora, el cual es común ciudades urbanizadas, a menos que de forma improvisada se realizaran construcciones informales.

Finalmente, en caso que el algoritmo detecte una invasión, se podrá emitir una alerta temprana para su posterior confirmación y notificación de las autoridades.

4.2 Herramientas, métodos y procedimientos a usar

Se elaboró la Tabla 2 para presentar las herramientas a emplear y procedimientos a seguir, asociados con cada resultado esperado de este proyecto.

Tabla 11: Resultados esperados, herramientas a usar y métodos a utilizar en la investigación.

Resultados esperados	Herramientas a utilizar	Métodos a utilizar
Preprocesamiento de imágenes satelitales	<ul style="list-style-type: none"> • SNAP 	<ul style="list-style-type: none"> • Apply-Orbit-File • Calibration • Subset • Speckle filter • Terrain correction • Linear to db
Modelos de detección de outliers pixeles basados en intensidades para identificar asentamientos informales	<ul style="list-style-type: none"> • Python • PyTorch • NumPy • Matplotlib • Sklearn • Scipy • Os 	Algoritmos de aprendizaje supervisado: <ul style="list-style-type: none"> • SVM • Logistic Regression • Random Forest

Modelos basados en redes neuronales para el reconocimiento de invasiones	<ul style="list-style-type: none"> • Glob • Gdal • Ogr • Osr • Cv2 	Algoritmos de aprendizaje semi supervisado: <ul style="list-style-type: none"> • OCSVM • Isolation Forest • Elliptic Envelop • Local Outlier Factor
--	---	---

Algunas de los softwares empleados, con sus características, se presentan a continuación:

SNAP

SNAP (*Sentinel Application Platform*) es un software desarrollado por Brockmann Consult, Skywatch, Sensor y C-S y distribuido gratuitamente por la ESA (European Spatial Agency). Es una plataforma ideal para el acceso a imágenes satelitales de distintos sensores, procesamiento y análisis de imágenes satelitales.

Python

Es un lenguaje de programación, de libre acceso, creado por Guido van Rossum. Es ampliamente utilizado en el desarrollo de software, ciencia de datos y machine learning. Las librerías empleadas para el desarrollo de este trabajo son: gdal, numpy, matplotlib, os, cv2, glob, sklearn y Pytorch.

QGIS

QGIS es un sistema de código abierto para la manipulación y lectura de información geográfica. Permite crear y abrir diferentes tipos de archivos georreferenciados, tales como capas vectoriales e imágenes de radar.

4.3 Obtención y preprocesamiento de imágenes satelitales SAR

Las imágenes de radar de la constelación Sentinel-1 pueden obtenerse de manera automatizada de la plataforma “ASF Data Search Vertex” [25]. Algunas de las características de las imágenes utilizadas son:

- Modo de Adquisición: IW. La elección se realizó en base a la Tabla 12
- Producto: GRD (Ground Resolution Distance)
- Dirección de vuelo: ascendente o descendente
- Resolución: 10 m

Es importante mencionar que, durante un análisis de identificación de invasiones, todas las imágenes deben ser adquiridas bajo los mismos parámetros. Es decir, mismo modo de adquisición, misma dirección de vuelo y mismo ángulo de incidencia. Usar un conjunto de imágenes con distintos parámetros de adquisición generará falsas alarmas.

Tabla 12: Modos de adquisición y aplicación. Fuente: [36]

Aplicación	Modo de adquisición			
	SM	IW	EW	WV
Zonas árticas y hielo marino				
Vigilancia de embarcaciones en mar abierto				
Vigilancia de la contaminación de hidrocarburos				
Vientos marinos				
Ámbito forestal				
Agricultura				
Cartografiado de deformaciones urbanas				
Cartografiado de inundaciones				
Análisis de terremotos				
Vigilancia de deslizamientos y volcanes				

Nota: Adaptado de IGN [36]

El preprocesamiento de las imágenes se realizará de forma inicial con la finalidad de calibrar, alinear, reducir el ruido y tamaño de las imágenes para tener un proceso estandarizado y obtener resultados precisos. Este preprocesamiento consiste en una serie de procesos que se aplican en el orden mostrado en la Figura 24.

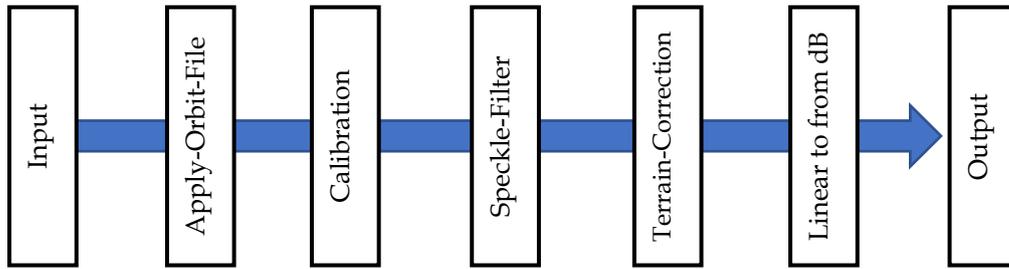


Figura 24: Flujo de trabajo para el preprocesamiento de imágenes SAR en el software SNAP

Apply-Orbit-File

Las imágenes satelitales contienen información con respecto a las coordenadas de la órbita del satélite al momento de la toma de la imagen. La precisión de dichas coordenadas pueden mejorarse debido a que el satélite es detectado por otros satélites y estaciones terrestres. Estas órbitas precisas se publican luego de varios días independientemente de la publicación de las imágenes. Por lo tanto, este proceso consiste en la actualización de las orbitas en las imágenes Sentinel-1. Como consecuencia, las coordenadas de las imágenes de radar contarán con una alta precisión. Se presentan los resultados de la aplicación de este proceso en la Figura 25.



Figura 25: Izquierda: Imagen con órbita estándar. Derecha: con órbita precisa

Radiometric Calibration

Con el objetivo de comprimir la información de las imágenes, estas se proporcionan en valores digitales de 16 bits por píxel. Es decir, los valores de los píxeles pueden variar entre 0 y 65535. Con el objetivo de recuperar la información física de la energía retrodispersada, se aplica el proceso de Calibración Radiométrica. Se presentan los resultados de la aplicación de este proceso en la Figura 26.

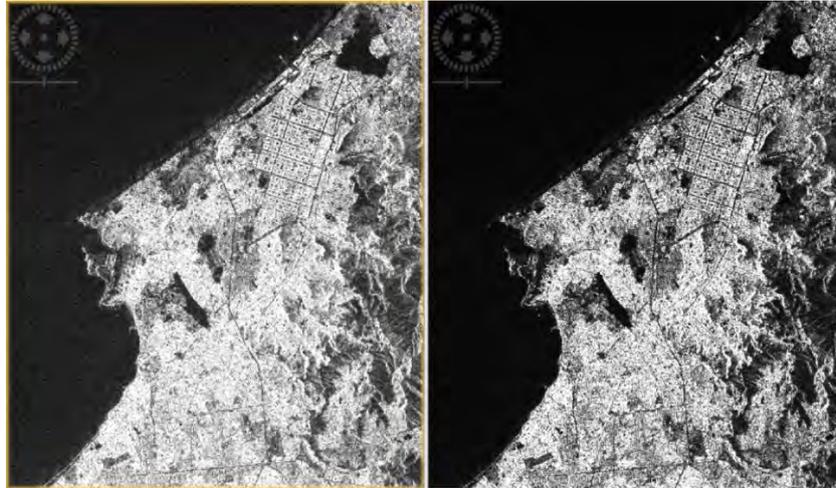


Figura 26: Izquierda: imagen Sentinel-1 en valores digitales. Derecha: Imagen luego de aplicar la calibración radiométrica

Speckle-Filter

El *Speckle* es un defecto aleatorio producido por una interferencia negativa o positiva generado por todos los dispersores dentro de cada pixel. El filtro *speckle* se aplica para reducir el ruido por en las imágenes. Se puede apreciar que el resultado de la aplicación de este filtro en la Figura 27 es el suavizado de las intensidades en la imagen.



Figura 27: Imagen Sentinel-1 antes (izquierda) y después (derecha) de aplicar el filtro Speckle

Terrain-Correction

Consiste en reducir el efecto de la topografía en los valores de la retrodispersión. Corrige las distorsiones geométricas en los valores de geolocalización. Estas se producen debido a que el sensor emite la señal de forma lateral. Por lo tanto, áreas con variaciones significativas presentarán deformaciones en las imágenes de radar. Para mitigar este efecto, se emplea un Modelo de Elevación Digital (DEM) en este proceso. Es importante

resaltar que el nivel de desempeño de este proceso dependerá mucho de la fecha de adquisición del DEM. Desafortunadamente, no se cuenta con un DEM de alta resolución de Lima. En este estudio se empleó el DEM SRTM 3 SEC, modelo satelital elaborado por la NASA el 2014 con resolución de 30m. Notar también que, luego de aplicar el proceso Terrain-Correction, la imagen resultante tiene una correcta orientación con respecto a un sistema de referencia de coordenadas esto se ve agravado por las condiciones accidentadas del terreno. Se presentan los resultados de la aplicación de este proceso en la Figura 28.

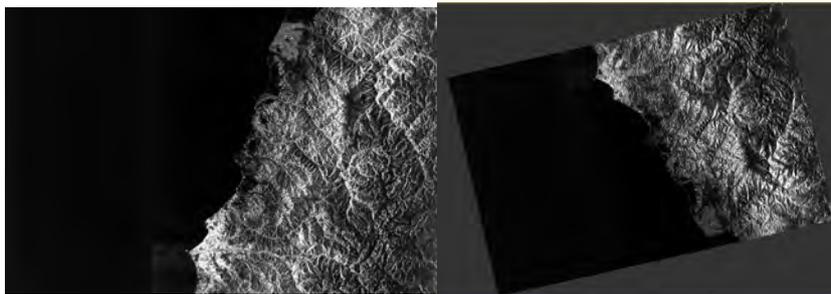


Figura 28: Imagen Sentinel-1 antes (izquierda) y después (derecha) de la aplicación del proceso Terrain correction. El eje vertical en la imagen de la izquierda representa la dirección de vuelo del satélite y el eje horizontal es la dirección perpendicular a la dirección de vuelo. El eje vertical y horizontal de la imagen derecha corresponden a las coordenadas latitud y longitud del sistema WGS84

Linear to from dB

Convierte en escala logarítmica los valores de retrodispersión. Es utilizado para comprimir la data y facilita la detección visual de cambios temporales. Se presentan los resultados de la aplicación del escalamiento logarítmico en la Figura 29.

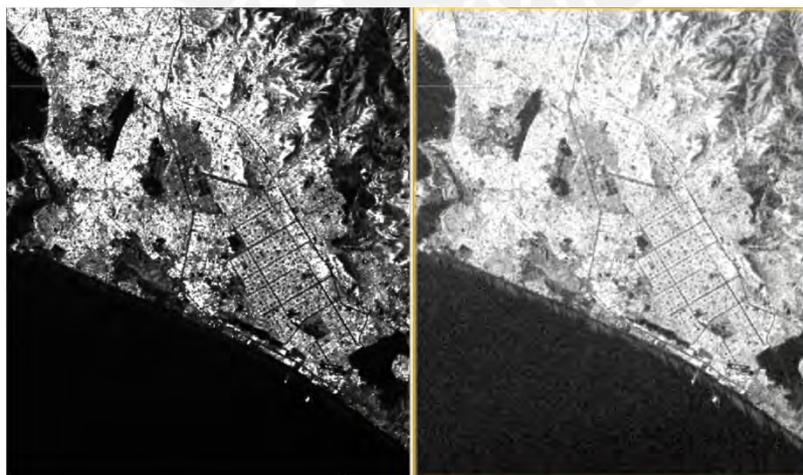


Figura 29: Imagen Sentinel-1 antes (izquierda) y después (derecha) de la aplicación del escalamiento logarítmico

Resampling y subset de imágenes empleando un sistema de grillas

Se ejecuto los siguientes pasos para reducir el tamaño de las imágenes, con el fin de reducir el coste computacional y enfocarse en las áreas de monitoreo, y normalizar el tamaño y resolución de las imágenes. Primero se definió una grilla en la ciudad de Lima en Qgis. Esta grilla se presenta en la Figura 30. Cada cuadrícula, dentro de las grillas, es cuadrada de extensión de 7.5 km.



Figura 30:(a) Grilla definida en la ciudad de Lima. (b) Cuadrículas de estudio

4.4 Detección de invasiones

Considerar un conjunto N de imágenes de radar procesadas $\{I_p\}_{p=1}^N$ donde cada imagen $I_p \in \mathbb{R}^3$ contiene $k \times l$ píxeles definidos en un espacio físico y temporal. Cada pixel de la imagen p contiene información de las coordenadas e índice de retrodispersión y se denota como $(I_{p,i,j})$, donde i y j denotan respectivamente la posición fila y columna del pixel en la imagen p . El análisis de detección se realizará a nivel de píxeles. Por lo tanto, una muestra arbitraria se representa de la siguiente manera:

$$\text{Ecuación 1 } x_k = (\bar{I}_{1,i,j}, \bar{I}_{2,i,j}, \dots, \bar{I}_{N,i,j})$$

$$\text{Ecuación 2 } = I_{q,i,j} - \frac{\sum_{l=1}^{N-1} I_{l,i,j}}{N-1}$$

Donde x_k es un vector que contiene la variación en el tiempo del coeficiente de retrodispersión de un pixel. Notar que en este análisis temporal, la imagen tomada en la

última fecha, I_N , es la imagen que podría contener una invasión de terreno. Se espera entonces que la variación temporal del valor de píxeles antes de la ocurrencia de la invasión es baja. Por lo tanto, para uniformizar todas las muestras, se ha removido el promedio de los valores del píxel en las fechas previa a la invasión.

Considerando el conjunto de M muestras $P = \{x_k\}_{k=1}^M$, el objetivo es asignar a cada muestra una etiqueta binaria, donde la etiqueta $y_k = 0$ indique que no ha ocurrido una invasión dentro del píxel, caso contrario, la etiqueta $y_k = 1$ indica una invasión de terreno. Para dicha tarea, se busca calibrar una función de clasificación bicategorica $g(x_k)$, la cual permite identificar si existió un cambio relevante en el índice de dispersión de cada píxel en la última imagen y por lo tanto se puede detectar si ocurrió una invasión repentina. Finalmente a los resultados predichos para la fecha se le aplica el operador morfológico de cierre y una función de filtro de cluster.

Para la elección de la función de clasificación, se realizará una evaluación experimental a un conjunto de métodos de clasificación de *machine learning*. Para el caso de métodos supervisados, se empleará un conjunto independiente $T = \{x_k, y_k\}$ cuyas etiquetas se conocen de antemano. Para el caso de métodos no-supervisados, se empleará directamente el conjunto P . Para las evaluaciones se empleará la matriz de confusión.

4.5 Limitaciones

Las imágenes SAR, debido a sus complejas proyecciones geométricas, tiene errores inatenuables que generan errores en la precisión y exactitud en los resultados del modelo a emplear. Además, se debe procurar evaluar invasiones de gran extensión para evitar confundirlo con ruido aleatorio de las imágenes de radar.

La solución que se plantea depende fuertemente al libre acceso de imágenes de radar. Actualmente el Gobierno Peruano no cuenta con un sistema para tomar imágenes de radar. Por lo tanto, dependemos de la posición de la Unión europea de proporcionar sus imágenes de forma gratuita.

Capítulo 5

Evaluación experimental

5.1 Caso de estudio: invasiones en el Morro Solar, Chorrillos, y Lomo de Corvina, Villa El Salvador del 2021

Según el diario La República, el pasado 5 y 12 de abril del 2021 se iniciaron las invasiones y lotizados de las zonas de El Morro Solar y Loma de Corvina ubicados en los distritos de Chorrillos y Villa el Salvador respectivamente según se puede apreciar en la Figura 31. Es importante recalcar según los estudios de microzonificación realizados por el CISMID en Villa del Salvador [69] y Chorrillos [70] que las áreas invadidas tiene un suelo poco acto para la construcción según se aprecia en la Figura 31 a). En la Figura 31 b) y c) se puede observar que se construyeron viviendas improvisadas a base de madera, esteras y plástico.

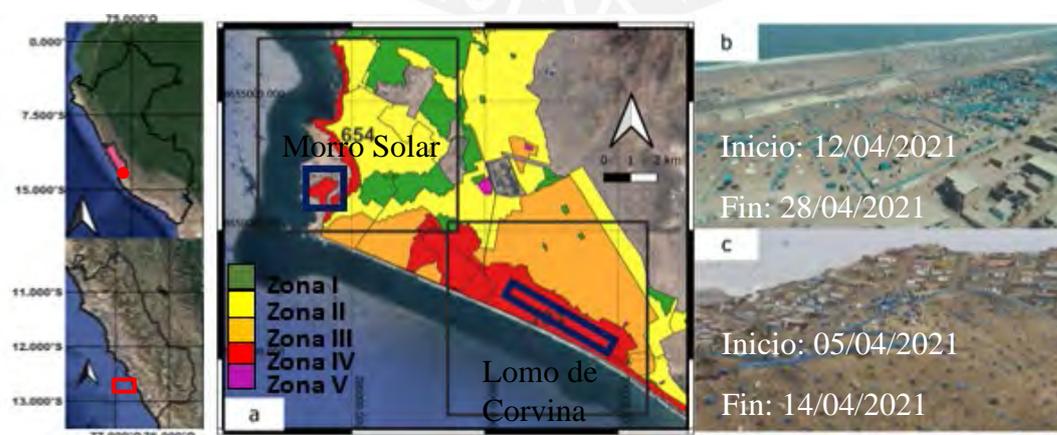


Figura 31: Áreas de estudio. (a) Localización de las áreas de estudio y tipo de zonas. Fuente: CISMID. (b) Invasión en Lomo de Corvina. Fuente: El Comercio. (c) Invasión del Morro Solar. Fuente: El Comercio

5.2 Descripción de los datos

En la Tabla 13 se muestra un resumen de las características de las imágenes empleadas del área de estudio y en la Figura 32 se muestra la distribución de la retrodispersión en la cuadrícula número 703, el cual contiene el área de Lomo de Corvina, durante las fechas de análisis. Se puede apreciar que la distribución de intensidades tiene una forma bimodal debido a que se esperan valores de intensidades menores a -15db en zonas no urbanas y en masas de agua y valores mayores en zonas urbanas.

Tabla 13: Fecha y características de las imágenes empleadas

Fechas		Características
22-09-20	01-02-21	<ul style="list-style-type: none"> • Producto: GRD (Ground Resolution Distance) • Orbita relativa: 18 • Dirección de vuelo: ascendente • Resolución: 10 m • Fuente: ASF Data Search Vertex
04-10-20	13-02-21	
16-10-20	25-02-21	
28-10-20	09-03-21	
09-11-20	21-03-21	
21-11-20	02-04-21	
03-12-20	14-04-21	
15-12-20	26-04-21	
27-12-20	08-05-21	
08-01-21	20-05-21	
20-01-21	01-06-21	
13-06-21		

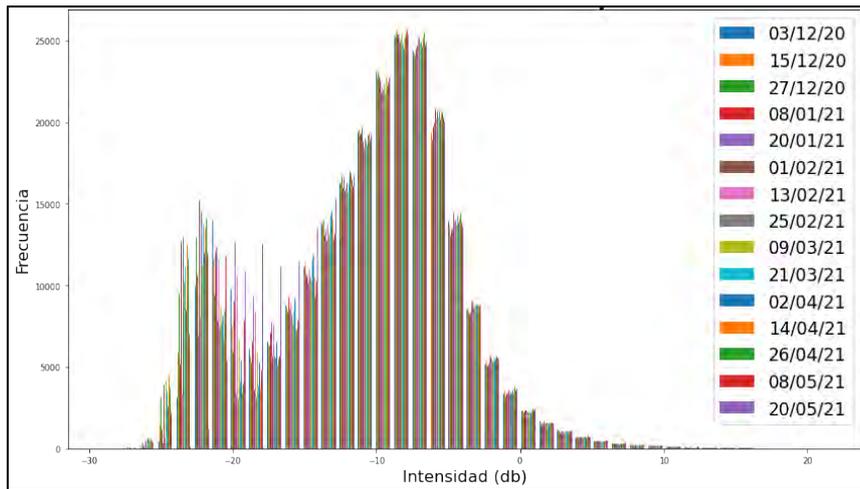


Figura 32: Distribución de la retrodispersión en el área de estudio.

En la Figura 33 se muestra la distribución espacial de la retrodispersión de un área urbana. Se puede apreciar que esta distribución es unimodal y se puede ajustar con una distribución de densidad gaussiana de la siguiente forma:

$$f(x) = \frac{1}{2 * \pi * 2.8} * e^{-\frac{(x+7.37)^2}{2*2.8^2}}$$

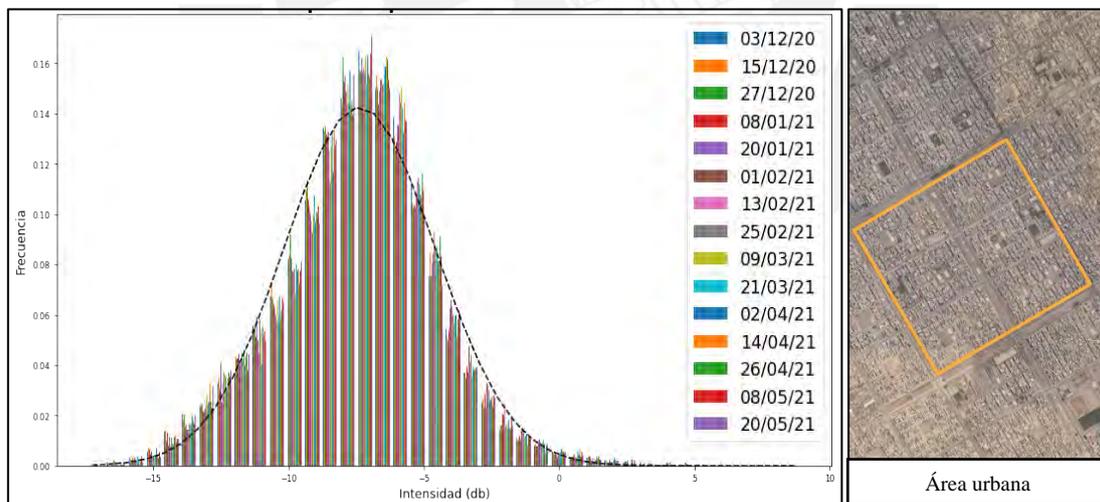


Figura 33: Distribución de la retrodispersión en una zona urbana.

En la Figura 34 se muestra la distribución espacial de la retrodispersión de un área no urbana. Se puede apreciar que esta distribución es unimodal y se puede ajustar con una distribución de densidad gaussiana de la siguiente forma:

$$f(x) = \frac{1}{2 * \pi * 1.92} * e^{-\frac{(x+19.05)^2}{2*1.92^2}}$$

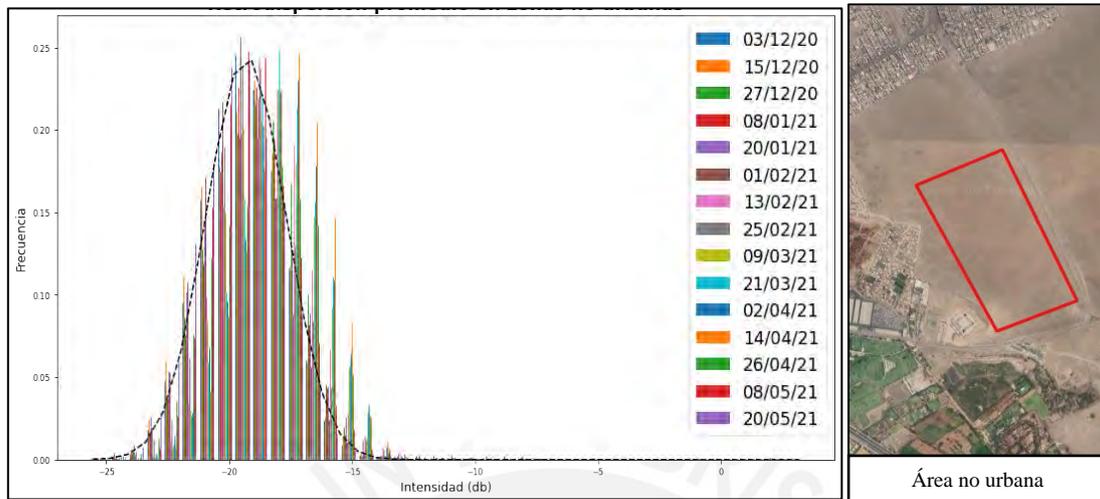


Figura 34: Distribución de la retrodispersión en una zona no urbana.

De los resultados mostrados previamente es importante destacar que en promedio la retrodispersión obtenida en zonas urbanas (-7.37 db) es mayor que en zonas no urbanas (-19.05 db) y que la desviación estándar obtenida en promedio en zonas no urbanas (1.92 db) resulto menor que en las zonas urbanas (2.8 db). Es importante destacar que se emplearon los polígonos mostaza y rojo presentes en las Figura 35 y Figura 36 para poder entrenar los algoritmos como áreas no invadidas.

En la Figura 35 se presenta la distribución de retrodispersiones del área invadida en Lomo de Corvina. Se puede apreciar la data un poco dispersa, esto se debe a que se presentan las distribuciones de retrodispersiones antes y durante la invasión. El polígono verde que se muestra en la figura mencionada representa la extensión de la invasión. Esta se obtuvo de [67], el cual se generó mediante imágenes tomadas por un dron, y se utilizará para validar las predicciones de áreas invadidas

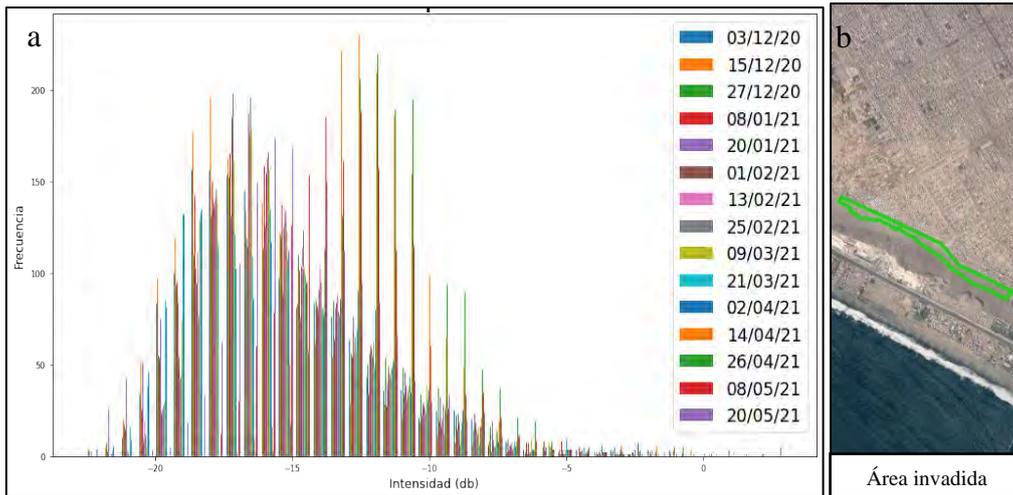


Figura 35: (a) Distribución de la retrodispersión del área invadida. (b) Polígono verde que representa el área invadida en Lomo de Corvina.

Se puede apreciar en la Figura 36 que la retrodispersión promedio antes de la invasión era de -16.8 db y durante la invasión -12.5 db respectivamente.

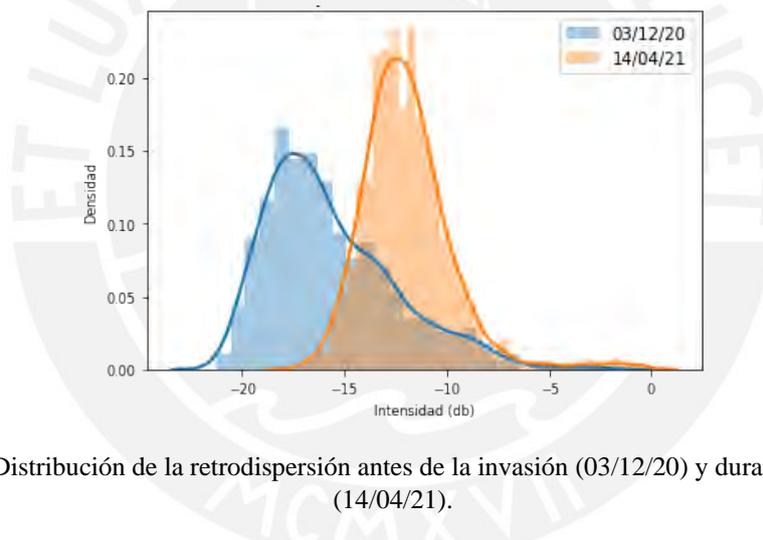


Figura 36: Distribución de la retrodispersión antes de la invasión (03/12/20) y durante la invasión (14/04/21).

En la Figura 37 se presenta la retrodispersión promedio en la zona invadida de Lomo de Corvina versus el tiempo. Como era de esperarse según la Figura 36 durante la invasión se produjo un incremento significativo de la retrodispersión.

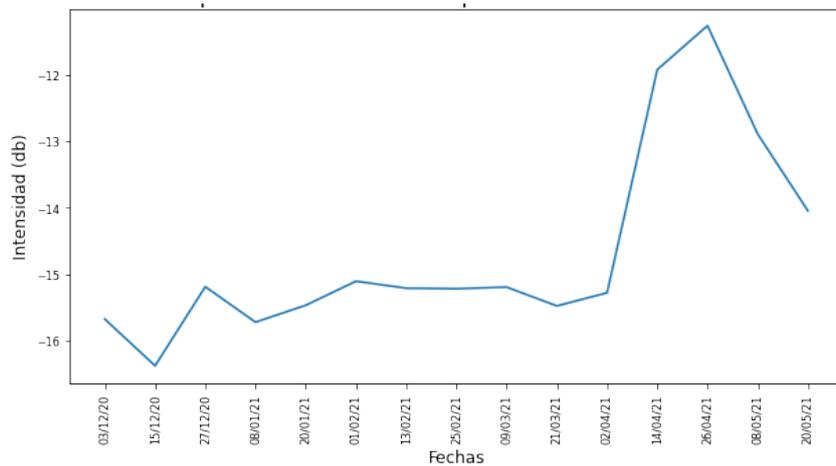


Figura 37: Variación de la intensidad promedio de la retrodispersión en el área invadida.

En la Figura 38 se presenta un polígono en la zona del Morro Solar. Dicha zona representa una zona interna de la invasión que fue delimitada manualmente. Se tomaron precauciones para no considerar los bordes de la extensión de la invasión por que no se contó con ninguna imagen aérea para su confirmación. La distribución de la retrodispersión luego de la invasión se muestra en color naranja y la distribución antes de la invasión se muestra en color azul. Nuevamente se observa el mismo patrón, hay un incremento del valor promedio de la retrodispersión en zonas no urbanas durante la invasión. Para la calibración de los algoritmos supervisados usará los pixeles dentro del polígono rojo como zonas invadidas para definirse el conjunto T (Ecuación 1).

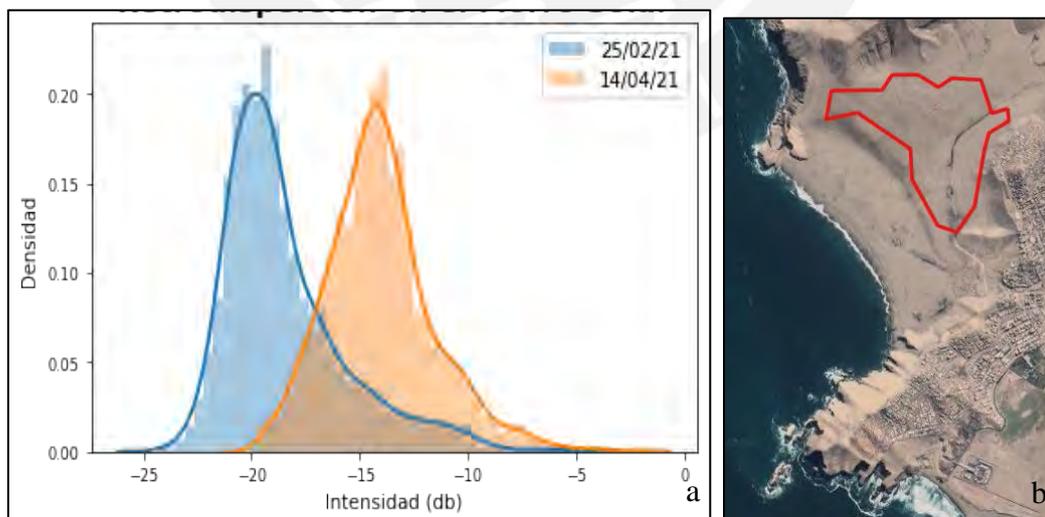


Figura 38: (a) Distribución de la retrodispersión en la zona del Morro Solar. (b) Polígono definido dentro del área invadida en el Morro Solar.

5.3 Resultados cuantitativos

En la presente sección se presentan los resultados de la aplicación de distintos algoritmos de *machine learning*. En la Tabla 14 se presenta las fechas de las imágenes empleadas para el entrenamiento tanto de los algoritmos supervisados y semi-supervisados. Se emplearán las métricas *overall*, *accuracy*, *recall*, *F1* y *kappa* para evaluar los resultados obtenidos.

Tabla 14: Fechas empleadas para el entrenamiento de los algoritmos.

Fechas
20-01-21
01-02-21
13-02-21
25-02-21
09-03-21
21-03-21
02-04-21
14-04-21

Árbol de decisión

En la Figura 39 se presenta el área predicha de la invasión en Villa Del Salvador. En la Tabla 15 se muestra los resultados de las métricas de validación de los resultados.



Figura 39: (a) Área predicha por el algoritmo de Árbol de decisión como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 15: Evaluación de los resultados predichos por el algoritmo de árbol de decisión

		Métricas				
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 559358 & 1593 \\ 232 & 1317 \end{pmatrix}$	0.9968	0.4526	0.8502	0.5907	0.5892	

En la Figura 40 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

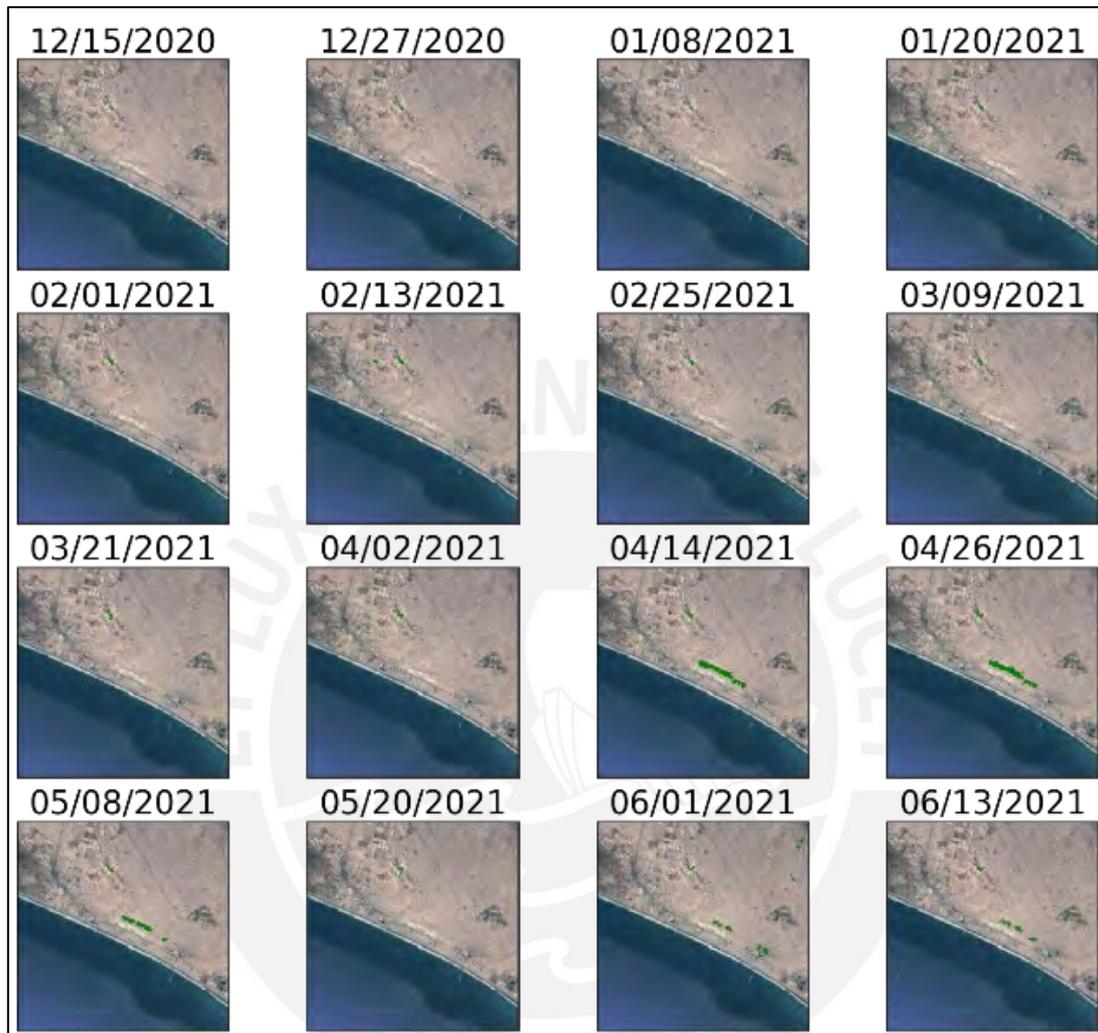


Figura 40: Resultados temporales mostrados por el algoritmo de Árbol de decisión.

Bayes Ingenuo Gaussiano

Se presenta el área predicha de la invasión en Villa Del Salvador en la Figura 41. En la Tabla 16 se muestra los resultados de las métricas de validación de los resultados.



Figura 41: (a) Área predicha por el algoritmo de Bayes ingenuo Gaussiano como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 16: Evaluación de los resultados predichos por el algoritmo de Bayes ingenuo Gaussiano

Métricas						
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 559112 & 1839 \\ 150 & 1399 \end{pmatrix}$	0.9965	0.4321	0.9032	0.5845	0.5829	

En la Figura 42 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

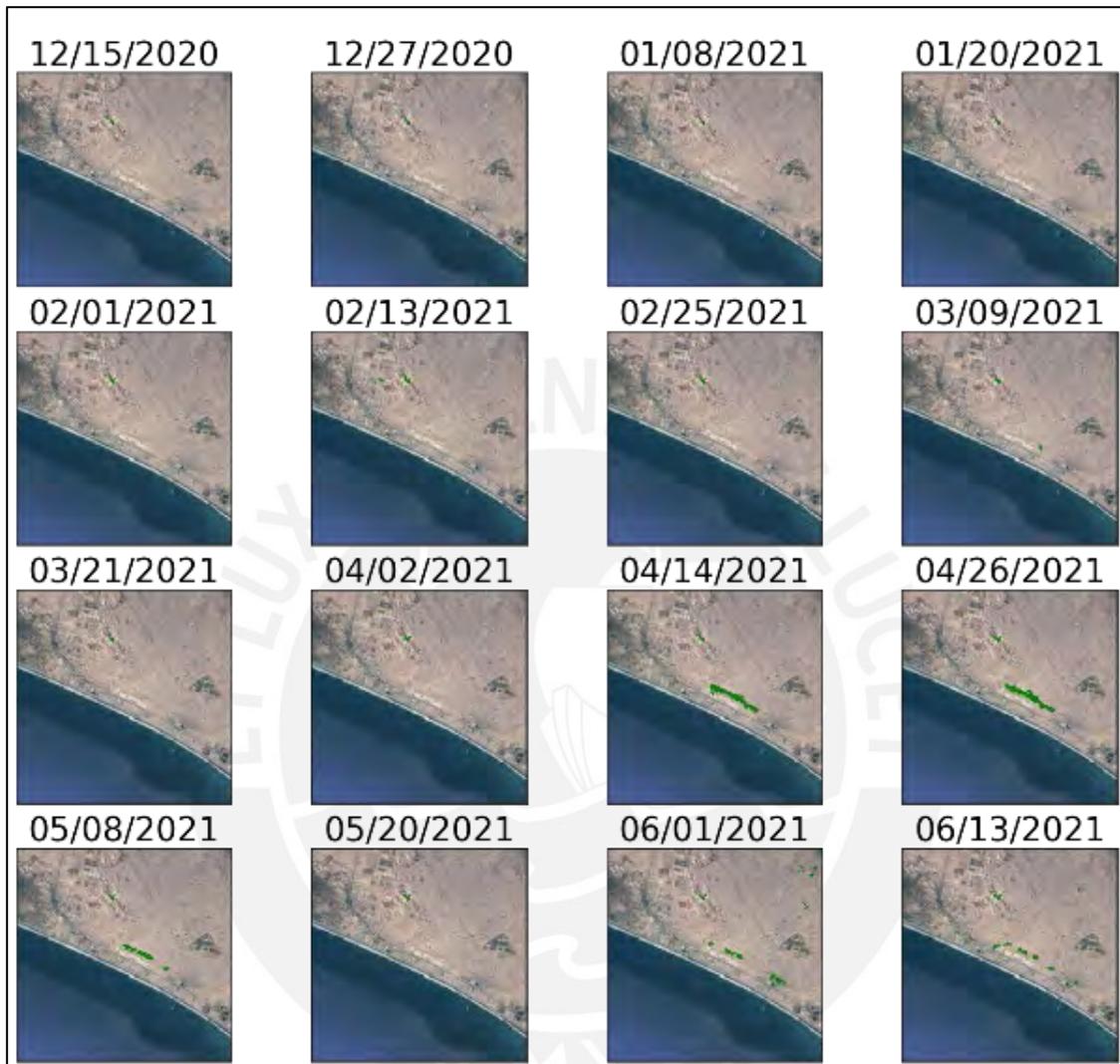


Figura 42: Resultados temporales mostrados por el algoritmo de Bayes ingenuo Gaussiano

Bosques Aleatorios

En la Figura 43 se presenta el área predicha de la invasión en Villa Del Salvador. En la Tabla 17 se muestra los resultados de las métricas de validación de los resultados.



Figura 43: (a) Área predicha por el algoritmo de Bosques aleatorios como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 17: Evaluación de los resultados predichos por el algoritmo de Bosques aleatorios

Métricas						
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 559006 & 1945 \\ 97 & 1452 \end{pmatrix}$	0.9964	0.4274	0.9374	0.5871	0.5856	

En la Figura 44 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

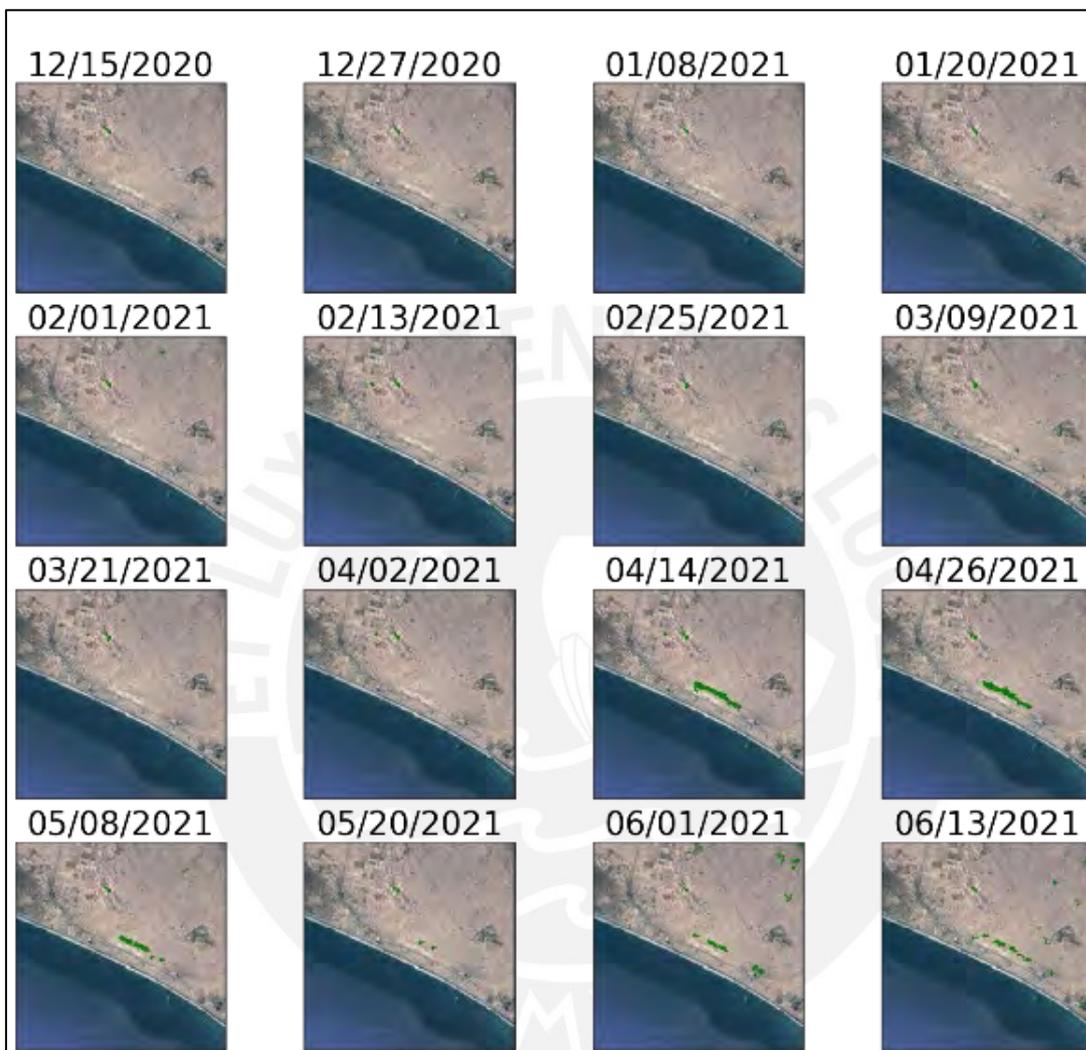


Figura 44: Resultados temporales mostrados por el algoritmo de Bosques aleatorios

Máquina de Soporte Vectorial

Se presenta el área predicha de la invasión en Villa Del Salvador en la Figura 45. En la Tabla 18 se muestra los resultados de las métricas de validación de los resultados.



Figura 45: (a) Área predicha por el algoritmo de Máquina de soporte vectorial como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 18: Evaluación de los resultados predichos por el algoritmo de Máquina de soporte vectorial

Métricas						
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 559536 & 1415 \\ 200 & 1349 \end{pmatrix}$	0.9971	0.4881	0.8709	0.6256	0.6242	

En la Figura 46 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

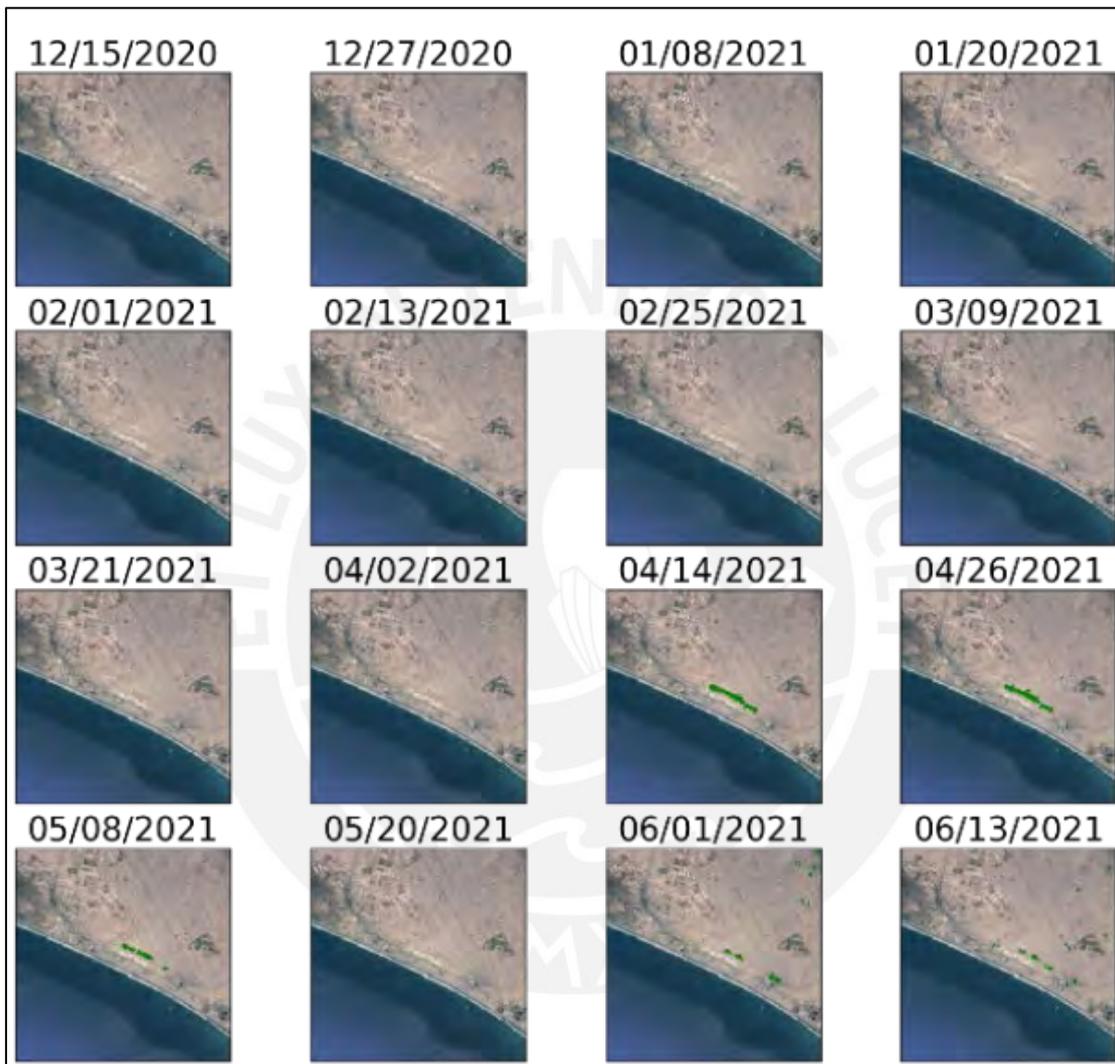


Figura 46: Resultados temporales mostrados por el algoritmo de Máquina de soporte vectorial

Descenso del Gradiente Estocástico

En la Figura 47 se presenta el área predicha de la invasión en Villa Del Salvador. En la Tabla 19 se muestra los resultados de las métricas de validación de los resultados.



Figura 47: (a) Área predicha por el algoritmo de Descenso del gradiente estocástico como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 19: Evaluación de los resultados predichos por el algoritmo de Descenso del gradiente estocástico

Métricas					
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa
$\begin{pmatrix} 558628 & 2323 \\ 127 & 1422 \end{pmatrix}$	0.9956	0.3797	0.9180	0.5372	0.5354

En la Figura 48 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

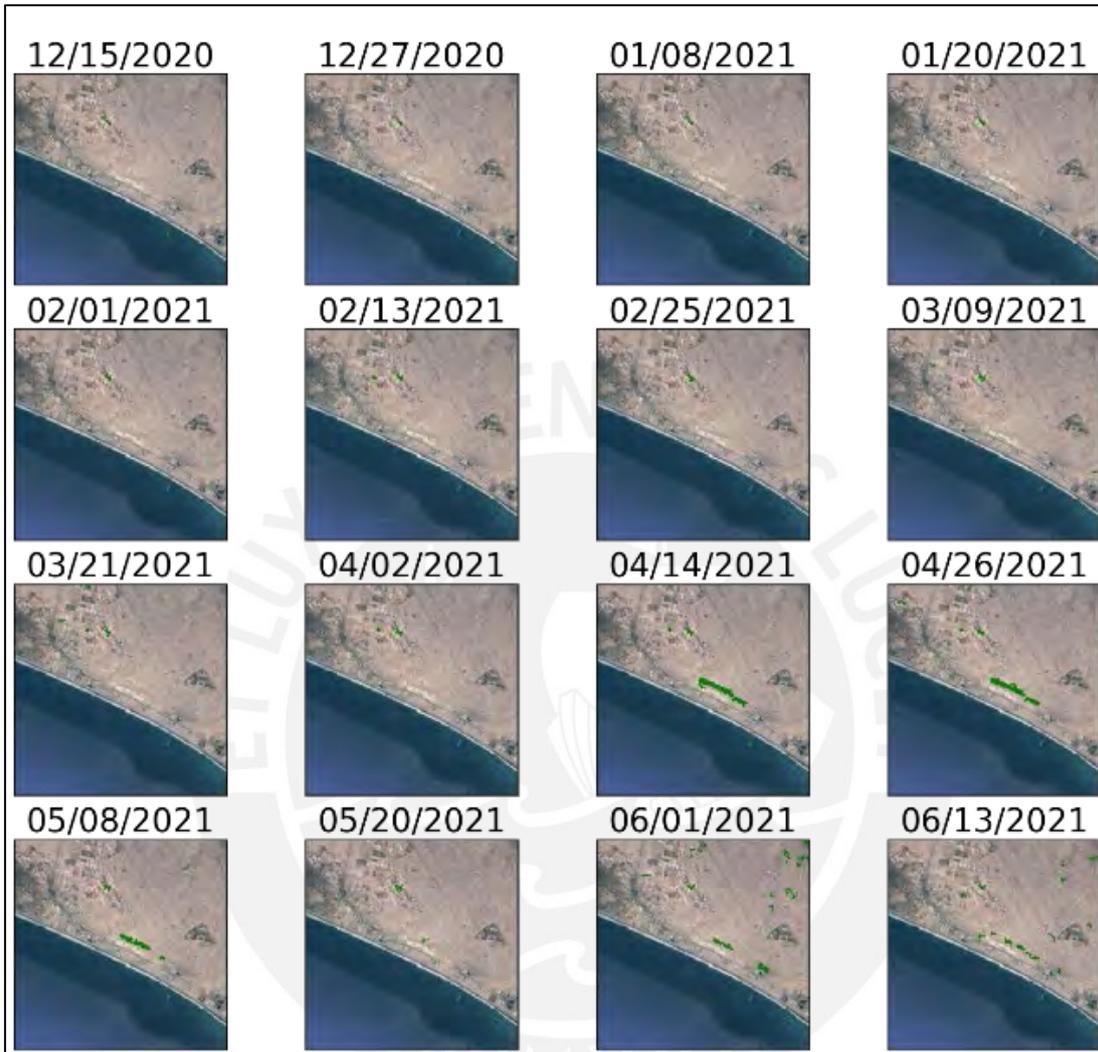


Figura 48: Resultados temporales mostrados por el algoritmo de Descenso del gradiente estocástico

K-Vecinos Cercanos

Se presenta el área predicha de la invasión en Villa Del Salvador en la Figura 49. En la Tabla 20 se muestra los resultados de las métricas de validación de los resultados.



Figura 49: (a) Área predicha por el algoritmo de K-vecinos cercanos como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 20: Evaluación de los resultados predichos por el algoritmo de K-vecinos cercanos

Métricas						
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 559262 & 1689 \\ 165 & 1384 \end{pmatrix}$	0.9967	0.4504	0.8932	0.5989	0.5974	

En la Figura 50 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

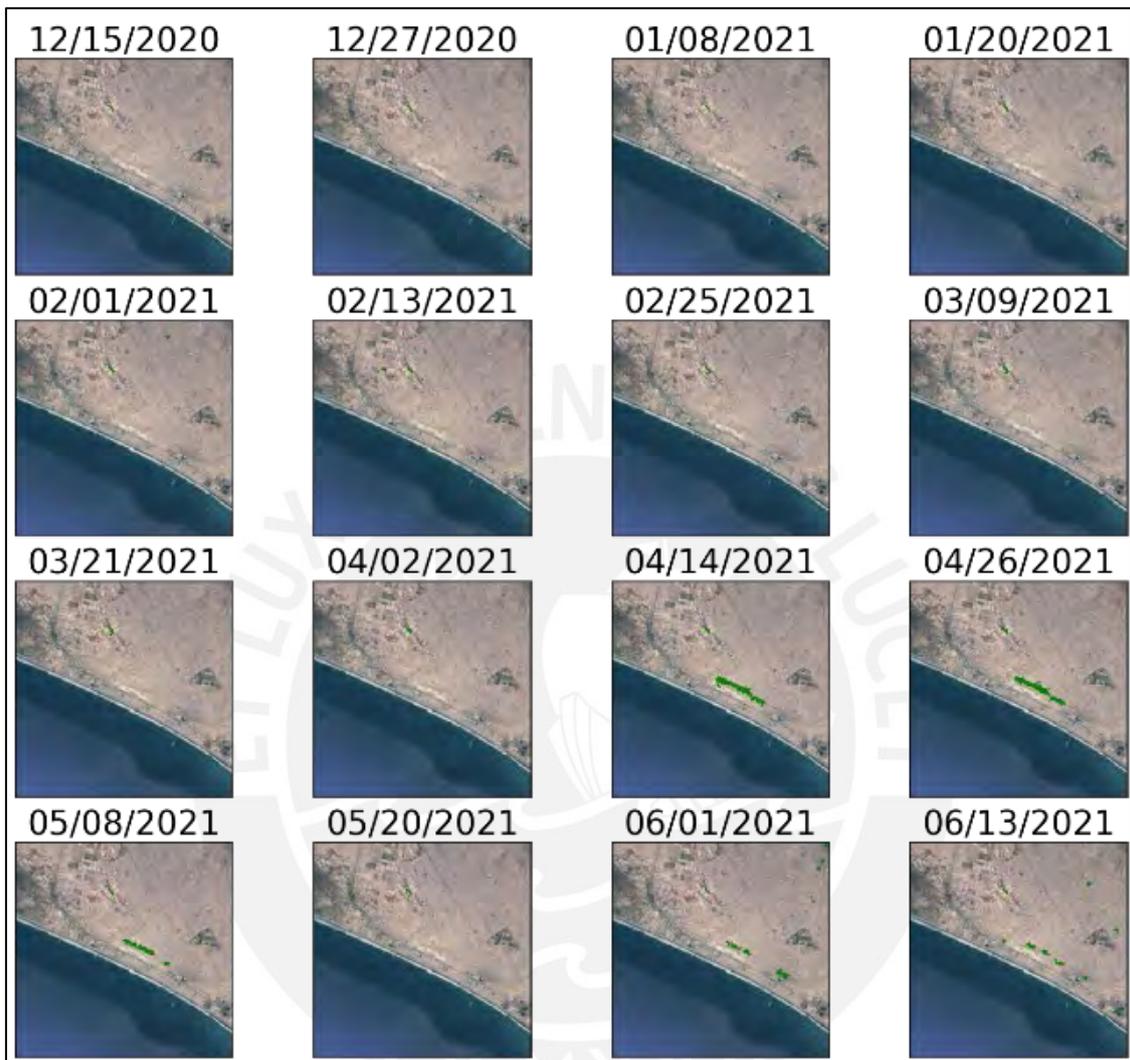


Figura 50: Resultados temporales mostrados por el algoritmo de K-vecinos cercanos

Regresión Logística

En la Figura 51 se presenta el área predicha de la invasión en Villa Del Salvador. En la Tabla 21 se muestra los resultados de las métricas de validación de los resultados.



Figura 51: (a) Área predicha por el algoritmo de Regresión logística como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 21: Evaluación de los resultados predichos por el algoritmo de Regresión logística

Métricas					
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa
$\begin{pmatrix} 558630 & 2321 \\ 123 & 1426 \end{pmatrix}$	0.9957	0.3806	0.9206	0.5385	0.5367

En la Figura 52 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

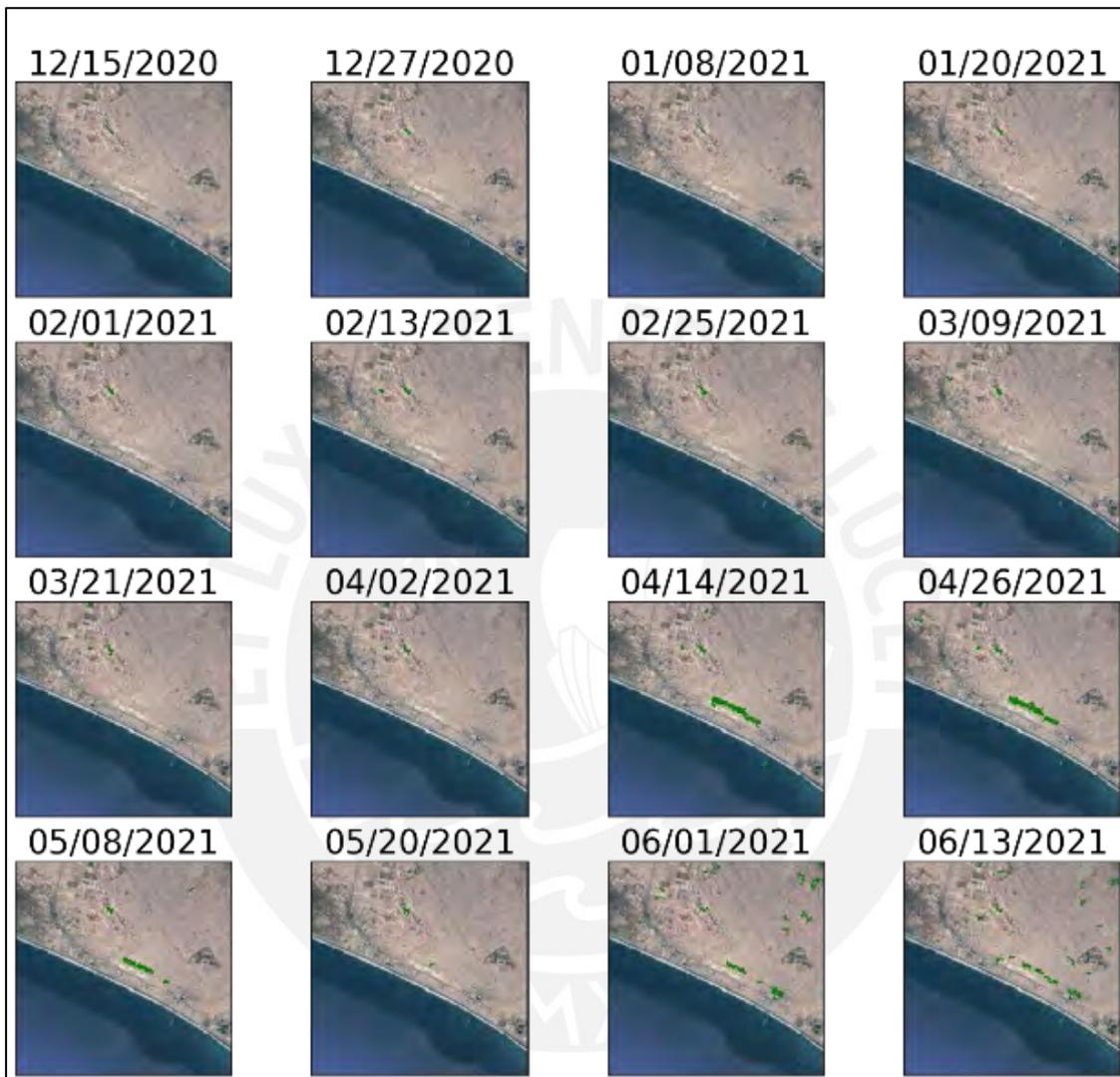


Figura 52: Resultados temporales mostrados por el algoritmo de Regresión logística

Envolvente Elíptica

Se presenta el área predicha de la invasión en Villa Del Salvador en la Figura 53. En la Tabla 22 se muestra los resultados de las métricas de validación de los resultados.



Figura 53: (a) Área predicha por el algoritmo de Envolverte elíptica como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 22: Evaluación de los resultados predichos por el algoritmo de Envolverte elíptica

Métricas						
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 557477 & 3474 \\ 381 & 1168 \end{pmatrix}$	0.9931	0.2516	0.7540	0.3773	0.3747	

En la Figura 54 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

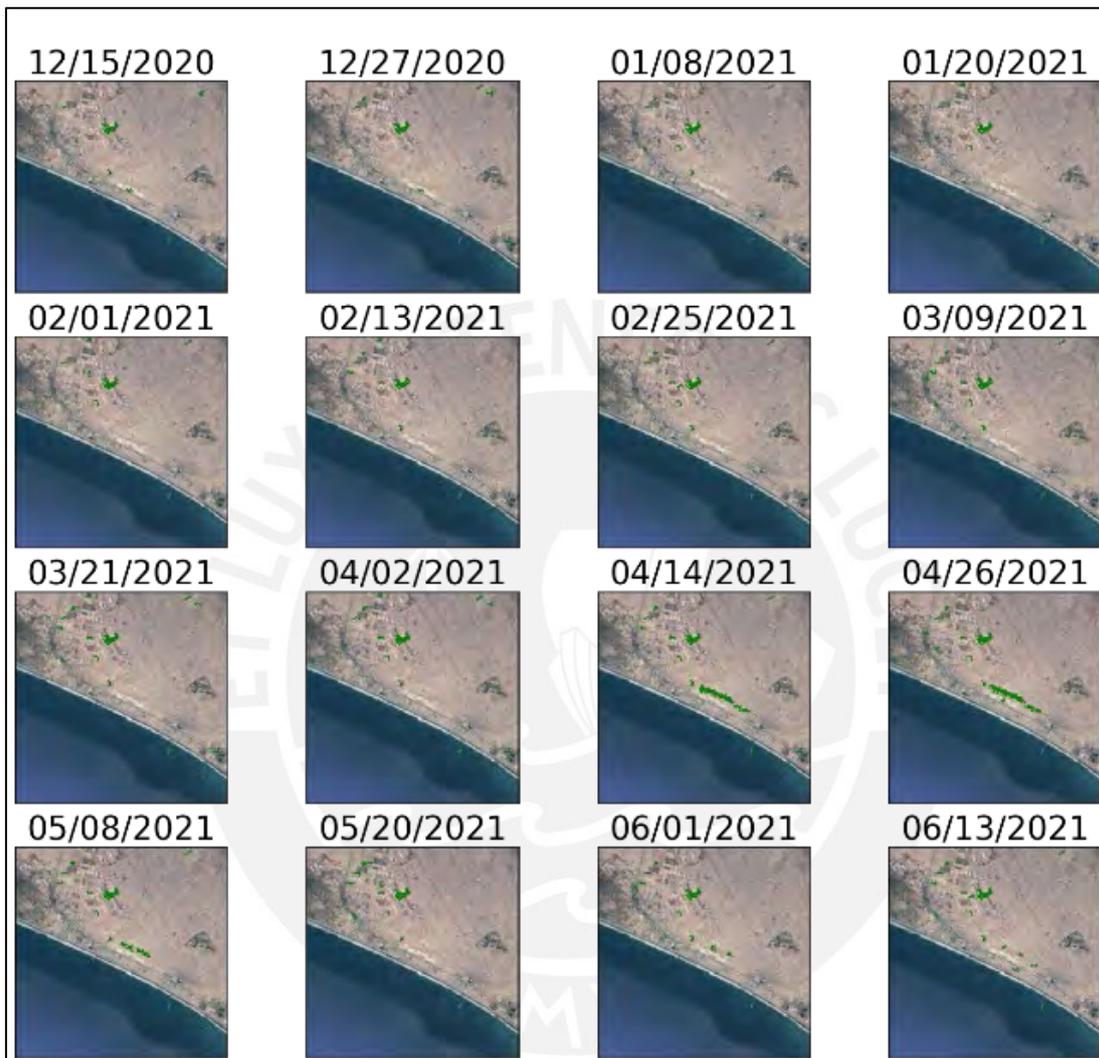


Figura 54: Resultados temporales mostrados por el algoritmo de Envolverte elíptica

Factor Local Atípico

En la Figura 55 se presenta el área predicha de la invasión en Villa Del Salvador. En la Tabla 23 se muestra los resultados de las métricas de validación de los resultados.

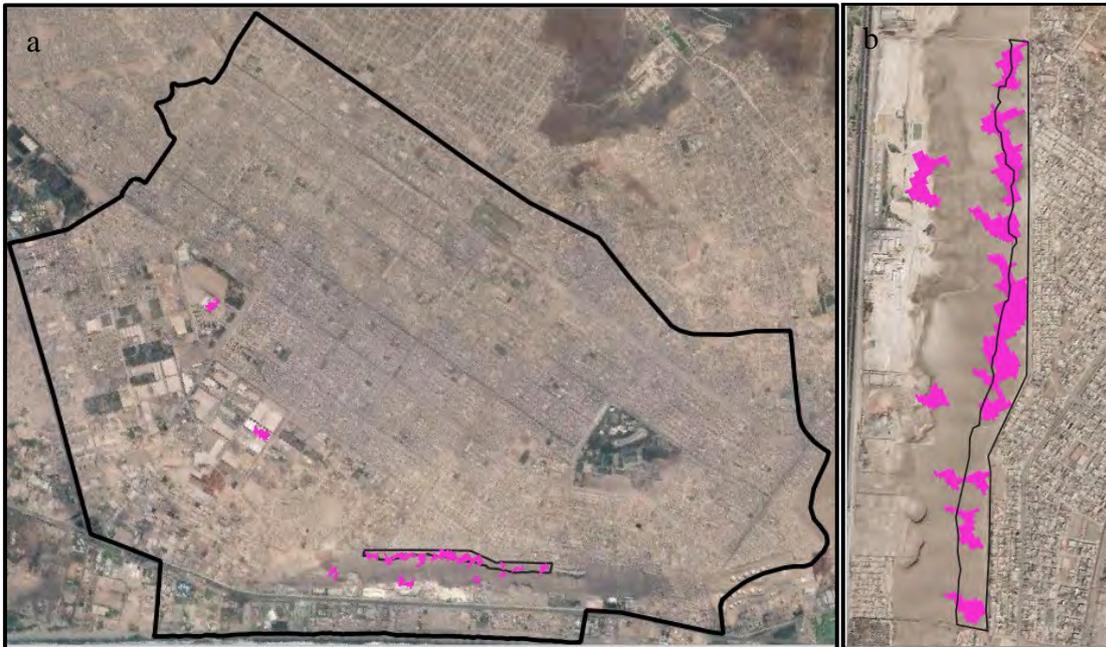


Figura 55: (a) Área predicha por el algoritmo de Factor atípico local como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 23: Evaluación de los resultados predichos por el algoritmo de Factor atípico local

Métricas						
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 559178 & 1773 \\ 727 & 822 \end{pmatrix}$	0.9956	0.3168	0.5307	0.3967	0.3946	

En la Figura 56 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

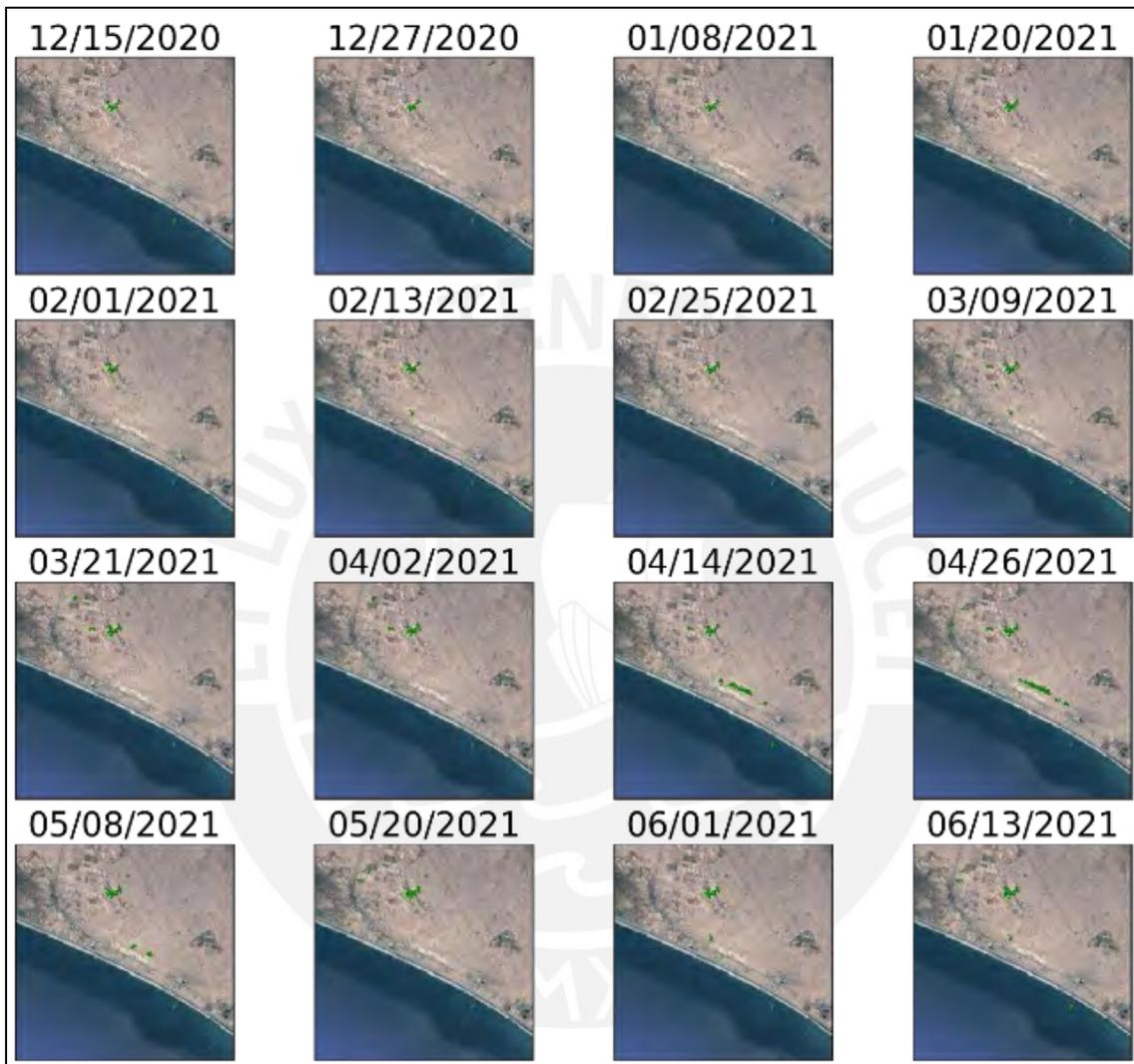


Figura 56: Resultados temporales mostrados por el algoritmo de Factor atípico local

Redes Neuronales

Se presenta el área predicha de la invasión en Villa Del Salvador en la Figura 57. En la Tabla 24 se muestra los resultados de las métricas de validación de los resultados.



Figura 57: (a) Área predicha por el algoritmo de Redes neuronales como invasiones dentro del distrito de Villa el Salvador. (b) Áreas predichas como invasiones en el sector de Lomo de Corvina

Tabla 24: Evaluación de los resultados predichos por el algoritmo de Redes neuronales

Métricas						
Matriz de confusión	Overall Accuracy	Precision	Recall	F1	Kappa	
$\begin{pmatrix} 558474 & 2477 \\ 205 & 1344 \end{pmatrix}$	0.9952	0.3517	0.8677	0.5006	0.4986	

En la Figura 58 se presenta los resultados mostrados por el algoritmo en la zona de Lomo de Corvina. Se puede apreciar un buen desempeño temporal al ubicar el periodo de las invasiones. En el caso de la invasión de Lomo de Corvina, el día 14 y 26 de abril se observan la detección de una gran área invadida y el resto de fechas no se observa mayores detecciones de áreas invadidas. Esto coincide con la realidad debido a que la invasión empezó el 12 de abril y el desalojo ocurrió el 28 de abril.

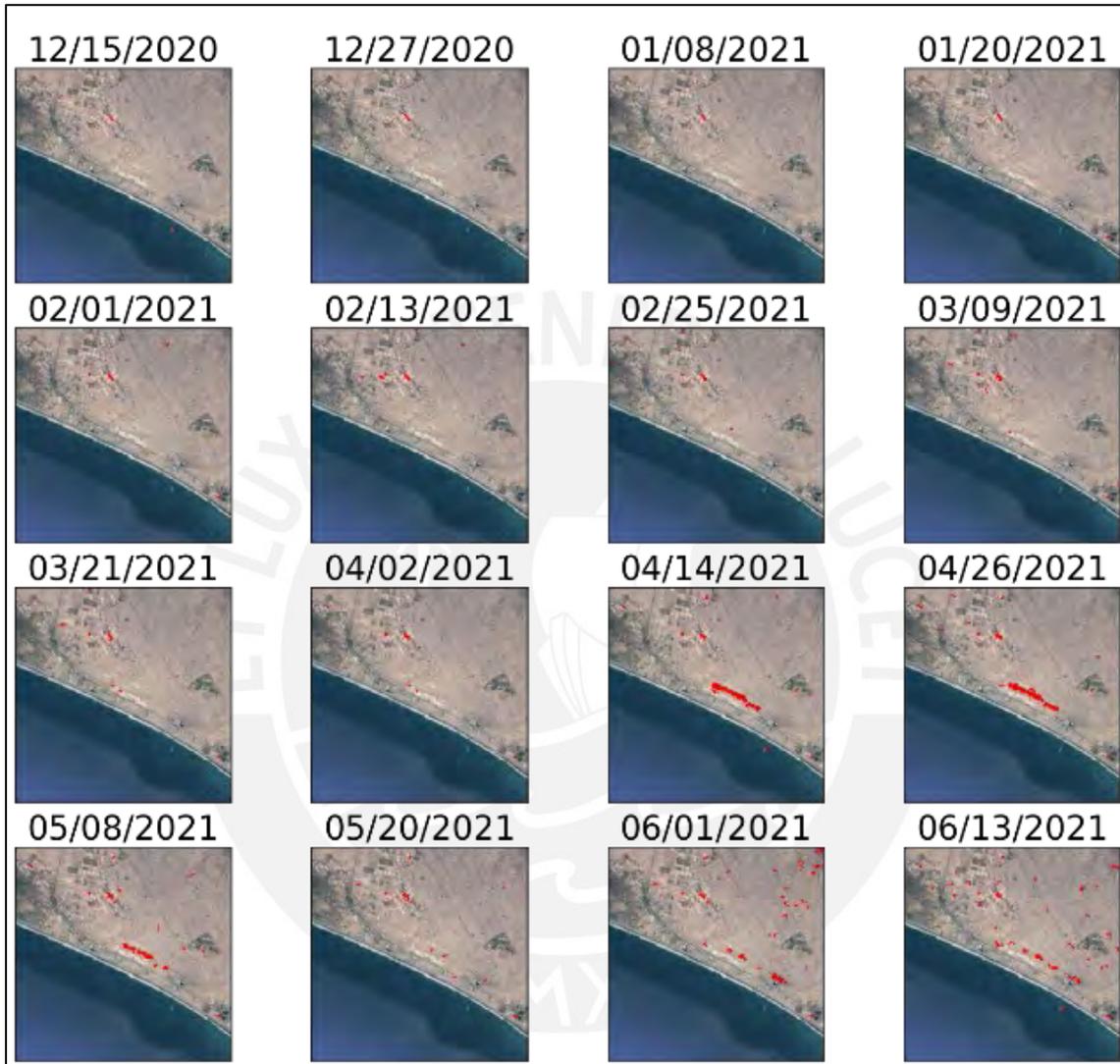


Figura 58: Resultados temporales mostrados por el algoritmo de Redes neuronales

5.4 Discusión de resultados

Como se puede apreciar en la Tabla 25, el algoritmo con mejor desempeño entre los algoritmos de clasificación supervisada en base a la métrica Kappa fue la Máquina de Soporte Vectorial (SVM) con un valor de 0.6242. Para el caso de los algoritmos semisupervisados, el algoritmo de mejor desempeño, en base a la métrica, fue el algoritmo de vecinos cercanos (NN) con un valor de 0.3946.

Tabla 25: Resumen de resultados

Resumen de resultados					
Algoritmo	Overall Accuracy	Precision	Recall	F1	Kappa
Decision Tree	0.9968	0.4526	0.8502	0.5907	0.5892
Gaussian Naïve Bayes	0.9965	0.4321	0.9032	0.5845	0.5829
Random Forest	0.9964	0.4274	0.9374	0.5871	0.5856
SVM	0.9971	0.4881	0.8709	0.6256	0.6242
SGD	0.9956	0.3797	0.9180	0.5372	0.5354
KNN	0.9967	0.4504	0.8932	0.5989	0.5974
LR	0.9957	0.3806	0.9206	0.5385	0.5367
Elliptic Envelop	0.9931	0.2516	0.7540	0.3773	0.3747
Local Outlier Factor	0.9956	0.3168	0.5307	0.3967	0.3946
Red Neuronal	0.9952	0.3517	0.8677	0.5006	0.4986

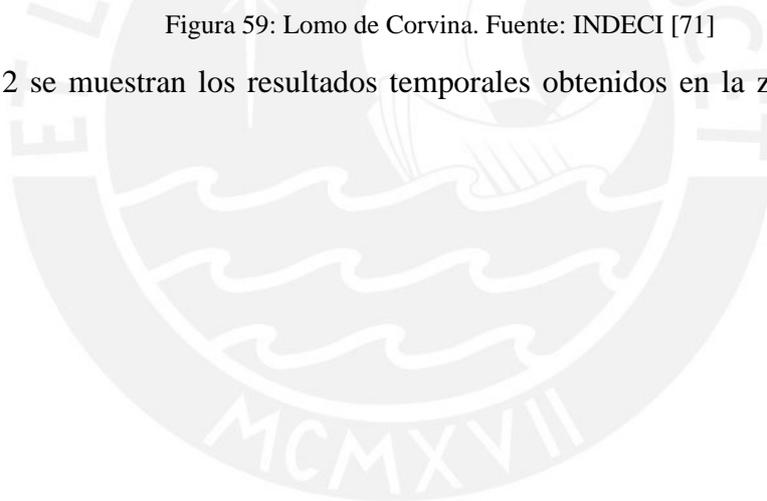
Finalmente se puede apreciar que en general los algoritmos de clasificación supervisada tuvieron mejor desempeño a comparación de los algoritmos semisupervisados. Es importante recalcar, también, que todos los algoritmos fueron capaces de predecir la aparición de una invasión. Es decir, hay una máxima precisión con respecto a la resolución temporal. Sin embargo, la precisión espacial es mediana a baja. La principal razón de estas imprecisiones se debe a la topografía accidentada en la zona

de estudio como se observa en la Figura 59. Dado que el principal objetivo es poder emitir una alerta de invasión, para su posterior confirmación visual, podemos aseverar que el uso de las imágenes de radar representa un gran potencial para el monitoreo continuo de zonas de alto riesgo con un costo mínimo.



Figura 59: Lomo de Corvina. Fuente: INDECI [71]

En el Anexo 2 se muestran los resultados temporales obtenidos en la zona del Morro Solar.



Capítulo 6

Conclusiones

6.1 Conclusiones

Los algoritmos aplicados sobreestiman el área invadida debido a las complejas proyecciones y transformaciones geométricas de las imágenes SAR. Como se puede apreciar en la Tabla 25 las precisiones obtenidas por los diversos algoritmos resultaron menores al 50%. Estos resultados no se deben por un mal desempeño de los algoritmos o un inadecuado entrenamiento, se dan debido a las distorsiones geométricas que poseen las imágenes de radar como el *layover* por lo que sobrestiman las áreas invadidas. No obstante, los algoritmos estiman en promedio más del 80% del área invadida lo cual es indicado por la métrica *recall*. Además, los resultados temporales muestran que los algoritmos detectan las zonas invadidas en las fechas de las invasiones. Por lo tanto, se puede emplear el análisis de píxeles empleando técnicas supervisada y semi-supervisada para el monitoreo de las invasiones con alta precisión temporal y referencia de su ubicación.

La aplicación de técnicas de clasificación supervisada muestra altos valores de exactitud en la determinación de las zonas invadidas y valores de precisión relativamente bajos. No obstante, técnicas semi-supervisadas a pesar de no obtener una precisión como el conjunto de técnicas anteriores, también son prometedoras para la identificación general de invasiones debido a que para el entrenamiento solo se requieren un conjunto de imágenes temporales de áreas detectadas como no invadidas. En cambio, las técnicas supervisadas requieren para su entrenamiento dos conjuntos de imágenes temporales de áreas detectadas como no invadidas e invadidas.

6.2 Recomendaciones

En general, en cualquier área de estudio se recomienda emplear algoritmos de aprendizaje semi-supervisado, los cuales deben calibrarse empleando data local, debido a que se adaptan perfectamente a las condiciones locales (topografía) y o requieren data de áreas invadidas localmente. Adicionalmente con el objetivo de asegurar una cobertura completa del área de estudio, recomendamos realizar el análisis dos veces, empleando imágenes tomadas en trayectoria ascendente en un análisis e imágenes en trayectoria descendente en la otra. Con esto aseguraremos que una menor superficie terrestre se vea afectada por las distorsiones geométricas.

Se recomienda emplear la metodología y monitorear las cuadrículas que se presentan en la Figura 60, en base a los suelos no aptos para la construcción. En caso de emplear computación paralela evaluar directamente toda la imagen sin necesidad de extraer cuadrículas.

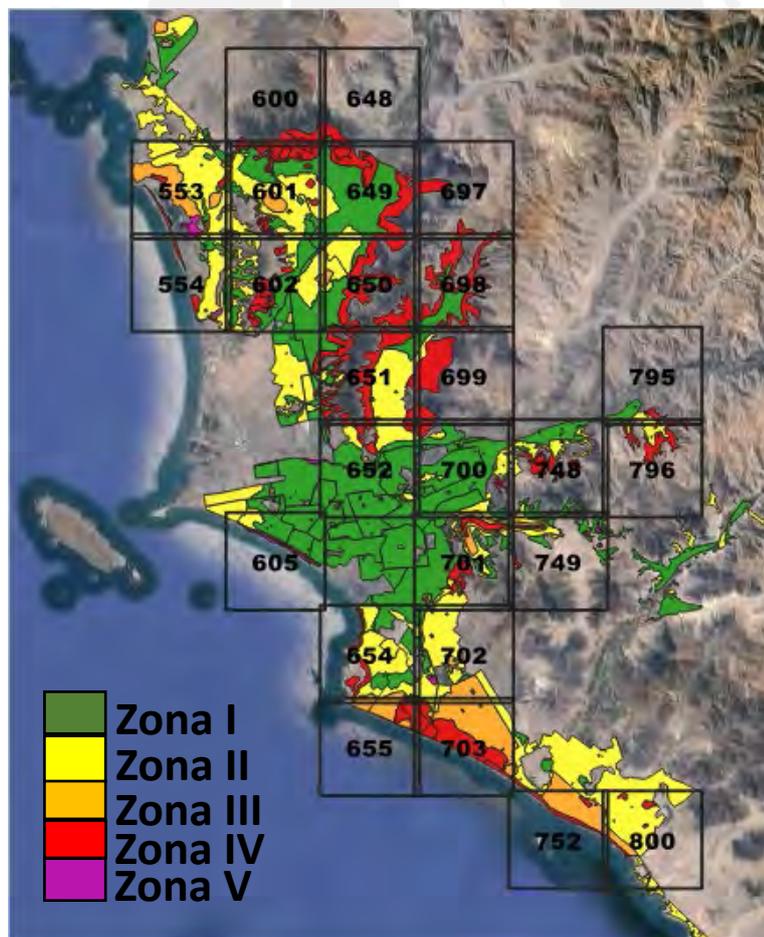


Figura 60: Sectores más crítico a monitorear según el tipo de zona

Nota: Adaptado de CISMID [68,69]

Referencias

- [1] ONU, « Departamento de asuntos económicos y sociales,» 2018. [En línea].
- [2] UN-Habitat, «Informal Settlements,» *UN-Habitat*, pp. 1-8, 2015.
- [3] I. Müller, H. Taubenböck, M. Kuffer y M. Wurm, «Misperceptions of Predominant Slum Locations? Spatial Analysis of Slum Locations in Terms of Topography Based on Earth Observation Data,» *remote sensing*, vol. 12, n° 15, p. 2474, 2020.
- [4] SENCICO, *Norma E.030 Diseño Sismoresistente*, Lima, 2018.
- [5] H. Tavera, *Evaluación del peligro asociado a los sismos y efectos secundarios en Perú*, Lima, 2014.
- [6] Z. Assarkhaniki, S. Sabri y A. Rajabifard, «Using open data to detect the structure and pattern of informal settlements: an outset to support inclusive SDGs' achievement,» *Big Earth Data*, vol. 5, n° 4, pp. 497-526, 2021.
- [7] C. Kuenzer, S. Dech y W. Wagner, *Remote sensing time series*, Springer, 2015.
- [8] ONU, «Transformar nuestro mundo: la Agenda 2030 para el desarrollo sostenible,» 2015.
- [9] P. Lamperein, «Aplicaciones de imágenes de RADAR, en la generación de información para la mitigación de riesgos naturales,» *Dialogo Andino*, pp. 35-43, 2004.
- [10] F. Deng, M. Rodgers, S. Xie, T. H. Dixon, S. Charbonnier, E. A. Gallant, C. M. L. Vélez, M. Ordoñez, R. Malservisi, N. K. Voss y J. A. Richardson, «High-resolution DEM generation from spaceborne and terrestrial remote sensing data for improved volcano hazard assessment. A case study at Nevado del Ruiz, Colombia,» *Remote sensing of environment*, vol. 233, 2019.
- [11] A. D. Vibhute, K. V. Kale, R. K. Dhumal y S. C. Mehrotra, «Soil type classification and mapping using hyperspectral remote sensing data,» *2015 International Conference on Man and Machine Interfacing (MAMI)*, pp. 1-4, 2015.

- [12] N. Kureel, J. Sarup y S. Matin, «Modelling vegetation health and stress using hyperspectral remote sensing data,» *Modeling Earth Systems and Environment*, vol. 8, pp. 733-748, 2022.
- [13] X. Li, B. Liu, G. Zheng, Y. Ren, S. Zhang, Y. Liu, L. Gao, Y. Liu, B. Zhang y F. Wang, «Deep-learning-based information mining from ocean remote-sensing imagery,» *National Science Review*, vol. 7, pp. 1584-1605, October 2020.
- [14] L. Moya, C. Geiß, M. Hashimoto, E. Mas, S. Koshimura and G. Strunz, "Disaster intensity-based selection of training samples for remote sensing building damage classification," *IEEE transactions on geoscience and remote sensing*, vol. 59, no. 10, pp. 8288-8304, 2021.
- [15] M. Kuffer, K. Pfeffer y R. Sliuzas, «Slums from Space—15 Years of Slum Mapping Using Remote Sensing,» *Remoting sensing*, vol. 8, n° 6, p. 455, 2016.
- [16] SCOPUS, «SCOPUS,» [En línea]. Available: <https://www.scopus.com/>. [Último acceso: 12 enero 2023].
- [17] A. Prado Naccha y M. Colonio Roque, «Evaluación de la política nacional de gestión de riesgo de desastres en el Perú,» 2016.
- [18] C. Wiese, J. Miyahiro y R. Marcés, *Desigualdad urbana en lima metropolitana*, 2016.
- [19] P. Muñoz y L. Rodriguez, «Viviendo en Laderas. Una política urbana/pública en la periferie de Lima (Parte 1),» 28 04 2015. [En línea]. Available: <http://www2.ual.es/RedURBS/BlogURBS/vivienda-en-laderas-parte-2/>. [Último acceso: 03 08 2022].
- [20] INDECI, Compendio estadístico gestión reactiva 2022, Lima, 2022.
- [21] INDECI, Información estadística de emergencias y daños, periodo 2003 al 2019, Lima, 2020.
- [22] INDECI-DIPRE, «Escenario sísmico para Lima Metropolitana y Callao: Sismo 8.8Mw,» 2017.
- [23] M. Matsuoka, S. Mito, S. Midorikawa, H. Miura, L. G. Quiroz, Y. Maruyama y M. Estrada, «Development of building inventory data and earthquake damage estimation in Lima, Peru for future earthquakes,» *Journal of disaster research*, vol. 9, n° 6, pp. 1032-1041, 2014.
- [24] Municipalidad de Lima, «Plan de prevención de riesgo y reducción del desastre de Lima Metropolitana 2019-2022,» Lima, 2019.
- [25] ASF, «ASF Data Search Vertex,» [En línea]. Available: <https://search.asf.alaska.edu/#/>.

- [26] Copernicus, «Copernicus Open Access Hub,» [En línea]. Available: <https://scihub.copernicus.eu/dhus/#/home>.
- [27] *Decreto Supremo N°038-2021-PCM.Normas Legales*, 2021.
- [28] D. Torres Obrgón, D. Perleche Ugás y A. Aiquipa Zavala, La producción del espacio urbano en Lima Metropolitana y el Callao: Entre las informalidades y la regulación (1961-2020), Primera ed., Lima: Urbes Lab, 2021.
- [29] M. Matos, «Las barriadas de Lima, 1957,» Instituto de Estudios Peruanos, Lima, 1977.
- [30] E. Vera, «El cruel negocio del tráfico de terrenos: un lucrativo delito se expande,» *El Comercio*, 25 mayo 2018.
- [31] N. Nilsson, *Introduccion to machine learning*, Standford University, 1998.
- [32] K. Heng Lai, D. Zha, J. Xu, Y. Zhao, G. Wang y X. Hu, «Revisiting Time Series Outlier Detection: Definitions and Benchmarks,» *NeurIPS*, 2021.
- [33] W. Bakker, W. Feriga, A. Gieske, B. Gorte, K. Grabmaier, C. Hecker, J. Horn, G. Huurneman, L. Janssen, K. Norman, F. Van der Meer, G. Parodi, C. Pohl, C. Reeves, F. Van Ruitenbeek, E. Schetselaar, K. Tempfli, M. Weir, E. Westinga y T. Woldai, *Principles of Remote Sensing*, Enschede: The International Institute for Geo-Information Science and Earth Observation, 2009.
- [34] S. Khorram, F. H. Koch, C. F. V. d. Wiele y S. A. C. Nelson, *Remote Sensing*, Raleigh: Springer, 2012.
- [35] E. A. Ustinov, *Sensitivity analysis in remote sensing*, Springer, 2015.
- [36] IGN, *El programa Copernicus aplicado a la producción y gestión de la información geoespacial*, Madrid, 2022.
- [37] T. Esch, H. Taubenbock, A. Roth, W. Heldens, A. Felbier, M. Thiel, M. Schmidt, A. Muller y S. Dech, «TanDEM-X mission-new perspectives for the inventory and monitoring of global settlement patterns.,» *Remote Sensing*, 2012.
- [38] N. J. Kraff, M. Wurm y H. Taubenböck, «The dynamics of poor urban areas - analyzing morphologic transformations across the globe using Earth observation data,» *Cities*, 2020.
- [39] M. Pesaresi, H. Guo, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann y T. Kemper, «A global human settlement layer from optical HR/VHR RS data: Concept and first results.,» *Remote Sensing*, 2013.
- [40] S. Shekhar, «Improving the slum planning through geospatial decision support system.,» *Remote Sensing*, pp. 76-90, 2014.

- [41] P. Zavala, «Aplicaciones de teledetección en el seguimiento del peligro sísmico,» *Diálogo Andino*, pp. 46-52, 2004.
- [42] C. Wang, Y. Zhang, T. Xie, L. C. Guo, J. Li y F. Shi, «A Detection Method for Collapsed Buildings Combining Post-Earthquake High-Resolution Optical and Synthetic Aperture Radar Images,» *remote sensing*, 2022.
- [43] S. Khorram, C. F. V. d. Wiele, F. H. Koch, S. A. C. Nelson y M. D. Potts, *Principles of applied remote sensing*, Springer, 2016.
- [44] J. A. Richards, *Remote sensing digital image analysis*, Quinta ed., Berlín: Springer, 2013.
- [45] J. A. Richards, *Remote sensing with imaging radar*, Springer, 2009.
- [46] ESA, «1st Advanced Training Course on Land Remote Sensing 2007,» 07 Setiembre 2007. [En línea]. Available: <https://earth.esa.int/eogateway/events/advanced-training-course-on-land-remote-sensing>. [Último acceso: 23 01 24].
- [47] H. Taubenbock y N. Kraff, «The physical face of slums: A structural comparison of slums in Mumbai, India, based on remotely sensed data,» *J. Hous. Built Environ*, pp. 15-38, 2014.
- [48] K. K. Owen y D. Wong, «An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics,» *Appl. Geogr*, pp. 107-118, 2013.
- [49] D. Kohli, M. Kuffer y C. M. Gevaert, «The Generic Slum Ontology: Can a Global Slum Repository be created?,» *2019 Joint Urban Remote Sensing Event (JURSE)*, pp. 1-4, 2019.
- [50] M. Kuffer, D. R. Thomson, G. Boo, R. Mahabir, T. Grippa, S. Vanhuyse, R. Engstrom, R. Ndugwa, J. Makau, E. Darin, J. P. d. Albuquerque y C. Kabaria, «The Role of Earth Observation in an Integrated Deprived Area Mapping “System” for Low-to-Middle Income Countries,» *Remote sensing*, vol. 12, 2020.
- [51] A. Fallatah, S. Jones, L. Wallace y D. Mitchell, «Combining Object-Based Machine Learning with Long-Term Time-Series Analysis for Informal Settlement Identification,» *MDPI*, 2022.
- [52] P. Hofmann, J. Strobl, T. Blaschke y H. Kux, «Detecting informal settlements from Quickbird data in Rio de Janeiro using an object based approach,» *ISPRS*.
- [53] L. Ella, F. Van Den Bergh, B. Van Wyk y M. Van Wyk, «A comparison of texture feature algorithms for urban settlement classification,» *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2008.
- [54] I. Tingzon, N. Dejito, R. Avell Flore, R. De Guzman, L. Carvajal, K. Zapata Erazo, I. E. Contreras Cala, J. R. D. Villaveces y R. Ghani, «Mapping New

Informal Settlements using Machine Learning and Time Series Satellite Images: An Application in the Venezuelan Migration Crisis,» *Arxiv*.

- [55] B. Graham-Hansen, P. Helber, I. Varatharajan, F. Azam, A. Coca-Castro, V. Kopackova and P. Bilinski, "Mapping Informal Settlements in Developing Countries using Machine Learning and Low Resolution Multi-spectral Data," *ACM*, 2019.
- [56] C. Geiß, Y. Zhu, C. Qiu, L. Mou, X. Xiang Zhu y H. Taubenböck, «Deep Relearning in the Geospatial Domain for Semantic Remote Sensing Image Segmentation,» *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [57] R. Ahmed Ansari, R. Malhontra y K. Mohan Buddhiraju, «Identifying Informal Settlements Using Contourlet Assisted Deep Learning,» *Sensors*, 2020.
- [58] P. Helber, B. Bischke y A. Dengel, «Artificial Intelligence in Earth Observation: Application in the Copernicus Programme».
- [59] J. Ajayakumar, A. Curtis, V. Rouzier, J. William Pape, S. Bempha, M. Taifur Alum, M. Alam, M. Rashid, A. Ali y J. Glenn Morris, «Exploring convolutional neural networks and spatial video for on-the-ground mapping in informal settlements,» *International Journal of Health Geographics*, 2021.
- [60] Y. Ban, *Multitemporal remote sensing*, Springer, 2016.
- [61] C. Kuenzer, M. Ottinger, M. Wegmann, H. Guo, C. Wang, J. Zhang, S. Dech y M. Wikelski, «Earth observation satellite sensors for biodiversity monitoring: potentials and bottlenecks,» *International Journal of Remote Sensing*, vol. 35, nº 18, pp. 6599-6647, 2014.
- [62] A. Ambugadu M. y N. Hosni, «Identifying informal settlements in satellite images for sustainable urban planning: a systematic review of methods available,» *Journal of Information System and Technology Management*, pp. 102-119, 2022.
- [63] A. M. Fallatah, «Mapping Informal Settlements in a Middle Eastern Environment Using Remote Sensing Techniques,» RMIT, 2020.
- [64] K. Klemmer, G. Yeboah, J. Port de Albuquerque y S. Jarvis, «Population mapping in informal settlements with high-resolution satellite imagery and equitable ground-truth,» *ML-IRL workshop at ICLR*, 2020.
- [65] O. Mboga y Nicholus, «Detection of informal settlements from VHR satellite images using convolutional neural networks,» Netherlands, 2017.
- [66] S. Mason y C. Fraser, «Image sources for informal settlement management.,» *Photogramm*, pp. 313-330, 1998.
- [67] L. Moya, F. Garcia, C. Gonzales, M. Diaz, C. Zavala and M. Estrada, "Brief communication: Radar images for monitoring informal urban settlements in

vulnerable zones in Lima, Peru," *Natural Hazards Earth System Sciences*, pp. 65-70, 2022.

- [68] Mahendra, Shivakumar y Praveen, «Pixel-based Classification of Multispectral Remotely Sensed Data Using Support Vector Machine Classifier,» *International journal of innovative research in electrical, electronics, instrumentation and control engineering*, vol. 3, pp. 94-98, 2015.
- [69] CISMID, «Microzonificación sísmica del distrito de Villa el Salvador,» Lima, 2011.
- [70] CISMID, «Microzonificación sísmica del distrito de Chorrillos,» 2010.
- [71] INDECI, «Peligro de deslizamiento del cerro de Lomo de Corvina dist. Villa el Salvador-Lima,» Lima, 2014.
- [72] J. A. Richards, *Radio Wave Propagation*, Berlin: Springer, 2008.



Anexo 1: Mapas de riesgo

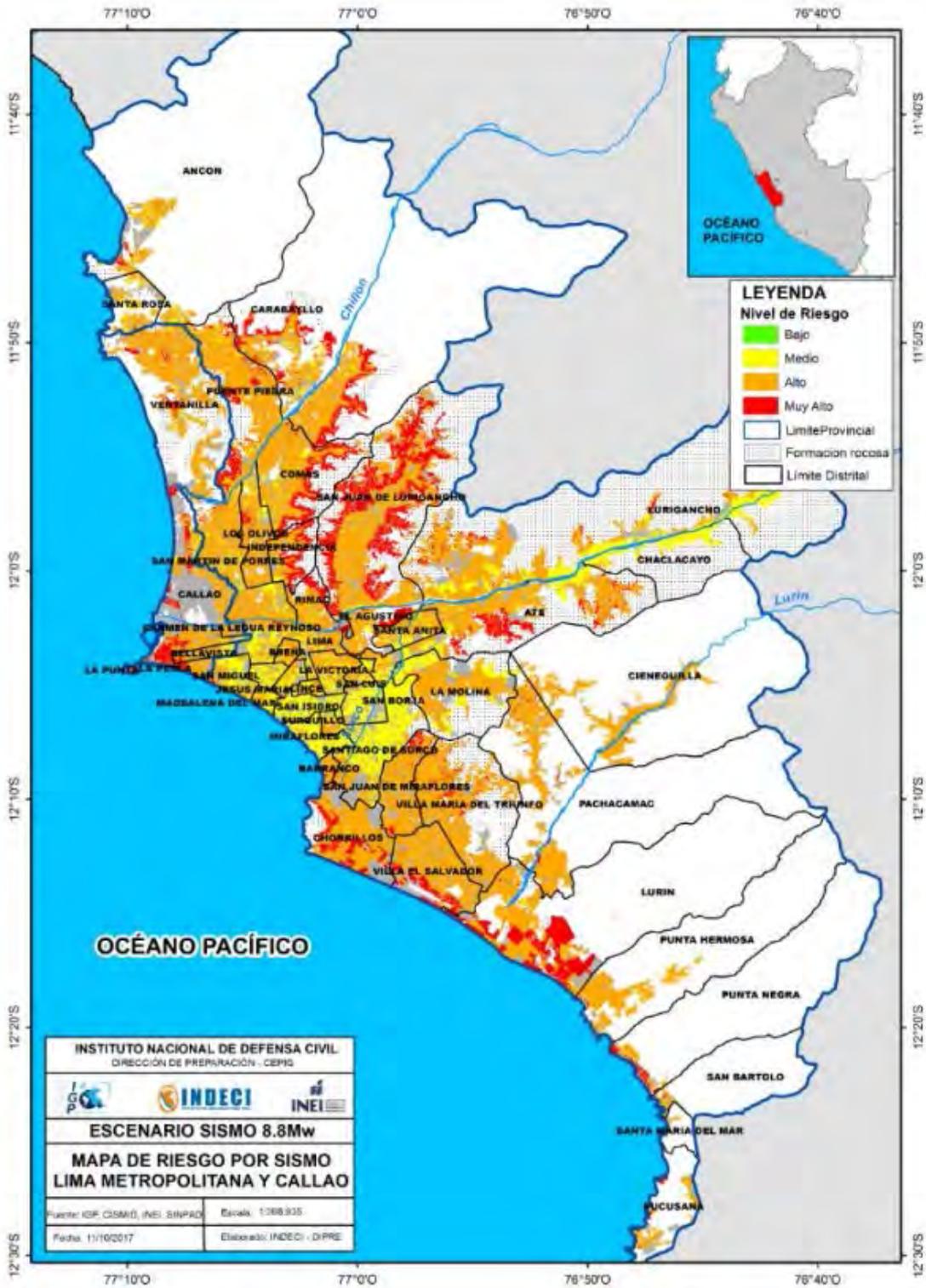


Figura 61: Mapa de riesgos por sismo en Lima Metropolitana y Callao. Fuente: INDECI-DIPRE

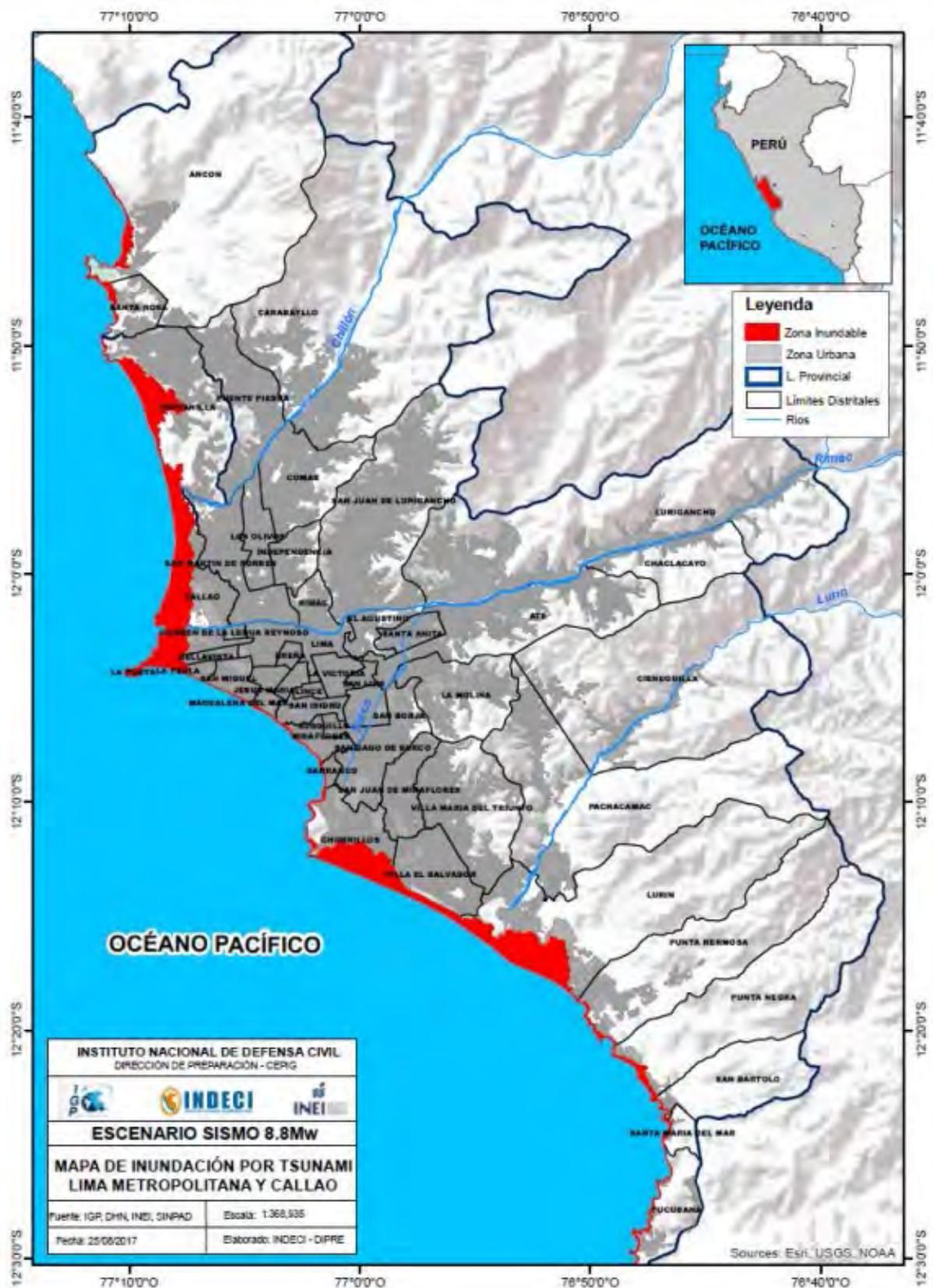


Figura 62: Mapa de inundación por tsunami en Lima metropolitana y Callao. Fuente: INDECI-DIPRE

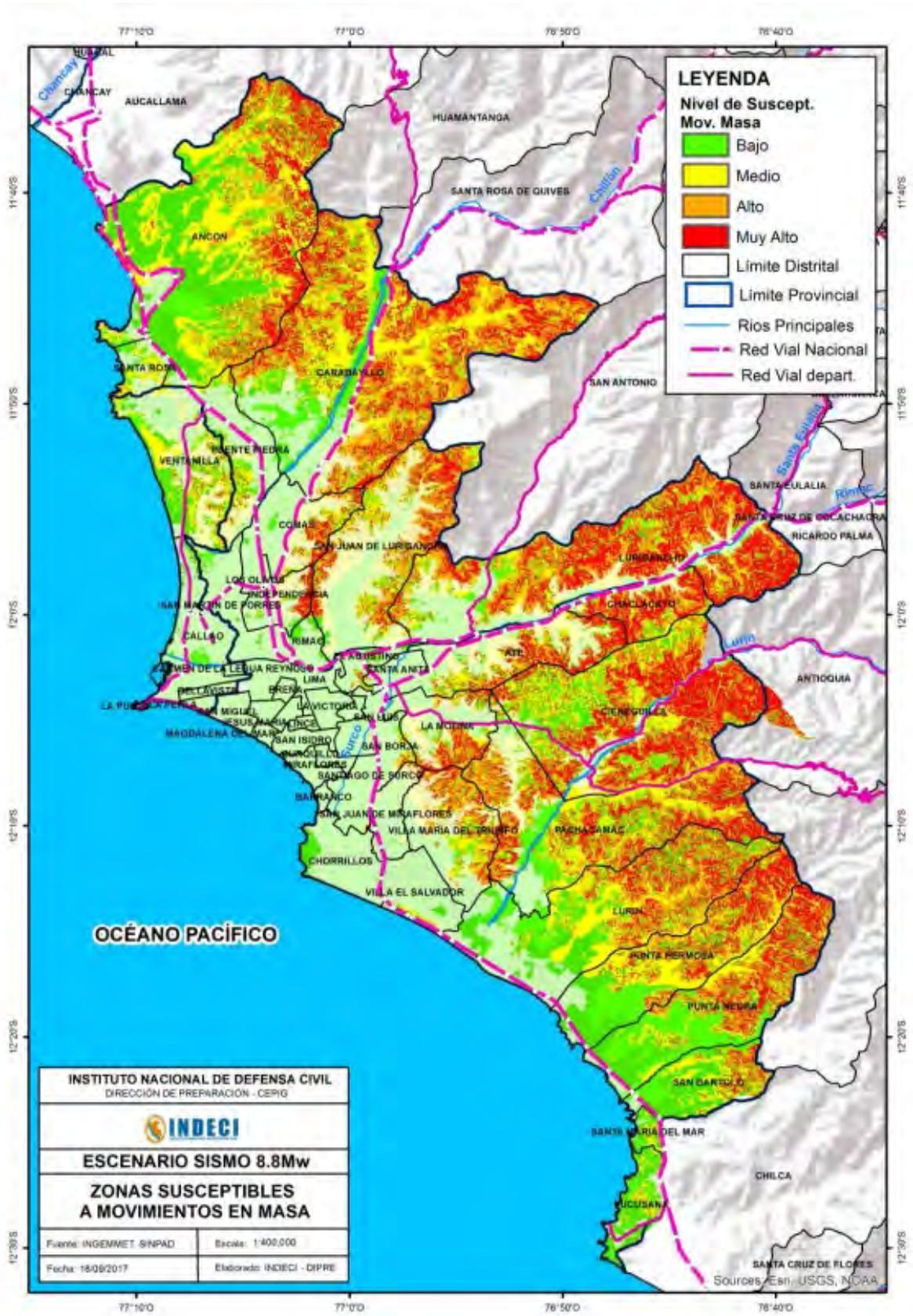


Figura 63: Zonas susceptibles a movimientos de masas. Fuente: INDECI-DIPRE

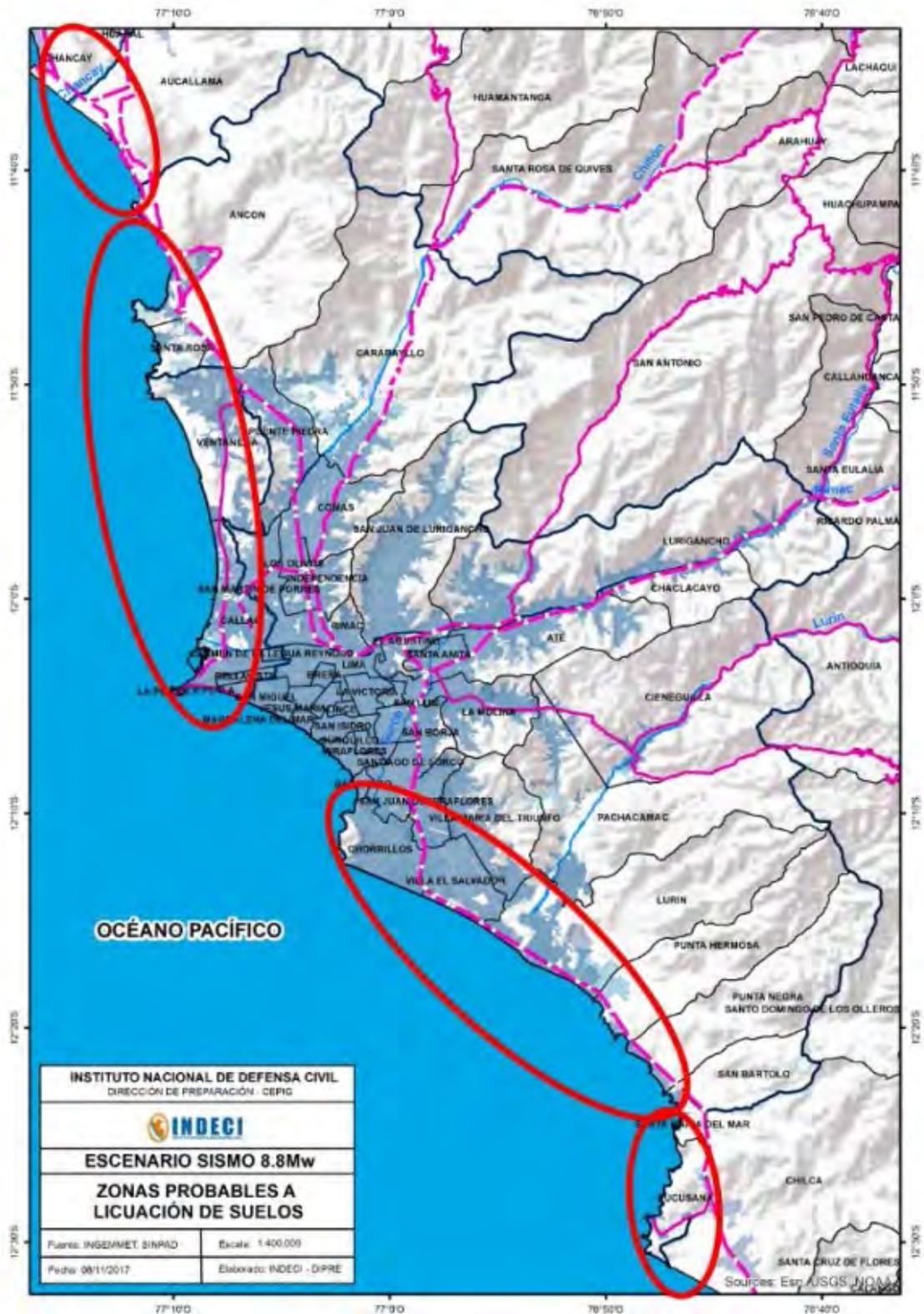


Figura 64: Zonas probables a licuación de suelos. Fuente: INDECI-DIPRE

Anexo 2: Resultados temporales en el Morro Solar

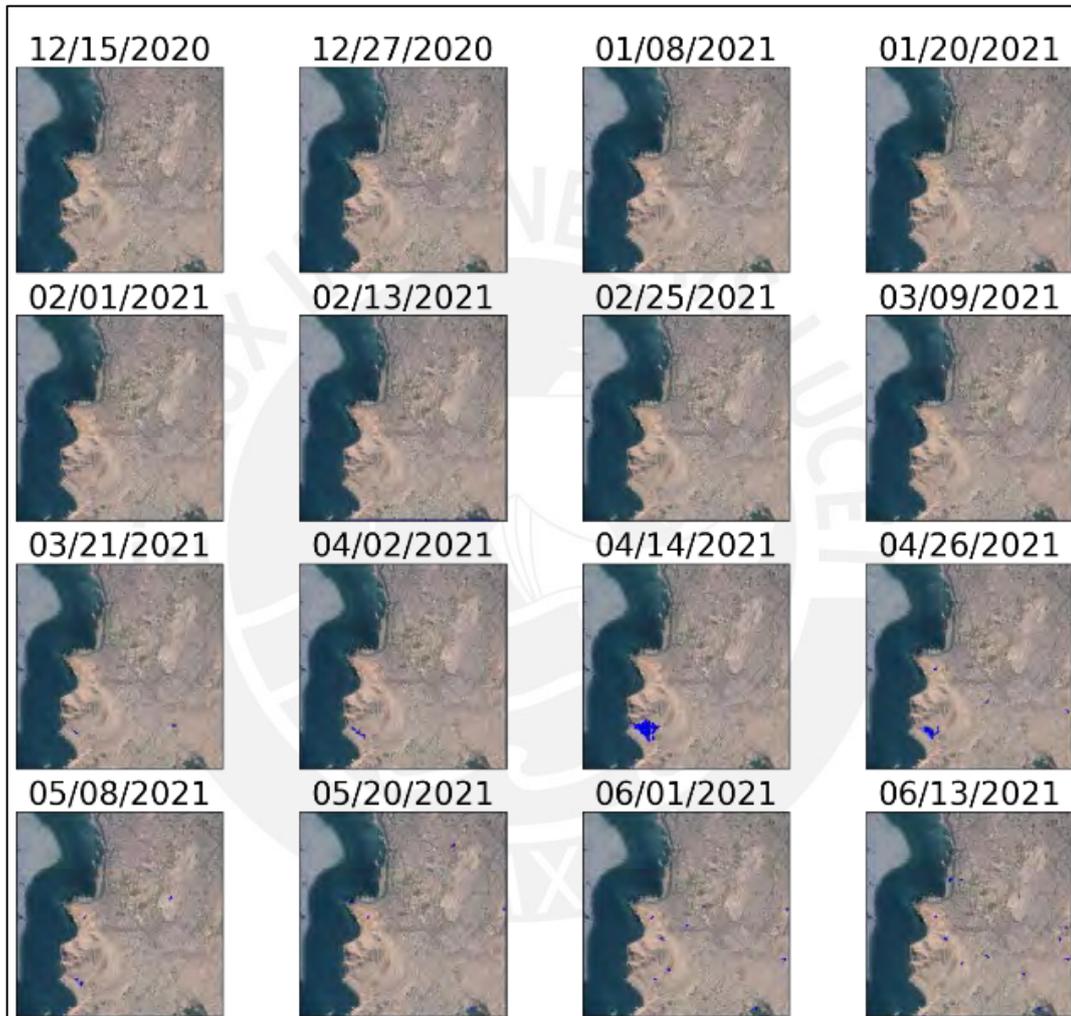


Figura 65: Resultados temporales mostrados por el Árbol de decisión

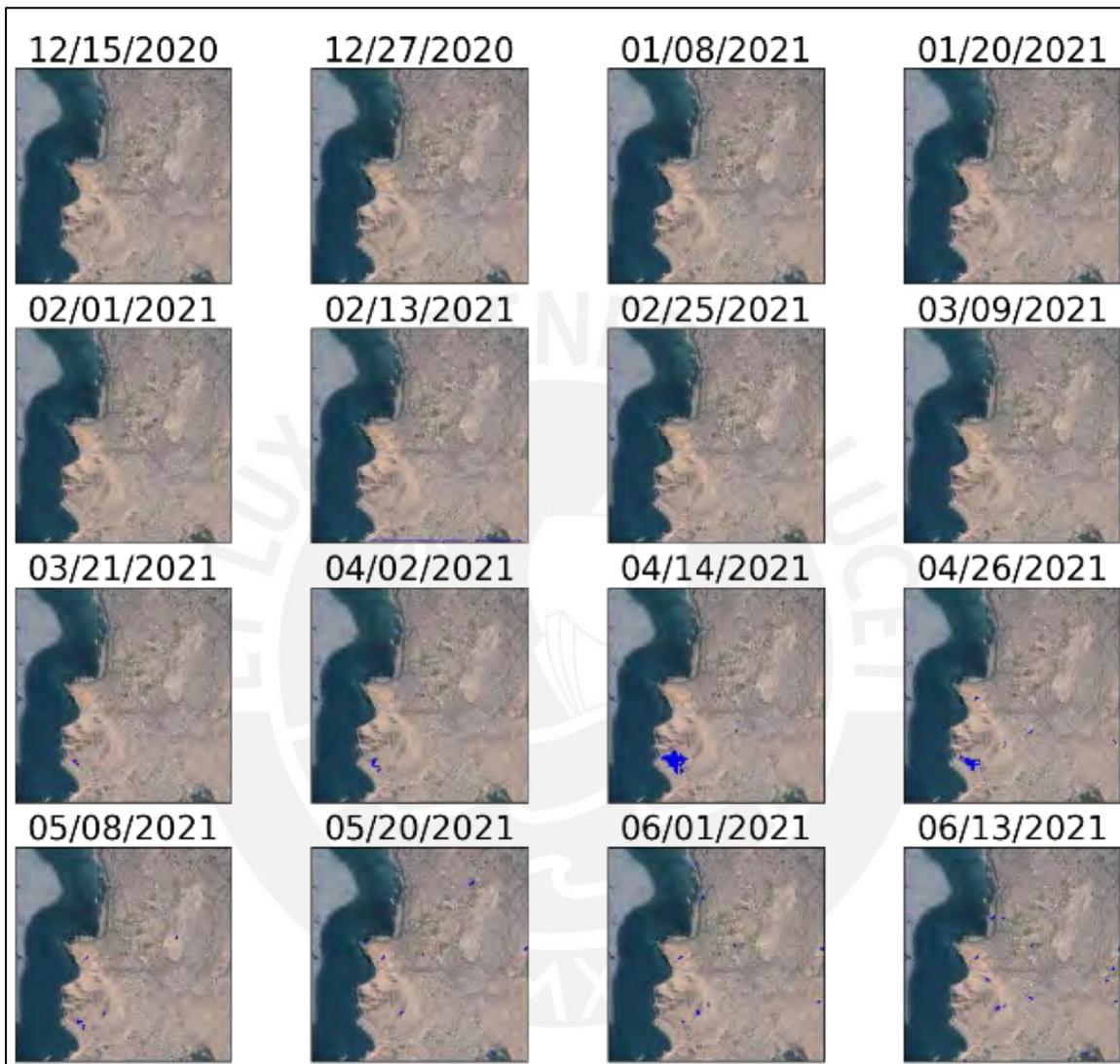


Figura 66: Resultados temporales mostrados por Naive Bayes Gaussiano

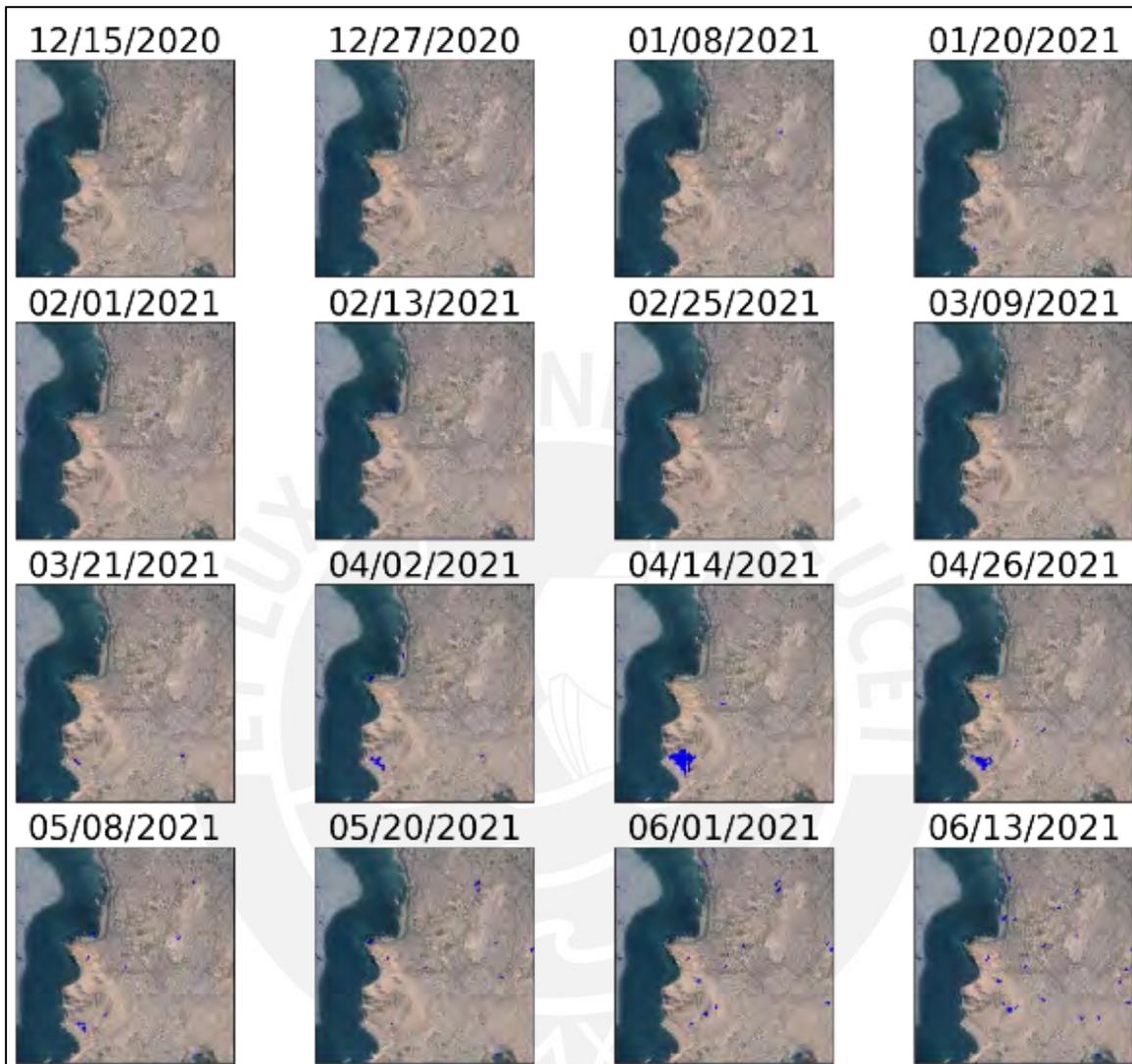


Figura 67: Resultados temporales mostrados por los Bosques aleatorios

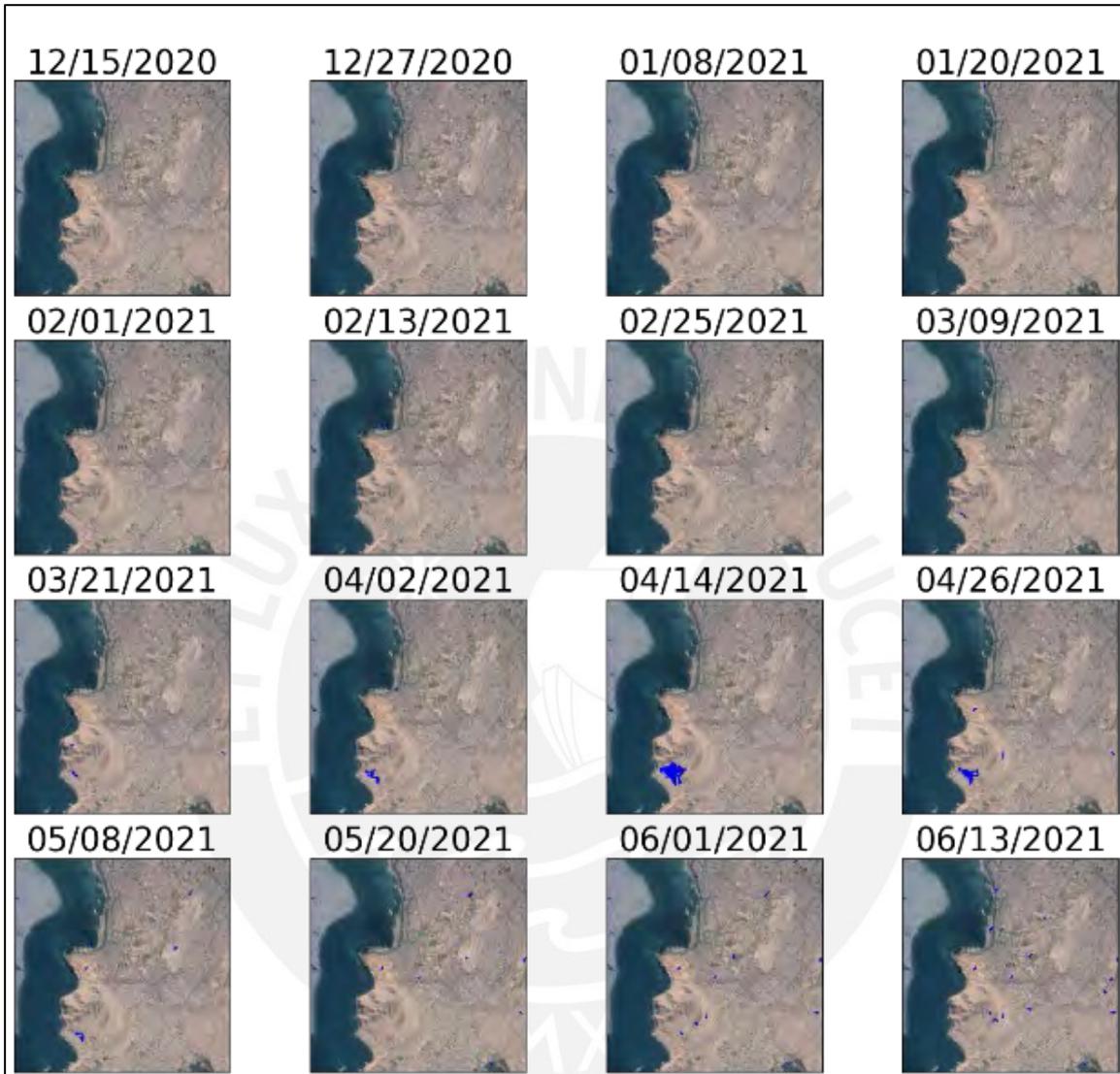


Figura 68: Resultados temporales mostrados por la Máquina de Soporte Vectorial

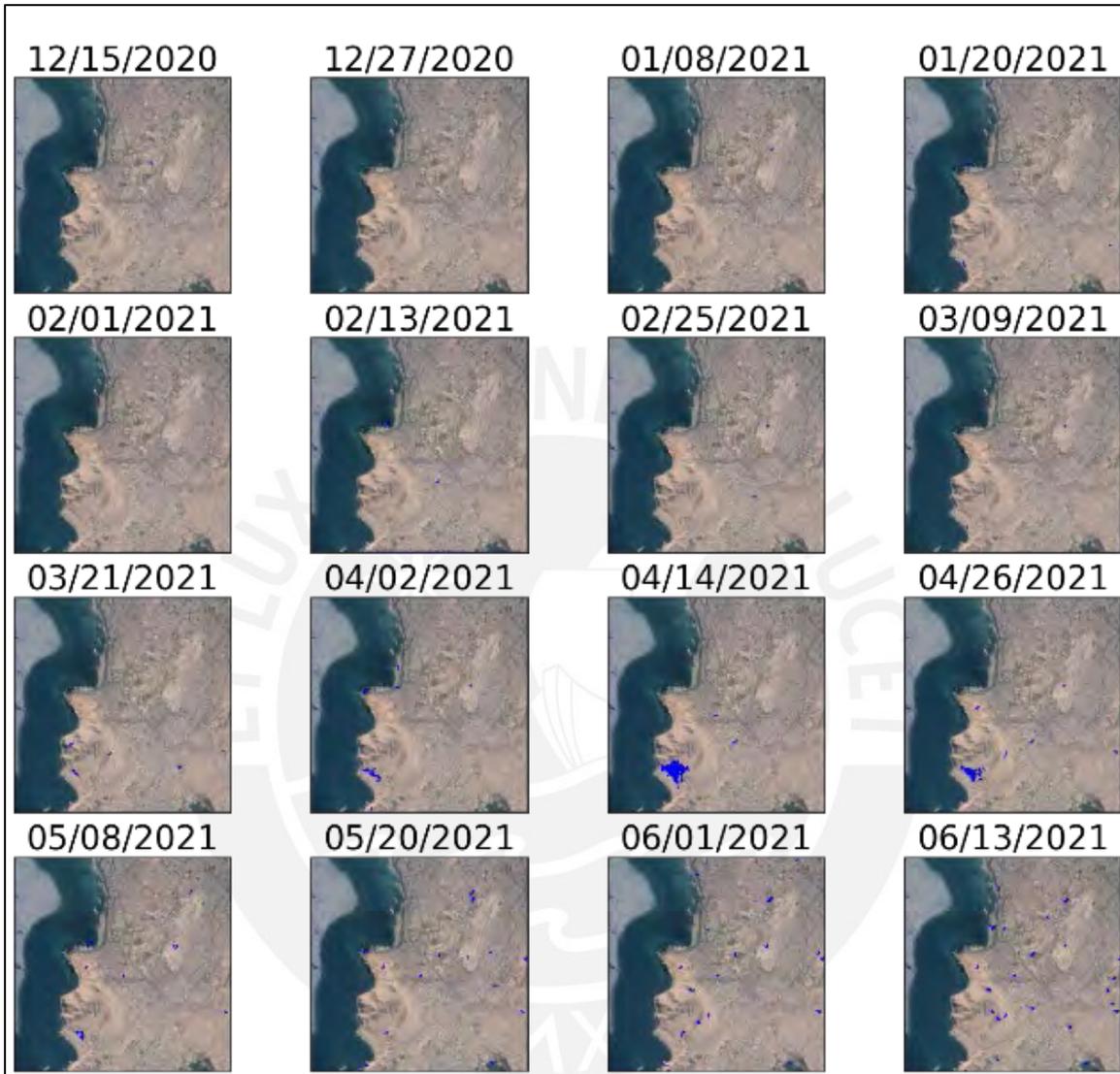


Figura 69: Resultados temporales mostrados por el Descenso del gradiente estocástico

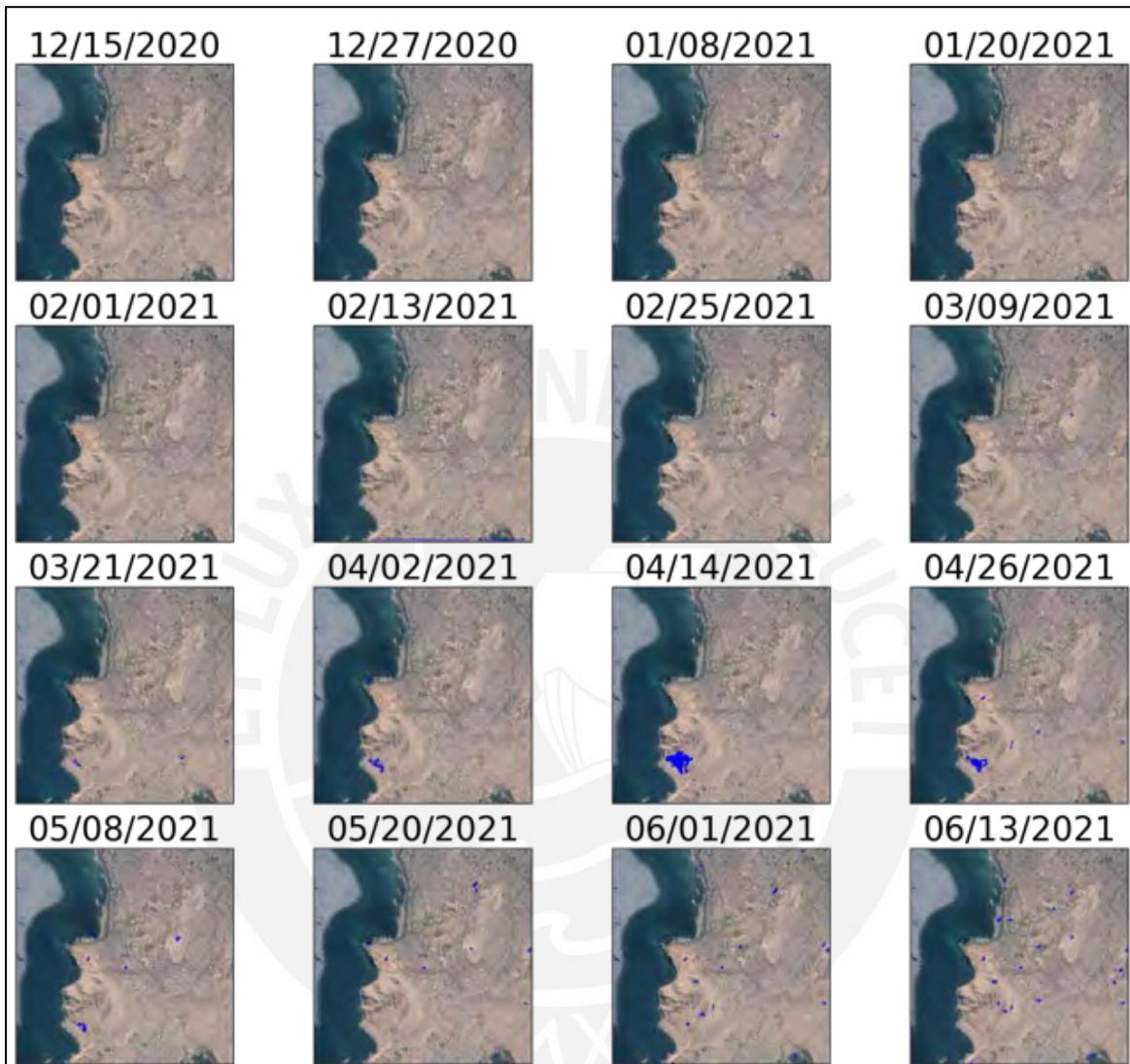


Figura 70: Resultados temporales mostrados por K-Vecinos Cercanos

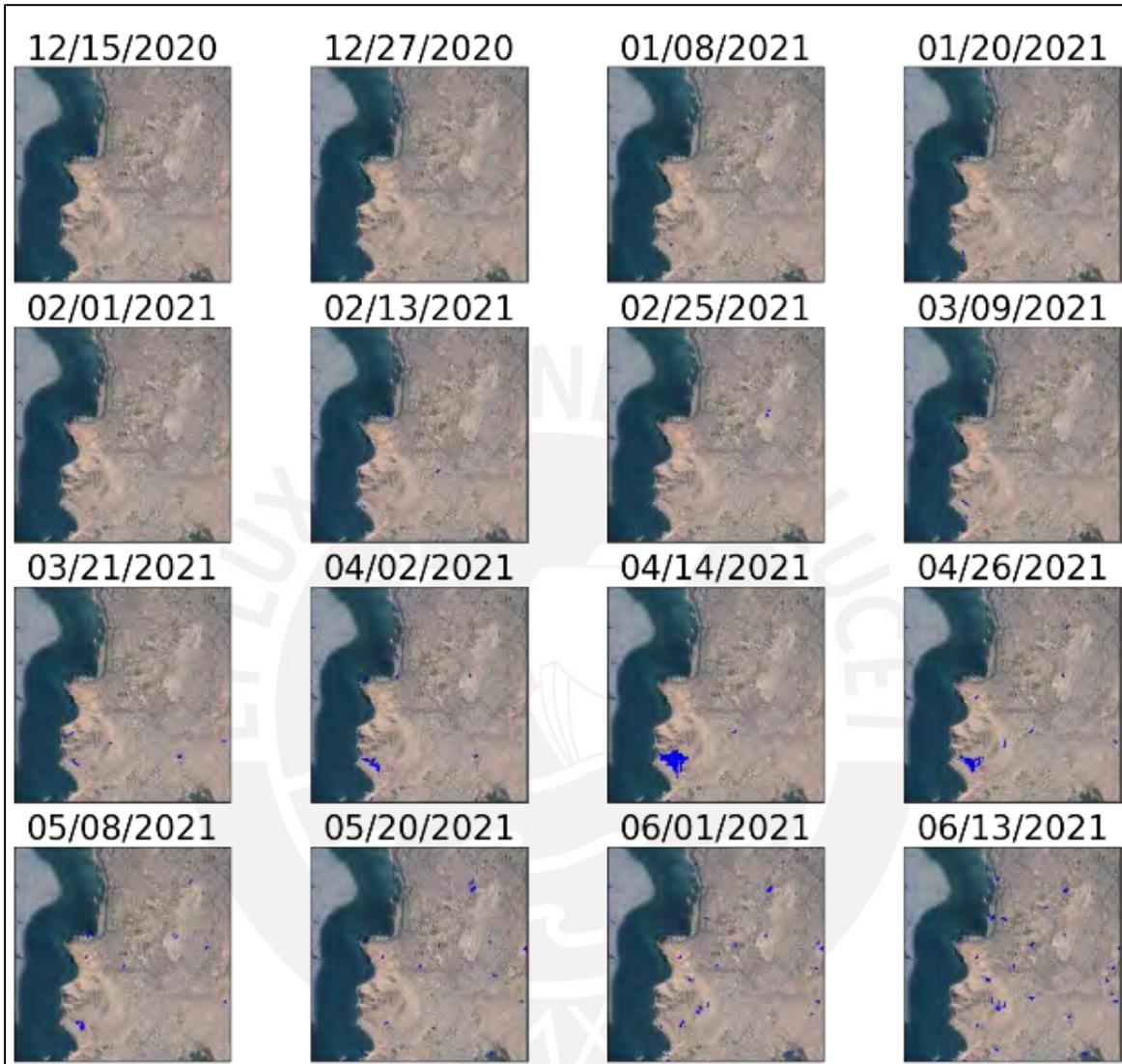


Figura 71: Resultados temporales mostrados por la regresión logística



Figura 72: Resultados temporales mostrados por la Envolverte Elíptica

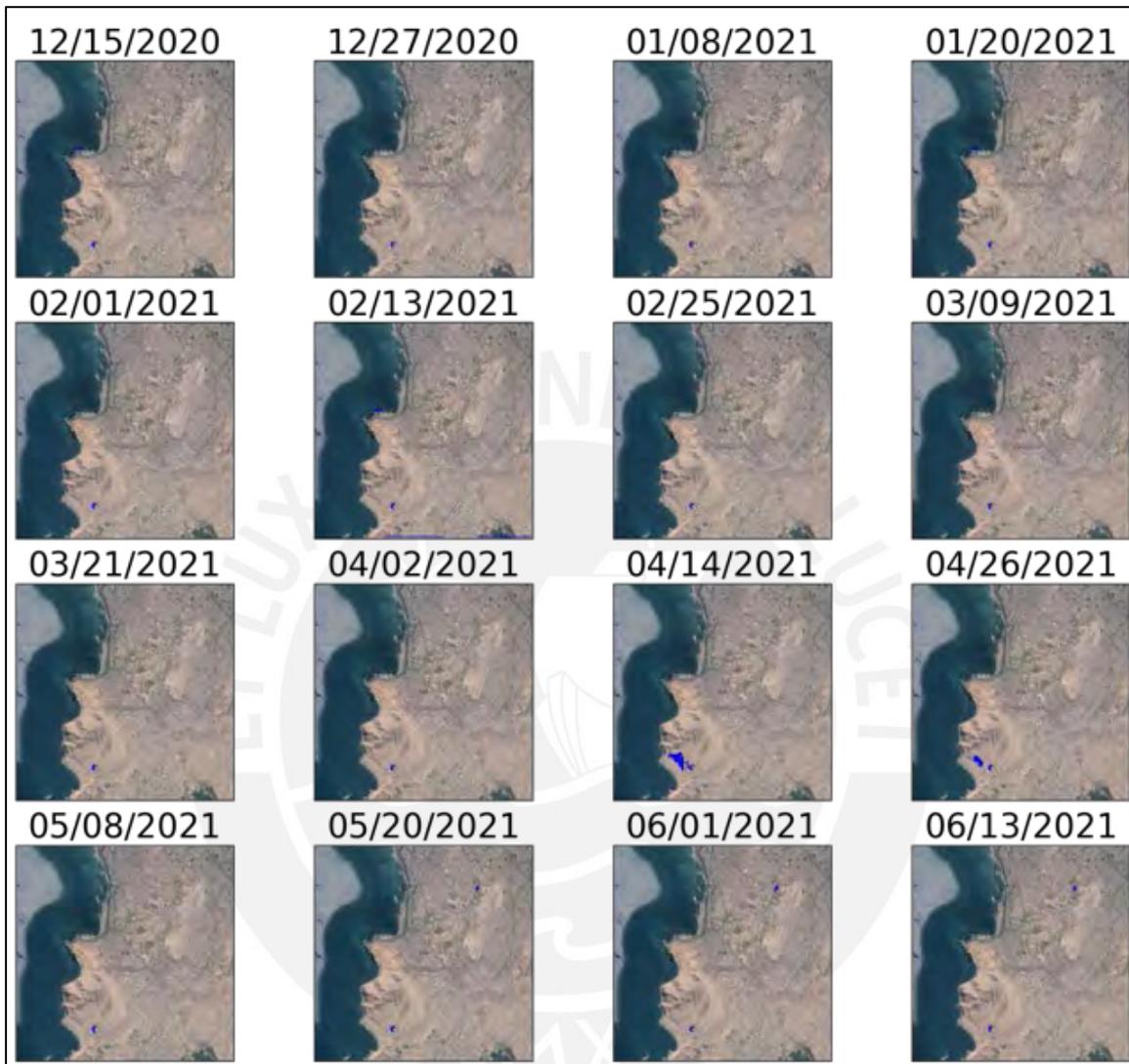


Figura 73: Resultados temporales mostrados por el Factor Atípico Local

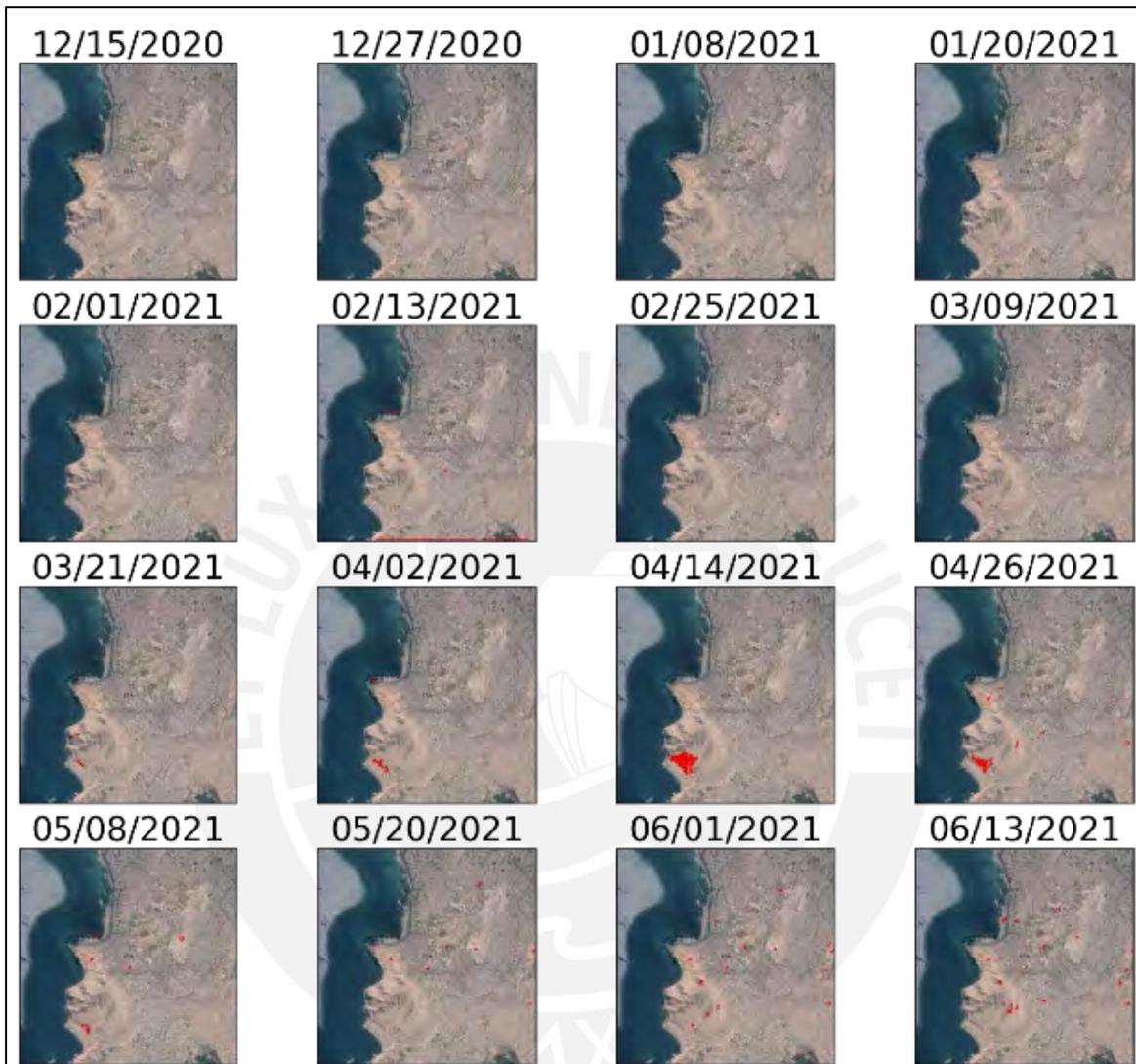


Figura 74: Resultados temporales mostrados por las redes neuronales