



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE
LA COMPUTACIÓN

ESTIMACIÓN DEL GRADO DE MARGINACIÓN EN
MÉXICO UTILIZANDO IMÁGENES SATELITALES Y
APRENDIZAJE PROFUNDO

T E S I S

QUE PARA OPTAR POR EL GRADO DE
MAESTRA EN CIENCIA E INGENIERÍA DE LA
COMPUTACIÓN

PRESENTA:
MARÍA FERNANDA GÓMEZ GALINDO

DIRECTOR DE TESIS:
DR. BORIS ESCALANTE RAMÍREZ
FACULTAD DE INGENIERÍA

CIUDAD UNIVERSITARIA, CIUDAD DE MÉXICO. MAYO 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A Dios por ser siempre la niña de sus ojos.

A mamá grande por guiarme.

A papá por bendecirme desde el cielo.

A mi mamá por ser mi inspiración.

A mi hermanito, espero poder ser siempre tu mejor ejemplo.

A mi tío por ser mi persona favorita y enseñarme a confiar en

Dios.

A mi tía y Valentina por estar siempre en mi corazón.

A José y al cachorro por ser los mejores compañeros de amor.

A mis amigos por ser mi sonrisa.

A mis asesores Boris Escalante y Jimena Olveres por apoyarme.

Al CONACYT por mi beca de maestría.

Al Programa de Apoyo a Proyectos de Investigación e Innovación

Tecnológica UNAM PAPIIT TA101121 y PAPIIT IV100420.

Resumen

En México los datos confiables disponibles para consulta sobre variables económicas y características socioeconómicas de la población son escasos y la mayoría son calculados con baja frecuencia. Esto representa un fuerte impedimento para realizar investigación y diseñar políticas públicas que impulsen el crecimiento económico del país. Además, en todo el mundo se han realizado grandes esfuerzos desde hace varios años para recolectar grandes bases de datos y desarrollar nuevas metodologías que puedan ser utilizadas para generar más información útil para diferentes actores

A través de esta investigación se demuestra que las técnicas de detección remota son útiles para identificar diversas características de la población. En particular, se observa como las imágenes satelitales tienen un poder predictivo razonable para clasificar el grado de marginación en la zona urbana, a partir de utilizar datos del Grado de Marginación e imágenes satelitales Landsat 7, de las 32 entidades federativas de la República Mexicana.

Para identificar características de las imágenes que explican hasta el 86.6% del grado de marginación se entrenan diferentes arquitecturas de redes neuronales convolucionales. Además, se prueban combinaciones de bandas de las imágenes satelitales y el procesa-

miento de refinado pancromático.

Este trabajo permite identificar a los sectores de la población con mayores rezagos a nivel Área Geoestadística Básica (AGEB) aproximadamente una vez cada dos semanas, periodicidad con la que se encuentran disponibles las imágenes satelitales. Esto muestra que existen enfoques prometedores, los cuales utilizan datos públicos y datos de entrenamiento limitados para estimar elementos socioeconómicos con imágenes satelitales y poder así disponer de información con mayor periodicidad.

Abstract

In Mexico, the reliable data available for consultation on economic variables and socioeconomic characteristics of the population are scarce and most are calculated with low frequency. This represents a strong impediment to carry out research and design public policies that promote the economic growth of the country. In addition, throughout the world, great efforts have been made for several years to collect large databases and develop new methodologies that can be used to generate more useful information for different actors.

The present work shows that remote sensing techniques are useful to identify various characteristics of the population. In particular, it is observed how satellite images have a reasonable predictive power to classify the degree of marginalization in the urban area, based on using data from the Degree of Marginalization and Landsat 7 satellite images of the 32 states of Mexico.

To identify characteristics of the images that explain up to 86.6 % of the degree of marginalization, different convolutional neural network architectures are trained. Additionally, band combinations of satellite imagery and panchromatic refinement processing are tested.

This work makes it possible to identify the sectors of the population with the greatest lags at the Basic Geostatistical Area (AGEB) level approximately once every two weeks, the periodicity with which satellite images are available. This shows that there are promising approaches, which use public data and limited training data to estimate socioeconomic elements with satellite images and thus make information available more regularly.

Índice general

Abreviaturas	1
1. Introducción	2
2. Estado del arte	6
3. Conjuntos de datos	11
3.1. Colección de imágenes satelitales	11
3.2. Datos del Índice de Marginación	15
4. Estimación del grado de marginación	18
4.1. Clasificación de imágenes	18
4.2. Modelo de predicción	21
5. Evaluación	31
5.1. Discusión de resultados	39

<i>ÍNDICE GENERAL</i>	7
6. Conclusiones	41
Bibliografía	43

Abreviaturas

AGEB - Área Geoestadística Básica.

CONAPO - Consejo Nacional de Población.

INEGI - Instituto Nacional de Estadística y Geografía.

PIB - Producto Interno Bruto.

ResNet - Red Neuronal Residual (Residual Neural Network, por sus siglas en inglés).

RF - Bosque Aleatorio (Random Forest, por sus siglas en inglés).

TempCNN - Redes Neuronales Convolucionales Temporales (Temporal Convolutional Neural Network, por sus siglas en inglés).

VGG - Grupo de Geometría Visual (Visual Geometry Group, por sus siglas en inglés).

Capítulo 1

Introducción

En México, los datos confiables sobre la economía y características socioeconómicas de la población son escasos, regularmente se publican con una frecuencia de baja utilidad.

Aunque, en nuestro país se realizan censos para recopilar información valiosa sobre el tamaño, la composición y la demografía de la población, esos conteos ocurren cada 5 y 10 años. Por el contrario, los satélites recogen datos a un ritmo constante durante todo el año y hay enfoques prometedores en el mundo que los estiman con esas imágenes. Utilizar técnicas de detección remota para estimar los datos en México permitiría tener referencias constantes y así entender mejor a la población para dirigir mejor los esfuerzos, maximizando los recursos disponibles e impulsando el crecimiento económico de México.

Los algoritmos de aprendizaje profundo demuestran tener buen desempeño en diferentes disciplinas como las ciencias sociales, las finanzas y otras. Muchas tareas han logrado obtener los mejores resultados en el mundo, por ejemplo en el procesamiento del lenguaje natural, en la medicina y la biología; en la generación de imágenes,

en los sistemas de recomendación, en los juegos de estrategia en tiempo real y en tareas de visión por computadora. Además, las redes neuronales convolucionales, desde su introducción por LeCun y cols. (1989), han sido eficientes en el reconocimiento y detección de objetos.

Este esfuerzo de investigación establece cómo las técnicas de detección remota son útiles para identificar características de la población. Las imágenes satelitales tienen un poder predictivo razonable para clasificar el grado de marginación en las zonas urbanas de México y así, identificar aquéllas con mayores rezagos, con una alta periodicidad, utilizando datos públicos y datos de entrenamiento limitados.

Para la clasificación de imágenes satelitales se utilizan cinco arquitecturas de redes neuronales convolucionales: AlexNet (Krizhevsky, 2014), VGG19 (Simonyan y Zisserman, 2015), SqueezeNet1 (Iandola y cols., 2016), Resnet50 (He, Zhang, Ren, y Sun, 2016), y Resnet152, cada una de estas arquitecturas está preentrenada en el conjunto de datos *ImageNet* (Deng y cols., 2009).

Además, para clasificarlas se utilizan datos del Grado de Marginación construidos por el Consejo Nacional de Población (CONAPO). Para su estimación se emplean los datos del Censo de Población y Vivienda publicado por el Instituto Nacional de Estadística y Geografía (INEGI) y para su definición se consideran algunos indicadores como: la carencia de bienes, acceso o no a educación y servicios de salud, así como la calidad de la vivienda. Tener datos con mayor periodicidad del grado de marginación en nuestro país es fundamental, pues es un fenómeno multidimensional que afecta a una persona, un grupo o a toda una colectividad. Sin embargo, estos datos se encuentran disponibles cada 5 años, siendo el último disponible el del año 2010.

Las imágenes satelitales son obtenidas de *Google Earth Engine*, que combina un catálogo de múltiples imágenes satelitales y con-

juntos de datos geospaciales, en específico, se usan las de Landsat 7.

Objetivo

El objetivo es demostrar que existen metodologías útiles para aumentar la periodicidad de los datos disponibles, sin necesidad de realizar censos o encuestas con mayor frecuencia, las cuales representan un alto costo monetario para los países. En nuestro país, los datos socioeconómicos regularmente se encuentran con baja periodicidad y la limitada frecuencia es un problema que afecta a los ciudadanos, empresas y gobierno. En particular, concierne a investigadores y formuladores de políticas públicas que fundamentan sus esfuerzos en estos datos que se reportan cada 5 o 10 años y que, en ocasiones, utilizar datos desactualizados provoca que no se llegue al resultado deseado o no se distribuyan los recursos eficientemente.

Contribución

Desarrollar una alternativa para estimar el Grado de Marginación a nivel AGEB en la zona urbana de México con una mayor periodicidad de la que actualmente se encuentra disponible, utilizando imágenes satelitales Landsat y redes neuronales convolucionales.

Asimismo, evaluar el desempeño de diferentes arquitecturas de redes neuronales convolucionales y distintas combinaciones de bandas de las imágenes satelitales.

Por último, demostrar como las redes neuronales convolucionales en conjunto con las imágenes satelitales sirven para estimar datos socioeconómicos en México sin hacer mediciones de encuestas

o censos. En este sentido, es de gran utilidad para identificar características en lugares con pocos recursos y, por tanto, con pocos datos.

Contenido

En el primer capítulo se desarrolla la introducción; en el segundo los antecedentes y la revisión de literatura; en el tercero se detalla el conjunto de datos del Grado de Marginación y las imágenes satelitales Landsat, en el cuarto se describe el análisis con aprendizaje profundo y la generación del modelo de predicción; en el quinto se presenta la evaluación del sistema, y en el último, las conclusiones.

Capítulo 2

Estado del arte

En esta sección se describen los antecedentes y revisión de literatura del procesamiento y clasificación de imágenes, y análisis predictivo con aprendizaje profundo.

De acuerdo con Konkle, Brady, Alvarez, y Oliva (2010) el cerebro humano puede almacenar miles de imágenes de escenas y objetos en la memoria visual y solamente requiere unos milisegundos para reconocer la categoría a la que pertenece. Para el humano es importante el procesamiento de retroalimentación en el reconocimiento visual y así logra aprender, al igual que los algoritmos de aprendizaje de máquina. Pero, ¿cuánto tendría que aprender un sistema artificial para alcanzar las capacidades de reconocimiento del ser humano? y, ¿esto es posible? (Zhou, Lapedriza, Xiao, Torralba, y Oliva, 2014)

Bengio (2009) menciona que aún no tenemos algoritmos que puedan entender y describir completamente ciertas escenas, excepto en entornos muy limitados y tampoco tenemos algoritmos que puedan inferir suficientes conceptos semánticos para interactuar con el ser humano.

No obstante, los avances en la resolución de tareas de visión computacional demuestran que el aprendizaje profundo es una herramienta poderosa, en donde las características aprendidas son efectivas para resolver problemas de esta índole.

LeCun y cols. (1990) popularizaron las redes neuronales convolucionales, las cuales probaron tener un buen desempeño al clasificar dígitos escritos a mano y destacaron que estas redes pueden aplicarse a problemas reales de reconocimiento de imágenes sin una etapa de preprocesamiento compleja de los datos.

Los autores He y cols. (2018) mencionan que desde la introducción de *AlexNet* (Krizhevsky, 2014; Krizhevsky, Sutskever, y Hinton, 2012), las redes neuronales convolucionales se convirtieron en el enfoque predominante para la clasificación de imágenes, desde entonces se proponen varias arquitecturas, como la *VGG* (Simonyan y Zisserman, 2015) y la *ResNet* (He y cols., 2016).

La red neuronal convolucional *AlexNet* (Krizhevsky, 2014; Krizhevsky y cols., 2012) incluye ocho capas: cinco capas convolucionales (*convolutional layers*) y tres capas completamente conectadas (*fully-connected layers*). Además, se utilizan Unidades Lineales Rectificadas (*ReLU*) en lugar de la función *tanh*, lo cual mejora el tiempo de entrenamiento, e incluyen el método de regularización ‘*dropout*’ que reduce el sobreajuste.

Simonyan y Zisserman (2015) evalúan redes convolucionales para clasificar imágenes y demuestran que la profundidad es beneficiosa para la precisión de la clasificación. Sus modelos logran tener un buen rendimiento en diferentes tareas e introducen la red neuronal convolucional *VGG16*, la cual muestra una mejora respecto a la *AlexNet*.

Con la introducción de la *ResNet*, los autores He y cols. (2016) observaron que una red que utilizaba más capas funcionaba peor en el conjunto de validación y en el de entrenamiento que otra que usaba menos capas con las mismas características entre ambos mo-

delos. Para lo cual se esperaría lo contrario, que las redes profundas con más capas fueran al menos tan buenas como una red con menos capas. Ante esto, introdujeron la variante de pasar por un bloque de dos capas convolucionales o bien, saltar las convoluciones para facilitar el aprendizaje, como se muestra en la figura 2.1.

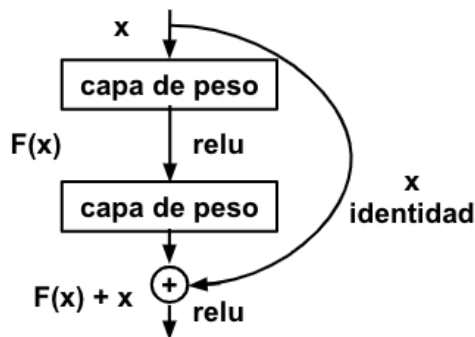


Figura 2.1: Aprendizaje residual.

Fuente: Elaboración propia, con información de He y cols., 2016.

Zeiler y Fergus (2014) reforzaron que las redes neuronales convolucionales tienen un buen desempeño en la clasificación de imágenes y asentaron que las características que se aprenden son patrones interpretables con propiedades como la invarianza creciente y la discriminación de clase a medida que se asciende por las capas.

En el mundo, se han realizado esfuerzos para utilizar datos extraídos de redes sociales, de registros telefónicos móviles, de encuestas socioeconómicas realizadas en cada país, y de tarjetas inteligentes para evidenciar características de la población en imágenes satelitales.

Jean y cols. (2016) extraen datos socioeconómicos de imágenes satelitales de alta resolución en cinco países africanos. Los autores utilizan aprendizaje por transferencia, en donde entrenan una

red neuronal convolucional con *ImageNet* (Deng y cols., 2009) y después, ajustan la red neuronal para predecir la intensidad de la luz nocturna en imágenes diurnas. Ellos concluyen que algunas características de estas imágenes, como el material del techo y la distancia a las áreas urbanas tienen relación con el nivel de gasto. También, los autores Xie, Jean, Burke, Lobell, y Ermon (2016), entrenan un modelo para predecir las luces nocturnas a partir de imágenes diurnas, en donde se aprenden características que son útiles para pronosticar la pobreza como las carreteras, edificios y tierras de cultivo.

Albert, Kaur, y Gonzalez (2017) analizan patrones en el uso del suelo en vecindarios urbanos con imágenes satelitales y redes neuronales convolucionales. Para entrenar utilizan el conjunto de datos de clasificación del suelo " *Urban Atlas*" que contiene 20 clases en 300 ciudades europeas y comparan diferentes arquitecturas que han evidenciado un buen desempeño en tareas de visión computacional. Además, Robinson, Hohman, y Dilkina (2017) entrenan redes neuronales convolucionales para predecir el total de la población en Estados Unidos a partir de imágenes Landsat.

Wu y Tan (2019), utilizan la *ResNet50* para clasificar la intensidad de la luz nocturna en imágenes satelitales para predecir dos indicadores económicos: el Producto Interno Bruto (PIB) y las ventas totales al por menor de bienes de consumo. Después, encuentran que las redes neuronales convolucionales combinadas con imágenes satelitales se pueden utilizar en la estimación de la pobreza regional (Wu y Tan, 2020).

Sun, Di, Sun, Wang, y Wu (2020), utilizan redes neuronales convolucionales con imágenes de *Google Earth Engine* para la estimación del PIB y mencionan que el método es aplicable a otros países, especialmente aquéllos en desarrollo, en donde los datos socioeconómicos son difíciles de obtener.

También, Pelletier, Webb, y Petitjean (2019) mencionan que

los algoritmos de clasificación tradicionales, como *Random Forest (RF)*, se han utilizado con éxito en estas tareas, pero no logran aprovechar el dominio temporal, mientras que las redes neuronales convolucionales tienen un mejor desempeño. Ellos, además, proponen una arquitectura de redes neuronales convolucionales llamadas *TempCNN*.

Otra metodología utilizada es la segmentación semántica, en dónde las arquitecturas de aprendizaje profundo se han aplicado con éxito en imágenes satelitales. Por ejemplo, en Wurm, Stark, Zhu, Weigand, y Taubenböck (2019), los autores utilizan imágenes de los sensores QuickBird, Sentinel-2 y TerraSAR-X para clasificar píxel por píxel en urbano, vegetación, agua y barrios marginados con aprendizaje por transferencia. Por otra parte, Iglovikov, Mushinskiy, y Osin (2017), como un desafío de detección de características en imágenes satelitales de Kaggle, utilizaron segmentación semántica para clasificar en edificios, estructuras, carreteras, agua, árboles, cultivos y vehículos con redes neuronales convolucionales.

Capítulo 3

Conjuntos de datos

3.1. Colección de imágenes satelitales

Las imágenes satelitales utilizadas en este trabajo de investigación son obtenidas de *Google Earth Engine*, una plataforma de *Google* que combina un catálogo de imágenes satelitales y datos geoespaciales con capacidades de análisis a escala planetaria. Se emplean imágenes Landsat, las cuales provienen de un programa conjunto entre la USGS y la NASA, esta colección de imágenes se encuentra disponible desde 1972 hasta nuestros días. Los satélites Landsat obtienen imágenes de toda la superficie de la Tierra, una vez cada dos semanas, a una resolución aproximada de 30 metros, dependiendo de las bandas.

Con frecuencia, dichas imágenes tienen interferencias que perturban su información, por ejemplo, fallos en los sensores que generan píxeles incorrectos, alteraciones en el movimiento del satélite que provocan distorsiones en la imagen e interferencias de la atmósfera que afectan sistemáticamente los valores de los píxeles. Para corregir estas perturbaciones se realiza correcciones ra-

diométricas, geométricas y atmosféricas.

La corrección radiométrica se refiere a la modificación de píxeles y del bandeo. Para arreglar los píxeles faltantes se puede utilizar la media de los valores anterior y posterior, o bien, el valor del píxel en una banda diferente. El bandeo se refiere a la aparición periódica de una banda más clara u oscura y se puede enmendar utilizando el histograma global de la imagen. La corrección geométrica subsana la distorsión espacial y la rotación de la imagen al transformar las coordenadas de los píxeles para que coincidan con las de la imagen de referencia. Por último, con la corrección atmosférica se eliminan las interferencias producidas por la interacción de la radiación con la atmósfera.

En específico se utilizan imágenes Landsat 7 dada la temporalidad disponible, desde 1999 hasta la fecha, las cuales están disponibles en tres niveles diferentes:

1. Tier 1: son los datos que cumplen con los requisitos de calidad geométrica y radiométrica.
2. Tier 2: son los datos que no cumplen con los requisitos de calidad geométrica y radiométrica.
3. RT: son los datos que aún no se han evaluado.

Al mismo tiempo, se encuentran en tres diferentes opciones:

- Reflectancia superficial: conjunto de datos corregidos atmosféricamente mediante LEDAPS ¹ e incluyen una máscara de nube, sombra, agua y nieve producida con CFMASK ², así como una máscara de saturación por píxel.

¹El Sistema de Procesamiento Adaptativo de Perturbación del Ecosistema Landsat (LEDAPS) procesa las imágenes utilizando métodos de corrección atmosférica desarrollados para Terra MODIS (Vermote y cols., 1997).

²Algoritmo que corrige la presencia de nubes, sombras de las nubes, nieve y hielo.

- Techo de la atmósfera: estos datos presentan una reflectancia calibrada de la parte superior de la atmósfera, en donde los coeficientes de calibración son extraídos de los metadatos ³ de la imagen.
- Imágenes en crudo

Es importante agregar que las imágenes de Landsat contienen nueve bandas:

Banda	Resolución	Descripción
B1	30 metros	Azul
B2	30 metros	Verde
B3	30 metros	Rojo
B4	30 metros	Infrarrojo cercano
B5	30 metros	Infrarrojo de onda corta 1
B6 1	30 metros	Infrarrojo térmico de baja ganancia
B6 2	30 metros	Infrarrojo térmico de alta ganancia
B7	30 metros	Infrarrojo de onda corta 2
B8	15 metros	Pancromática

Cuadro 3.1: Bandas y resolución de las imágenes satelitales Landsat 7.

Fuente: *Google Earth Engine*.

Las bandas de las imágenes tienen aplicaciones diferentes. De acuerdo con los “Aspectos Técnicos de las Imágenes Landsat” del INEGI⁴, la banda azul se utiliza para el mapeo de aguas costeras y la diferenciación entre suelo y vegetación; la banda verde para la

³Los metadatos son datos que describen los datos espaciales, es decir, son archivos de información que contienen las características básicas de los datos.

⁴Documento disponible en: <https://www.inegi.org.mx/temas/imagenes/imgLANDSAT/>

vegetación y observar la calidad del agua; la roja muestra la diferenciación de especies vegetales, áreas urbanas y uso del suelo; la banda infrarrojo cercano ubica la delimitación de cuerpos de agua, el mapeo geomorfológico y geológico, la identificación de áreas de incendios, agricultura y vegetación; la banda infrarrojo térmico mide la humedad, la diferenciación entre nubes y nieve e identifica la agricultura y vegetación. Por último, la banda pancromática, banda en blanco y negro, normalmente se encuentra disponible en una alta resolución espacial y puede combinarse con otras de menor resolución y así generar una imagen multibanda de mayor resolución.

En la figura 3.1 se muestra el territorio mexicano compuesto por imágenes satelitales correspondientes a las bandas roja, verde y azul.

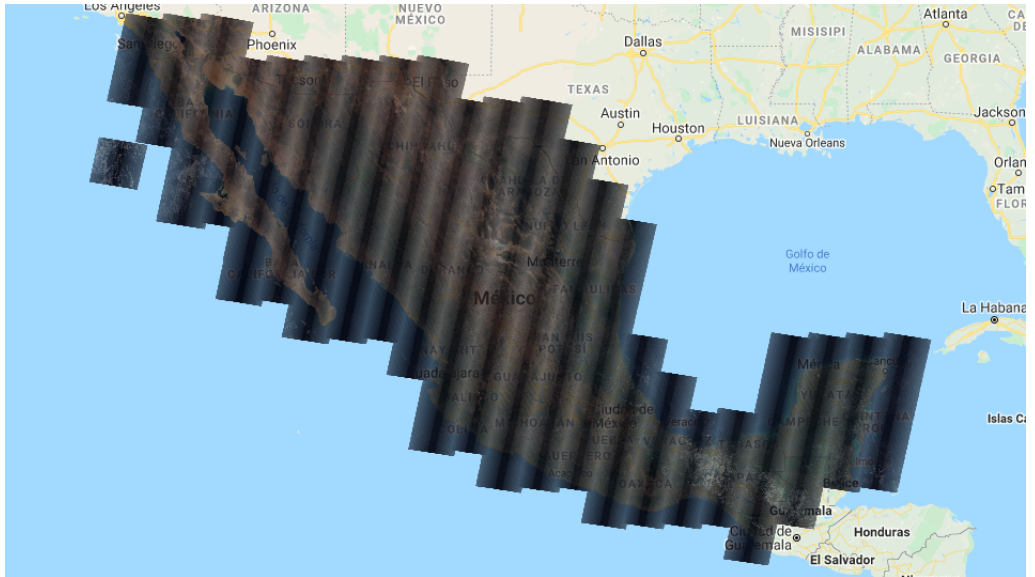


Figura 3.1: Mapa de la República Mexicana compuesto por imágenes satelitales correspondientes a las bandas roja, verde y azul (RGB).

Fuente: Elaboración propia, con información de *Google Earth Engine*.

3.2. Datos del Índice de Marginación

El Índice de Marginación es una estimación realizada por el CONAPO del territorio nacional mexicano, desde 1990 hasta el 2015, que refleja las carencias de la población mexicana. El índice mide diferentes aspectos como falta de acceso a la educación y a los servicios de salud, la residencia en viviendas inadecuadas y la falta de bienes.

Esta estimación se encuentra disponible en diferentes niveles de desagregación de acuerdo con el Marco Geoestadístico Nacional

del Instituto Nacional de Estadística y Geografía (INEGI) las cuales son: Área Geoestadística Estatal (AGEE), Área Geoestadística Municipal (AGEM), Localidad Geoestadística y Área Geoestadística Básica (AGEB).

Además, para construir el índice se utilizan cuatro dimensiones y diez indicadores socioeconómicos que se obtienen del Censo de Población y Vivienda publicado por el INEGI. En el cuadro 2.1 se muestran las dimensiones y los indicadores:

Dimensiones	Indicadores
Educación	% Población de 6 a 14 años que no asiste a la escuela.
	% Población de 15 años o más sin educación básica completa.
Salud	% Población sin derechohabiencia a los servicios de salud.
	% Hijos fallecidos de las mujeres de 15 a 49 años de edad.
Vivienda	% Viviendas particulares habitadas sin drenaje conectadas a la red pública o fosa séptica.
	% Viviendas particulares habitadas sin excusado con conexión de agua.
	% Viviendas particulares habitadas sin agua entubada dentro de la vivienda.
	% Viviendas particulares habitadas con piso de tierra.
	% Viviendas particulares habitadas con algún nivel de hacinamiento.
Bienes	% Viviendas particulares habitadas sin refrigerador.

Cuadro 3.2: Dimensiones e indicadores considerados en la construcción del índice de Marginación.

Fuente: Elaboración propia con información de CONAPO.

Para este trabajo de investigación se utilizaron los datos del Grado de Marginación por AGEB urbana del año 2010, los cuales son los últimos disponibles para clasificar las imágenes satelitales

Capítulo 4

Estimación del grado de marginación

4.1. Clasificación de imágenes

La clasificación de imágenes es el proceso de asignar una categoría a una imagen a partir de un conjunto de categorías predefinidas. De acuerdo con Kadhim y Abed (2020) es el proceso de categorización dependiendo del objeto o de su significado semántico, de modo que se pueden catalogar en métodos basados en características bajas, medias y altas.

Para este trabajo se utilizaron imágenes satelitales Landsat 7 de los 32 estados de la República Mexicana del año 2010. Este conjunto de datos presentan una reflectancia calibrada de la parte superior de la atmósfera, en donde los coeficientes de calibración son extraídos de los metadatos de la imagen y cumplen con los requisitos de calidad geométrica y radiométrica.

Se realizaron diferentes combinaciones de bandas que permitie-

ron diferenciar entre zonas vegetales, masas de agua, fenómenos naturales o provocados por el hombre, y usos de suelo. Para llevar a cabo la composición se eligen las bandas y su orden, se efectúan combinaciones lineales con el mismo peso, o bien, cada banda con uno diferente. La composición más habitual es la denominada RGB que es el resultado de juntar las bandas roja, verde y azul en la misma proporción.

Las combinaciones ¹ de bandas que se utilizan en este trabajo son (figura 4.1):

- RGB: roja, verde y azul de 30 metros de resolución cada una.
- Falso color para zonas urbanas: infrarrojo de onda corta 2, infrarrojo de onda corta 1 y roja de 30 metros de resolución cada una.
- Infrarrojo para vegetación: infrarrojo cercano, roja y verde de 30 metros de resolución cada una.
- Uso del suelo: infrarrojo cercano, infrarrojo de onda corta 1 y roja de 30 metros de resolución cada una.

¹Se realiza la combinación de las bandas utilizando cada banda en la misma proporción.

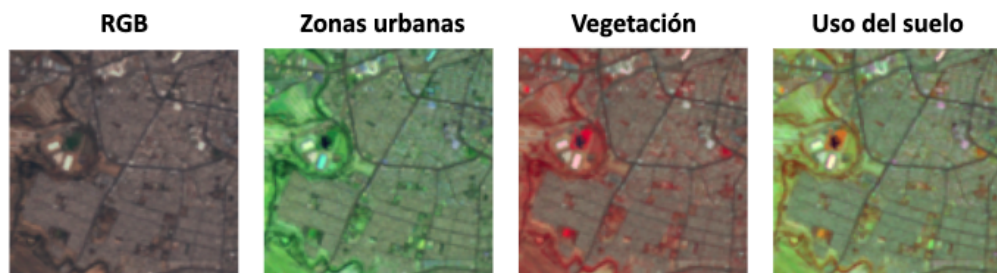


Figura 4.1: Combinaciones de bandas.

Fuente: Elaboración propia, con información de *Google Earth Engine*.

Las imágenes se etiquetan con los datos del Grado de Marginación por AGEB urbana para 2010 en las siguientes tres clases:

- Baja marginación.
- Media marginación.
- Alta marginación.

Dado que una imagen puede tener más de un AGEB, para asignarle una categoría, se tomó el nivel de marginación predominante, ya que se cuenta con la localización de los elementos geográficos y sus atributos asociados.

Nuestra implementación del tamaño de las imágenes sigue la práctica de los autores He y cols. (2016); Krizhevsky (2014); Krizhevsky y cols. (2012); Simonyan y Zisserman (2015) que utilizan imágenes de $224 \times 224 \times 3$ como datos de entrada para las redes neuronales convolucionales. Siguiendo lo anterior, las imágenes satelitales se cortan en $224 \times 224 \times 3$ píxeles y solamente se utilizan aquellas en las que al menos la mitad de sus píxeles contienen zonas urbanas. Se obtienen un total de 2,198 distribuidas en las 32 entidades federativas, siendo Colima y Nayarit los estados con menos

imágenes y el Estado de México, con más de éstas. En la figura 4.2 se ejemplifica el conjunto de datos obtenidos.

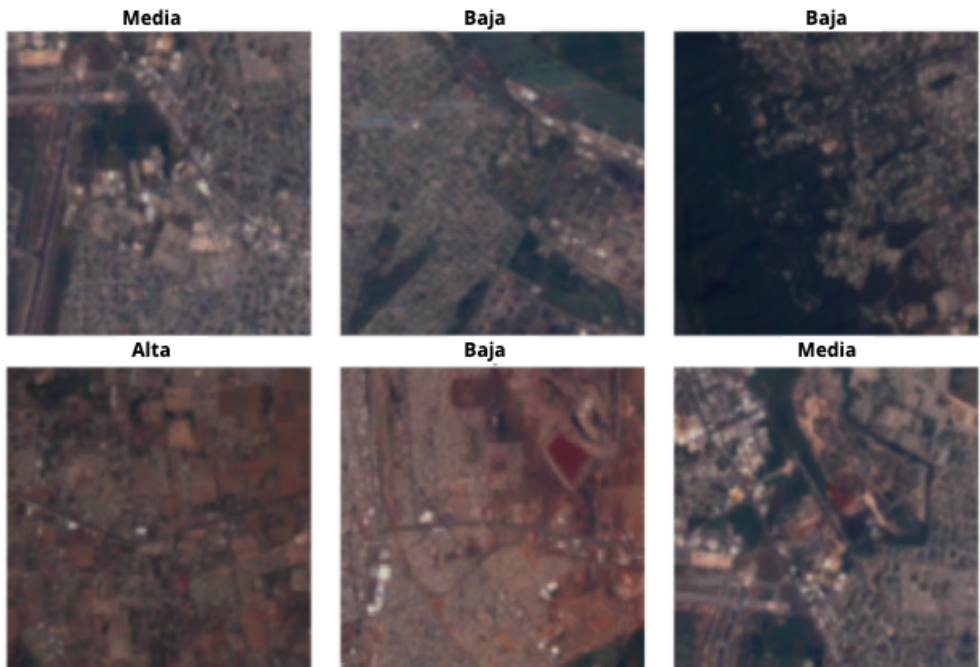


Figura 4.2: Imágenes de referencia para el conjunto de datos correspondientes a las bandas roja, verde y azul (RGB) y sus etiquetas. Fuente: Elaboración propia, con información de *Google Earth Engine* y CONAPO.

4.2. Modelo de predicción

De acuerdo con O'Shea y Nash (2015), las redes neuronales son sistemas computacionales inspirados en los sistemas nerviosos biológicos. Se componen principalmente de nodos interconectados llamados neuronas que trabajan entrelazados de manera distribui-

da, aprendiendo conjuntamente de la entrada dada, con el objetivo de optimizar la salida. Las redes presentan capas de entrada, las cuales reciben la información de entrada; las capas ocultas, que toman decisiones a partir de la capa anterior, y la capa de salida, que nos da el resultado final.

En la figura 4.3 se muestra una estructura de tres capas, compuesta por una de entrada, una oculta y otra de salida.

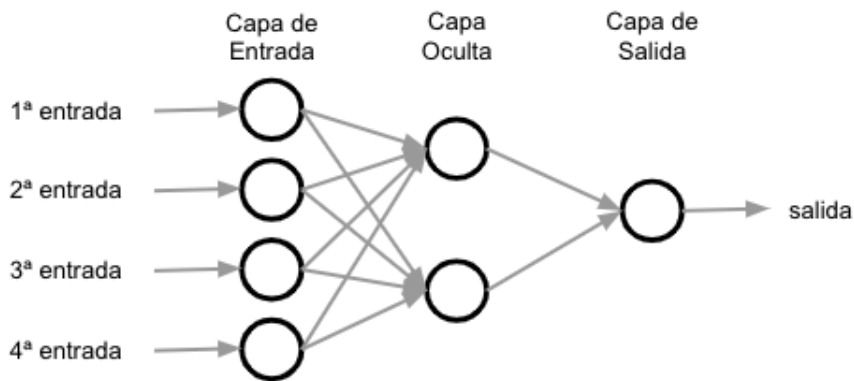


Figura 4.3: Estructura simple de red neuronal.

Fuente: Elaboración propia, con información de O'Shea y Nash, 2015.

Hinton, Srivastava, Krizhevsky, Sutskever, y Salakhutdinov (2012) mencionan que hay varios factores que son responsables de mejorar la precisión de las redes neuronales:

- Grandes conjuntos de datos de entrenamiento.
- Implementaciones de GPU disponibles.
- Estrategias de regularización.

- Profundidad de las redes neuronales. (Simonyan y Zisserman, 2015)

De acuerdo con Goodfellow, Bengio, y Courville (2016), las redes convolucionales son redes neuronales que utilizan la convolución en lugar de la multiplicación general de matrices en alguna de sus capas. Las redes neuronales convolucionales están inspiradas en la estructura del sistema visual de los seres vivos, su estructura típica está conformada por capas que incluyen una capa convolucional (*convolutional layer*), una de agrupación (*pooling layer*) y otras de conexión completas (*full connection layers*) (Bengio, 2009).

La capa convolucional (*convolutional layer*) es una capa extractora de características en donde se realiza la operación de convolución, en esta capa se puede incluir la activación ReLU que cambia todos los valores negativos a cero. Esta capa es parte fundamental de las redes neuronales convolucionales, cada filtro en la dimensionalidad espacial de la entrada produce un mapa de activación 2D. Las activaciones son los núcleos que se activan cuando ven una característica específica en una posición determinada.

La convolución es un promedio ponderado, con el cual obtenemos nuevas funciones que proporcionan una estimación suavizada (figura 4.4). La operación de convolución generalmente se indica con un asterisco como en la función 4.1:

$$s(t) = (x * w)(t) \tag{4.1}$$

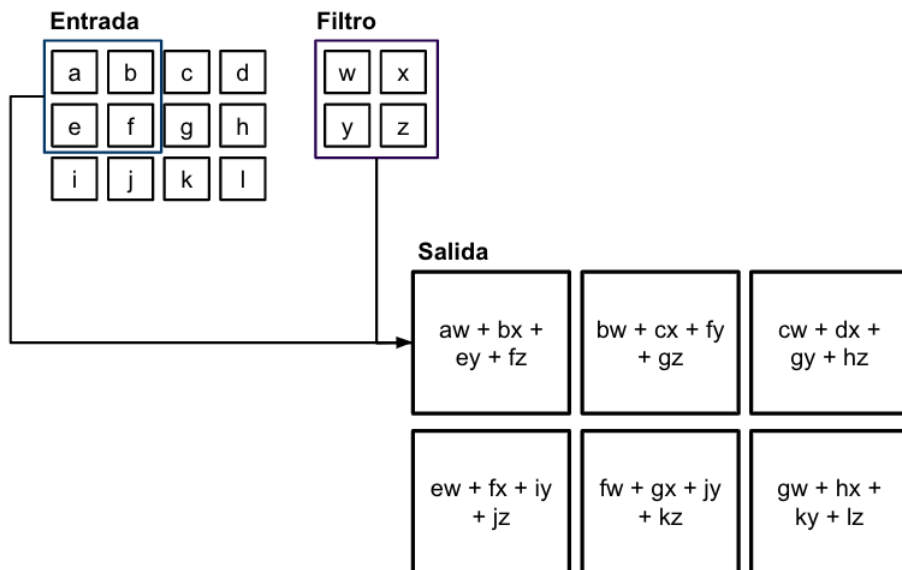


Figura 4.4: Ejemplo de una convolución 2-D.

Fuente: Elaboración propia, con información de Goodfellow y cols., 2016.

Howard y Gugger (2020) muestran que para aplicar una convolución a una imagen a color, es decir con varias bandas, se necesitan tantos filtros como bandas, ya que no usamos el mismo filtro para todas. Entonces, en cada una, multiplicamos los elementos de la ventana por los del filtro correspondiente y los agregamos, para después sumar los resultados obtenidos en las diferentes bandas como se ejemplifica en la figura 4.5.

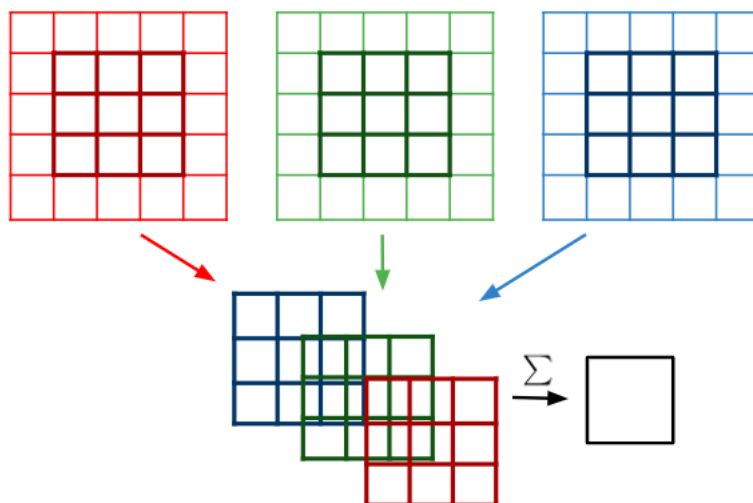


Figura 4.5: Convolución a una imagen de múltiples bandas.
Fuente: Elaboración propia, con información de Howard y Guggen, 2020.

La capa de agrupación (*pooling layer*) se utiliza para reducir el volumen espacial de la imagen después de la convolución, para que no sea computacionalmente costoso. Por otro lado, las capas de conexión completas (*full connection layers*) conectan neuronas de una capa con las neuronas de otra capa y se utilizan para clasificar imágenes entre diferentes categorías.

En los últimos años, las técnicas de aprendizaje profundo utilizadas con imágenes en tareas como la detección y clasificación de objetos han mejorado notablemente. Estas son más efectivas cuando tienen muchos datos de entrenamiento etiquetados, pero la desventaja es que aumenta el tiempo de cálculo (Lavin y Gray, 2016).

Sin embargo, en México los datos de marginación son escasos, por lo que nuestro conjunto de datos es más pequeño que los que

normalmente se utilizan en aplicaciones de aprendizaje profundo. Por lo tanto, aunque las redes neuronales convolucionales podrían entrenarse para estimar directamente el grado de marginación, la escasez que se tiene de datos de entrenamiento dificulta la aplicación de las técnicas.

Por lo anterior, se utiliza aprendizaje por transferencia para estimar los resultados. Zeiler y Fergus (2014) muestran cómo un modelo entrenado con *ImageNet* puede generalizarse bien a otros conjuntos de datos. Por lo que, primero, se empezó con un modelo de red neuronal convolucional preentrenado en *ImageNet*, el cual es un conjunto de datos que tiene más de un millón de imágenes de entrenamiento etiquetadas en cerca de mil categorías (Deng y cols., 2009). Con esto, el modelo aprende a clasificar cada imagen e identifica características comunes en las tareas de visión computacional.

Al utilizar el modelo preentrenado, se elimina la última capa y se reemplaza con una o más con diferentes pesos del tamaño apropiado para nuestro conjunto de datos (Howard y Gugger, 2020). Es decir, se ajusta la red neuronal convolucional a la nueva tarea de clasificación de imágenes, y se entrena para estimar el grado de marginación correspondiente a las imágenes diurnas de entrada. En este paso, el modelo decide la importancia de las características aprendidas para mejorar los resultados de la nueva tarea.

Los autores Zeiler y Fergus (2014), muestran como una red neuronal convolucional entrenada con *ImageNet* va aprendiendo filtros conforme avanza en sus capas. En la primera capa se observa que el modelo descubre pesos y componentes básicos que representan bordes verticales, horizontales y diagonales, pero en la segunda se observan características como patrones simples, esquinas y círculos, de esta forma, conforme se avanza se identifican componentes de nivel superior como ruedas de automóviles, perros y patas de animales como se muestra en la figura 4.6.

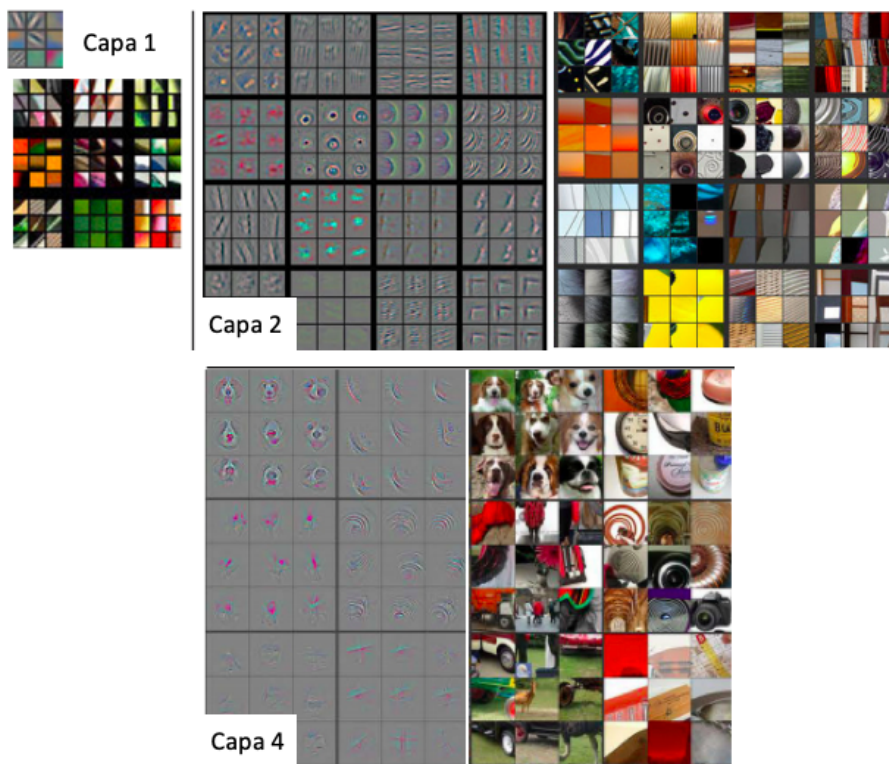


Figura 4.6: Filtros aprendidos por las capas convolucionales. Del lado izquierdo se observan las imágenes de los pesos aprendidos y del lado derecho se muestran las imágenes de entrenamiento que coinciden con las características aprendidas.

Fuente: Zeiler y Fergus, 2014.

Para la evaluación cuantitativa de los resultados utilizamos la exactitud y la entropía cruzada categórica. La exactitud es una métrica para evaluar modelos de clasificación que nos muestra la proporción de las predicciones que el modelo realizó correctamente, como se observa en la función 4.2:

$$Exactitud = \frac{NPC}{NTP} \quad (4.2)$$

En donde NPC se refiere al número de predicciones correctas y NTP al número total de predicciones.

La exactitud también se puede calcular en términos de positivos y negativos cuando el problema es de clasificación binaria (función 4.3):

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.3)$$

En donde VP se refiere a verdaderos positivos, VN a verdaderos negativos, FP a falsos positivos y FN a falsos negativos.

La entropía cruzada categórica es una función de pérdida que se utiliza para optimizar los modelos y el objetivo es minimizar esta función. La entropía cruzada se define como (función 4.4):

$$EntropiaCruzada = - \sum_{i=1}^n t_i \log(p_i) \quad (4.4)$$

En donde t_i es la etiqueta de verdad y p_i es la probabilidad *Softmax* para la clase i .

Además, se obtiene la matriz de confusión que nos permite visualizar el desempeño del algoritmo, observando los aciertos y errores que tiene nuestro modelo. En donde las columnas de la matriz representan las predicciones de cada clase y las filas representan los valores observados. En la figura 4.7 se muestra una matriz de confusión binaria:

		Valores predicción	
Valores reales	Verdaderos positivos	Falsos negativo	
	Falsos positivo	Verdaderos negativos	

Figura 4.7: Matriz de confusión binaria.

Fuente: Elaboración propia.

En este trabajo los datos se dividen en el conjunto de entrenamiento (80 %) y el conjunto de validación (20 %). Lo anterior nos permite probar que el modelo está aprendiendo de los datos de entrenamiento y se generalizan a los datos de validación.

Se utilizan diferentes arquitecturas de redes neuronales convolucionales preentrenadas con *ImageNet* para realizar el aprendizaje por transferencia en *Python* (Van Rossum y Drake Jr, 1995), y se emplea *fast.ai* (Howard y Gugger, 2020) la cual es una biblioteca popular para agregar la funcionalidad de nivel superior a *PyTorch*. Mientras que *PyTorch* proporciona las operaciones básicas.

Las arquitecturas que se utilizan son las siguientes:

- *AlexNet* (Krizhevsky, 2014; Krizhevsky y cols., 2012).
- *ResNet50* (He y cols., 2016).
- *ResNet152*.
- *SqueezeNet1* (Iandola y cols., 2016).

- *VGG19* (Simonyan y Zisserman, 2015).

Al experimentar con las diferentes arquitecturas, los mejores resultados se obtienen con la *ResNet152*. Después, se prueba esta arquitectura con las distintas combinaciones de bandas de las imágenes satelitales:

- RGB.
- Falso color para zonas urbanas.
- Infrarrojo para vegetación.
- Uso del suelo.

Por último, para obtener mejores resultados se realiza la técnica de *pansharpening* o refinado pancromático sobre las imágenes al utilizar el método *HSV Sharpening*, el cual se basa en intercambiar el valor de los datos (la ‘v’ de HSV) por la contraparte de mayor resolución, mientras que el tono y la saturación (la ‘h’ y la ‘s’) se proporcionan por las imágenes de menor resolución. Con esto se mejoran a través de un proceso en donde se utiliza como referencia la banda pancromática que tiene mejor resolución (15 metros contra 30 metros). Esta técnica permite combinar la banda pancromática con las otras bandas de menor calidad y genera una nueva imagen multibanda a color y preserva la mayor nitidez disponible.

Capítulo 5

Evaluación

Para la evaluación cuantitativa de los resultados obtenidos, se utiliza el porcentaje de imágenes correctamente clasificadas y la entropía cruzada categórica en el conjunto de validación, el 20 % del conjunto total, y se prueban diferentes arquitecturas de redes neuronales en imágenes con las bandas RGB (roja, verde y azul). En el cuadro 5.1 se expone que los mejores resultados se obtienen con la arquitectura *ResNet152*, al tener una exactitud de 69.7% y una entropía cruzada categórica de 0.74.

	Exactitud	Entropía Cruzada Categórica
ResNet50	67.8 %	0.77
ResNet152	69.7 %	0.74
SqueezeNet1	63.8 %	0.83
VGG19	67.1 %	0.79
AlexNet	54.1 %	1.13

Cuadro 5.1: Exactitud y entropía cruzada categórica utilizando diferentes arquitecturas de redes neuronales e imágenes RGB.

Fuente: Elaboración propia.

Una vez que se elige la arquitectura con mejor desempeño, *ResNet152*, se evalúa con diferentes combinaciones de bandas. Se utilizan las bandas RGB (roja, verde y azul), zonas urbanas (infrarrojo de onda corta 2, infrarrojo de onda corta 1 y roja), vegetación (infrarrojo cercano, roja y verde) y uso de suelo (infrarrojo cercano, infrarrojo de onda corta 1 y roja). En el cuadro 5.2 se muestra que los mejores resultados se obtienen con la combinación de bandas de falso color para zona urbana, al tener una exactitud de 74.7% y una entropía cruzada categórica de 0.62.

	Exactitud	Entropía Cruzada Categórica
RGB	69.7 %	0.74
Zonas urbanas	74.7 %	0.62
Vegetación	67.0 %	0.95
Uso de suelo	65.6 %	0.93

Cuadro 5.2: Exactitud y entropía cruzada categórica utilizando la arquitectura ResNet152 y diferentes combinaciones de bandas.

Fuente: Elaboración propia.

Para finalizar, se realiza el procesamiento de refinado pancromático en las imágenes con combinación de bandas para zona urbana como se muestra en la figura 5.1.



Figura 5.1: Ejemplo de imágenes correspondientes a las bandas de falso color para zona urbana con y sin refinado pancromático.

Fuente: Elaboración propia.

En la evaluación utilizando la *ResNet152* y la combinación de bandas para zona urbana con refinado pancromático se observa que se obtienen mejores resultados añadiendo esta técnica a las imágenes (cuadro 5.3).

	Exactitud	Entropía Cruzada Categorica
Sin refinado pancromático	74.7%	0.62
Con refinado pancromático	86.6%	0.35

Cuadro 5.3: Exactitud y entropía cruzada categorica utilizando refinado pancromático.

Fuente: Elaboración propia.

Después, se obtiene la matriz de confusión sobre el conjunto de validación para el modelo que obtuvo mejores resultados, utilizando *ResNet152* con la combinación de bandas de falso color para zona urbana con refinado pancromático. Las columnas representan las imágenes que el modelo predijo como alta, baja y media marginación. Entonces, en la diagonal de la matriz se observa las que se clasificaron correctamente y las celdas fuera de la diagonal representan las que lo hicieron incorrectamente.

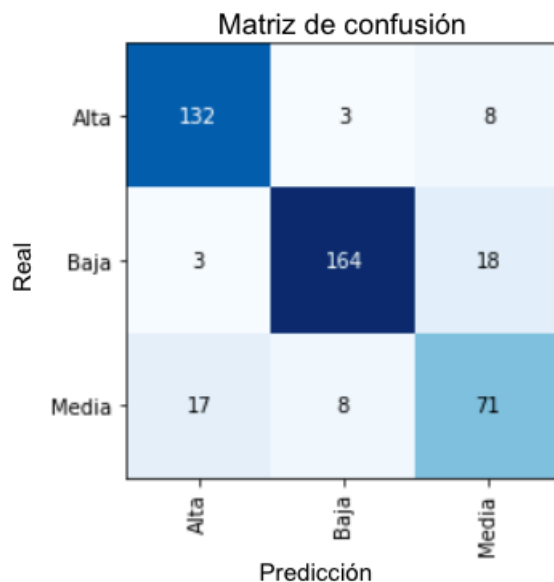


Figura 5.2: Matriz de confusión utilizando *ResNet152* y bandas de falso color para zona urbana con refinado pancromático.

Fuente: Elaboración propia.

En la figura 5.2 se muestra cómo se logra clasificar con éxito la alta y baja marginación en las imágenes satelitales y que cuando estas clases se clasifican de forma errónea, en su mayoría se clasifican como media marginación. En las figuras 5.3 y 5.4 podemos examinar un ejemplo de los resultados obtenidos.



Figura 5.3: Resultados obtenidos con la *ResNet152* y las bandas de falso color para zona urbana.

Fuente: Elaboración propia.

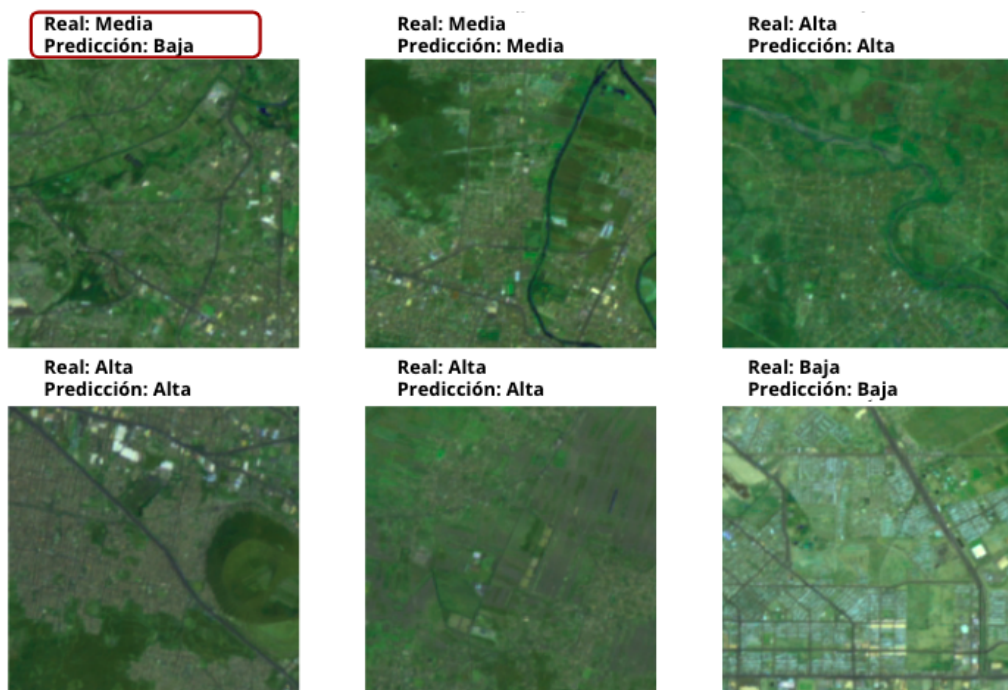


Figura 5.4: Resultados obtenidos con la *ResNet152* y las bandas de falso color para zona urbana.

Fuente: Elaboración propia.

Se muestran otras métricas (cuadro 5.4) y la curva ROC¹ (figura 5.5) para los resultados obtenidos utilizando *ResNet152* con la combinación de bandas de falso color para zona urbana con refinado pancromático.

¹Acrónimo de *Receiver Operating Characteristic*. Es una representación gráfica de la sensibilidad frente a la especificidad, o bien, de la proporción de verdaderos positivos frente a la de falsos positivos.

	Precisión	Exhaustividad	<i>F1-score</i>	Muestra
Alta	87 %	92 %	89 %	143
Baja	94 %	89 %	91 %	185
Media	73 %	74 %	74 %	96
Exactitud				87 %
Prom. macro	85 %	85 %	85 %	424
Prom. ponderado	87 %	87 %	87 %	424

Cuadro 5.4: Métricas utilizando *ResNet152* con la combinación de bandas de falso color para zona urbana con refinado pancromático. Fuente: Elaboración propia.

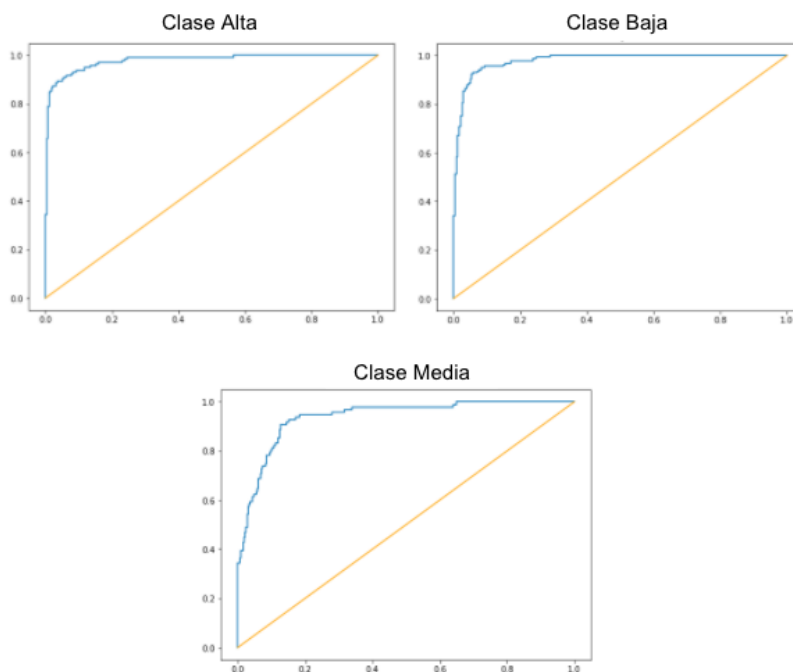


Figura 5.5: Curva ROC (uno contra el resto) utilizando *ResNet152* y bandas de falso color para zona urbana con refinado pancromático. Fuente: Elaboración propia.

5.1. Discusión de resultados

Se observa que a partir de una imagen satelital es posible extraer características con redes neuronales convolucionales y obtener información sobre el Grado de Marginación con una exactitud del 86.6 %².

Se utilizan redes neuronales convolucionales ya que son los algoritmos que mejor desempeño han mostrado en estas tareas. Asimismo, dada la naturaleza del problema es importante tomar en cuenta las bandas que componen la imagen satelital y su posición en la imagen, y las redes neuronales convolucionales son capaces de utilizar esta información para llegar al resultado.

Además, se logra clasificar con éxito la alta y baja marginación en las imágenes satelitales y cuando se clasifican de forma errónea, en su mayoría se clasifican como media marginación. Lo anterior, ya que la diferencia de características es mayor entre la alta marginación y baja marginación que respecto a la media marginación. Diferenciar estos niveles en las imágenes es un problema complejo incluso para el ser humano.

A nivel nacional, se lograron procesar para la detección del Grado de Marginación 2,198 imágenes en la zona urbana. Se clasifica solamente la zona urbana porque los datos del Grado de Marginación a nivel AGEB se encuentran disponibles para esta cobertura geográfica. Es importante resaltar que los datos utilizados, las imágenes satelitales y el Grado de Marginación, son datos gratuitos.

Las bandas de las imágenes del satélite Landsat 7 se encuentran disponibles en una resolución de 30 metros y la banda pancromática en una resolución de 15 metros. En el trabajo, se encuentra que

²Se obtiene esta exactitud utilizando ResNet152 y la combinación de bandas de falso color para zona urbana (infrarrojo de onda corta 2, infrarrojo de onda corta 1 y roja) con procesamiento de refinado pancromático.

la detección de zonas marginadas está relacionada con la resolución espacial de las imágenes satelitales, ya que cuando se utiliza el procesamiento de refinado pancromático se tiene un mejor desempeño del algoritmo. Para mejorar este desempeño se propone explorar con imágenes satelitales de mejor resolución.

Por último, es posible seleccionar bandas multiespectrales específicas que identifiquen mejor las áreas marginadas. El mejor desempeño se obtiene utilizando 3 bandas multiespectrales, la banda infrarroja de onda corta 2, la infrarroja de onda corta 1 y la roja.

Capítulo 6

Conclusiones

La investigación demuestra como las imágenes satelitales tienen un poder predictivo razonable para clasificar en alta, media y baja marginación la zona urbana de México. Además, se evidencia como las redes neuronales convolucionales tienen un buen desempeño en un entorno con datos de entrenamiento limitados.

Las imágenes satelitales utilizadas de las 32 entidades de la República Mexicana fueron tomadas del satélite Landsat 7 del año 2010, las cuales se obtuvieron de *Google Earth Engine* y para clasificar dichas imágenes utilizamos los datos del Grado de Marginación de CONAPO para zona urbana.

Además, se utilizan cinco arquitecturas de redes neuronales convolucionales preentrenadas con *ImageNet*: AlexNet (Krizhevsky, 2014), VGG19 (Simonyan y Zisserman, 2015), SqueezeNet1 (Iandola y cols., 2016), Resnet50 (He y cols., 2016) y Resnet152.

Con este enfoque se demuestra que las imágenes satelitales diurnas Landsat se pueden utilizar para hacer estimaciones bastante precisas sobre el grado de marginación en la zona urbana de México, permitiendo tener información con mayor frecuencia. Con esta

propuesta se obtienen, a partir de una imagen satelital de cualquier entidad federativa -las cuales están disponibles aproximadamente cada dos semanas- información sobre si en ese lugar determinado se tiene baja, media o alta marginación con una exactitud del 86.6 % al usar la red neuronal convolucional *ResNet152* y la combinación de bandas de falso color para zona urbana (infrarrojo de onda corta 2, infrarrojo de onda corta 1 y roja) con procesamiento de refinado pancromático.

Al tener esta información con mayor periodicidad permitirá a los formuladores de políticas públicas identificar a la población con mayores rezagos para incluirlos en los programas de desarrollo económico y social, y de esta manera, maximizar los recursos disponibles a fin de generar el mayor beneficio posible en la sociedad mexicana.

Asimismo, ya que se tiene ubicada geográficamente a la población por su nivel de marginación, se puede analizar en conjunto con estadísticas poblacionales para hacer política pública focalizada. Por lo tanto, es importante contar con herramientas eficientes para la detección automática de datos socioeconómicos en nuestro país.

Se propone como trabajo futuro, explorar combinaciones de bandas que ayuden a mejorar la predicción y realizar un mapa digital con los resultados obtenidos del Grado de Marginación a nivel AGEB para un año reciente, que sirva como fuente de información para todo el público y se integre con otras capas de información o con otras imágenes.

Bibliografía

- Albert, A., Kaur, J., y Gonzalez, M. C. (2017). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. En *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (p. 1357–1366). New York, NY, USA: Association for Computing Machinery. Descargado de <https://doi.org/10.1145/3097983.3098070> doi: 10.1145/3097983.3098070
- Bengio, Y. (2009). Learning deep architectures for ai. *Found. Trends Mach. Learn.*, 2(1), 1–127. Descargado de <https://doi.org/10.1561/22000000006> doi: 10.1561/22000000006
- Deng, J., Dong, W., Socher, R., Li, L., Kai Li, y Li Fei-Fei. (2009). Imagenet: A large-scale hierarchical image database. En *2009 ieee conference on computer vision and pattern recognition* (p. 248-255).
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. MIT Press. (<http://www.deeplearningbook.org>)
- He, K., Zhang, X., Ren, S., y Sun, J. (2016). Deep residual learning for image recognition. En *2016 ieee conference on computer vision and pattern recognition (cvpr)* (p. 770-778).

- He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., y Li, M. (2018). *Bag of tricks for image classification with convolutional neural networks*.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., y Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, *abs/1207.0580*. Descargado de <http://arxiv.org/abs/1207.0580>
- Howard, J., y Gugger, S. (2020). *Deep learning for coders with fastai and pytorch*. O'Reilly Media. Descargado de <https://books.google.com.mx/books?id=yATuDwAAQBAJ>
- Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., y Keutzer, K. (2016). *Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5mb model size*.
- Iglovikov, V., Mushinskiy, S., y Osin, V. (2017). Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv preprint arXiv:1706.06169*.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., y Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. Descargado de <https://science.sciencemag.org/content/353/6301/790> doi: 10.1126/science.aaf7894
- Kadhim, M., y Abed, M. (2020, 01). Convolutional neural network for satellite image classification. En (p. 165-178). doi: 10.1007/978-3-030-14132-5_13
- Konkle, T., Brady, T. F., Alvarez, G., y Oliva, A. (2010). Scene memory is more detailed than you think. *Psychological Science*, *21*, 1551 - 1556.
- Krizhevsky, A. (2014). *One weird trick for parallelizing convolutional neural networks*.
- Krizhevsky, A., Sutskever, I., y Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. En *Proceedings of the 25th international conference on neural information processing systems - volume 1* (p. 1097–1105).

- Red Hook, NY, USA: Curran Associates Inc.
- Lavin, A., y Gray, S. (2016). Fast algorithms for convolutional neural networks. En *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (p. 4013-4021).
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., y Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 541-551.
- LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., y Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. En D. S. Touretzky (Ed.), *Advances in neural information processing systems 2* (pp. 396-404). Morgan-Kaufmann. Descargado de <http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf>
- O'Shea, K., y Nash, R. (2015). An introduction to convolutional neural networks. *CoRR*, abs/1511.08458. Descargado de <http://arxiv.org/abs/1511.08458>
- Pelletier, C., Webb, G. I., y Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5), 523.
- Robinson, C., Hohman, F., y Dilkina, B. (2017, 11). A deep learning approach for population estimation from satellite imagery. En (p. 47-54). doi: 10.1145/3149858.3149863
- Simonyan, K., y Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*.
- Sun, J., Di, L., Sun, Z., Wang, J., y Wu, Y. (2020). Estimation of gdp using deep learning with npp-viirs imagery and land cover data at the county level in conus. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1400-1415. doi: 10.1109/JSTARS.2020.2983331
- Van Rossum, G., y Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

- Vermote, E., ElSaleous, N., Justice, C., Kaufman, Y., Privette, J., Remer, L., ... Tanre, D. (1997, 07). Atmospheric correction of visible to middle-infrared eos-modis data over land surfaces: Background, operational algorithm and validation. *Journal of Geophysical Research-Atmospheres*, 102, 17131-17141. doi: 10.1029/97jd00201
- Wu, P., y Tan, Y. (2019). Estimation of economic indicators using residual neural network resnet50. En *2019 international conference on data mining workshops (icdmw)* (pp. 206–209).
- Wu, P., y Tan, Y. (2020). Estimation of poverty based on remote sensing image and convolutional neural network. *Advances in Remote Sensing*, 8(4), 89–98.
- Wurm, M., Stark, T., Zhu, X. X., Weigand, M., y Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 59 - 69. Descargado de <http://www.sciencedirect.com/science/article/pii/S0924271619300383> doi: <https://doi.org/10.1016/j.isprsjprs.2019.02.006>
- Xie, M., Jean, N., Burke, M., Lobell, D., y Ermon, S. (2016). Transfer learning from deep features for remote sensing and poverty mapping. En *Proceedings of the thirtieth aaii conference on artificial intelligence* (p. 3929–3935). AAAI Press.
- Zeiler, M. D., y Fergus, R. (2014). Visualizing and understanding convolutional networks. En D. Fleet, T. Pajdla, B. Schiele, y T. Tuytelaars (Eds.), *Computer vision – eccv 2014* (pp. 818–833). Cham: Springer International Publishing.
- Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., y Oliva, A. (2014). Learning deep features for scene recognition using places database. En Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, y K. Q. Weinberger (Eds.), *Advances in neural information processing systems 27* (pp. 487–495). Curran Associates, Inc. Descargado de <http://papers.nips.cc/>

paper/5349-learning-deep-features-for-scene
-recognition-using-places-database.pdf