



UNIVERSIDAD NACIONAL DE COLOMBIA

Aplicación de una red neuronal convolucional para la predicción de mallas de población en el área metropolitana de Bogotá

Aníbal Montero Leguizamón

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2021

Aplicación de una red neuronal convolucional para la predicción de mallas de población en el área metropolitana de Bogotá

Aníbal Montero Leguizamón

Trabajo final presentado como requisito parcial para optar al título de:
Magíster en Ingeniería de Sistemas y Computación

Director:
Ph.D. Luís Fernando Niño Vásquez

Línea de Investigación:
Sistemas Inteligentes
Grupo de Investigación:
Laboratorio de Investigación en Sistemas Inteligentes - LISI

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2021

Un hombre del pueblo de Neguá, en la costa de Colombia, pudo subir al alto cielo.

A la vuelta contó. Dijo que había contemplado desde arriba, la vida humana.

Y dijo que somos un mar de fueguitos.

-El mundo es eso- revelo -un montón de gente, un mar de fueguitos.

Cada persona brilla con luz propia entre todas las demás.

No hay dos fuegos iguales. Hay fuegos grandes y fuegos chicos y fuegos de todos los colores. Hay gente de fuego sereno, que ni se entera del viento, y gente de fuego loco que llena el aire de chispas. Algunos fuegos, fuegos bobos, no alumbran ni queman; pero otros arden la vida con tanta pasión que no se puede mirarlos sin parpadear, y quien se acerca se enciende.

El Mundo. Tomado de *El libro de los abrazos*

Escrito por Eduardo Galeano

Agradecimientos

Una vez más, la Universidad Nacional de Colombia ha puesto luz a mis propósitos, me ha brindado un segundo hogar y me ha hecho un orgulloso partícipe de su institución.

Quiero agradecer al profesor Luís Fernando Niño por guiarme, desde mi primera semana de clases de maestría, hasta mis últimos espacios con la Universidad en el grupo de investigación, rumbo al conocimiento con el que seguiré abriendo puertas.

Agradezco también al Grupo de investigación y desarrollo geoestadístico del Departamento Administrativo Nacional de Estadística (DANE) por el espacio laboral que me motivó a realizar esta investigación.

A mi familia, y en especial a mi madre, Nelly, a quienes les debo su permanente atención y la inspiración para continuar mi formación académica.

Resumen

Aplicación de una red neuronal convolucional para la predicción de mallas de población en el área metropolitana de Bogotá

Saber cuántas personas viven en un área determinada y saber en dónde habitan específicamente, son preguntas tradicionalmente abordadas desde la Demografía. El presente trabajo plantea la alternativa de utilizar imágenes satelitales para predecir el número de habitantes a partir de mallas de población. Se extrajo un conjunto de imágenes de Landsat 8, a partir de un diseño muestral proporcional al tamaño (PPS) aplicado sobre una malla de población censal del 2018 en Colombia. Se entrenó la arquitectura LeNet-5 modificada para realizar regresión sobre número de habitantes por celda a partir del conjunto de imágenes obtenido. Se realizaron pruebas del modelado sobre una submuestra de la malla de población en Colombia y sobre la malla de población correspondiente a los municipios que componen el área metropolitana de Bogotá en 2018, arrojando MAEs de 947,8 y 1.181,9, respectivamente, igualando e incluso superando los resultados encontrados en el estado del arte.

Palabras clave: Malla de Población, Sensores Remotos, Aprendizaje Profundo, Redes Neuronales Convolucionales.

Abstract

Predicting a population grid in the Bogotá metropolitan area, based on convolutional neural networks

Knowing how many people live in an area and knowing where they live specifically are questions commonly approached through Demography. The present work proposes the using of satellite images to predict the number of inhabitants based on population grids as an alternative approach. A Landsat 8 images dataset was generated using a Probability Proportional to Size (PPS) sample extracted on a 2018 census population grid in Colombia. A LeNet-5 architecture was modified to predict the number of inhabitants per cell and trained with the previous image dataset obtained. The trained model was tested with a subsample of the population grid in Colombia and the population grid corresponding to the municipalities of the Bogotá metropolitan area in 2018. The model reached MAEs of 947.8 and 1181.9, respectively. These results equal and even exceed the performance found in the state of the art.

Keywords: Population grid, Remote Sensing, Deep Learning, Convolutional Neural Networks.

Este Trabajo Final de Maestría fue calificado en septiembre de 2021 por el siguiente evaluador:

Fabio Augusto González Osorio Ph.D.
Profesor Departamento de Ingeniería de Sistemas e Industrial
Facultad de Ingeniería
Universidad Nacional de Colombia

Contenido

Agradecimientos	vii
Resumen	ix
Listas de abreviaturas y símbolos	xvii
1. Introducción	1
2. Antecedentes	3
2.1. Predicción de población a partir de Datos de Sensores Remotos	3
2.2. Población representada en geometría de puntos y mallas	5
2.3. <i>Deep Learning</i> y Redes Neuronales Convolucionales	6
2.3.1. Predicción de población con Redes Neuronales Convolucionales	7
3. Objetivos	9
3.1. Objetivo general	9
3.2. Objetivos específicos	9
4. Fuentes de datos	10
4.1. Malla de población del censo de población del 2018 en Colombia	10
4.2. Google Earth Engine	11
4.2.1. Proyecto Landsat 8	12
5. Metodología	14
5.1. Obtención de imágenes satelitales	14
5.1.1. Muestreo de celdas sobre la malla de población	14
5.1.2. Descarga de imágenes por celda en Google Earth Engine	17
5.1.2.1. Método de composición de imágenes por celda	17
5.1.2.2. Automatización de descarga	18
5.1.2.3. Verificación de descarga	19
5.2. Preprocesamiento	20
5.2.1. Normalización	20
5.2.2. Aumento de datos	21
5.2.3. Conjunto de datos obtenido	21

5.3. Modelado	22
5.3.1. Experimentación	24
5.3.2. Implementación final	25
5.3.3. Validación del modelo final	26
5.3.3.1. Herramientas tecnológicas utilizadas	27
6. Resultados	29
6.1. Comparativa en el desempeño de predicción	29
6.2. Análisis exploratorio de los residuales	31
6.2.1. Mapeo de los residuales para el área metropolitana de Bogotá en 2018	31
6.2.2. Exploración de la predicción en casos particulares de imágenes	33
6.3. Predicción de población en el área metropolitana de Bogotá	35
6.3.1. Comparación de las predicciones con las series de proyecciones de po- blación del DANE	36
7. Conclusiones y recomendaciones	39
7.1. Conclusiones	39
7.2. Recomendaciones	40
A. Anexo: Productos generados	42
Bibliografía	43

Lista de Figuras

1-1. Desagregación de proyecciones en superficies de densidad de población [1]	1
2-1. Densidad de población en Europa, representada en malla (izquierda) y en división político-administrativa [2, 3]	5
2-2. Arquitectura VGG modificada por Doupe et al. [4]	8
2-3. Arquitectura VGG modificada por Robinson et al. [4]	8
4-1. Histograma y diagrama de Pareto de la distribución de la población del CNPV2018 en las celdas de la malla del DANE	11
4-2. Cuadrícula de Landsat 8 en Colombia. Tomada de http://arcg.is/0vWm5K	12
5-1. Distribución muestral del numero de celdas según su número de habitantes	15
5-2. Distribución espacial de la muestra de celdas sobre el territorio terrestre colombiano. Elaboración propia	16
5-3. Interfaz gráfica de la aplicación web de GEE	17
5-4. Formulario emergente para confirmar la descarga de las imágenes individuales desde GEE	18
5-5. Verificación de la coincidencia entre la visualización: con una de las bandas del espectro visible en Python, con la imagen desplegada en el panel de Google Maps y con la ubicación de la celda sobre una capa de información de OpenStreetMap	20
5-6. Obtención de ocho imágenes a partir de una, aplicando espejos y giros de 90, 180 y 270 grados	21
5-7. Separación del conjunto de datos de imágenes para realizar entrenamiento, validación y prueba	22
5-8. Arquitectura VGG-A [5]	23
5-9. Arquitectura LeNet-5 [6]	24
5-10. Evolución de la función de costo MAE durante el entrenamiento de la red LeNet-5 luego de la afinación de los hiperparámetros	26
6-1. Diagrama de dispersión entre número de habitantes y el residual obtenido, evaluado en el conjunto de prueba del área metropolitana de Bogotá del 2018	31
6-2. Mapeo de los residuales sobre el área metropolitana de Bogotá de 2018	32

6-3. Ejemplos de predicción en casos con residuales pequeños, casos con sobre-estimación y casos con sub-estimación	34
--	----

Lista de Tablas

2-1. Evolución de las características de las imágenes aéreas	4
5-1. Comparación de las arquitecturas CNN utilizadas	25
6-1. Comparativa de las métricas de desempeño predictivo del modelo obtenido contra los reportes de los trabajos de referencia	30
6-2. Comparativa de las métricas de desempeño predictivo del modelo obtenido por rangos en el subconjunto de validación para Colombia en 2018	30
6-3. MAE para el área metropolitana de Bogotá en 2018, desagregado por municipio	35
6-4. Predicciones por municipio y periodo en el área metropolitana de Bogotá comparadas con las proyecciones de población del DANE	37
6-5. Diferencia entre las predicciones por municipio y período en el área metropo- litana de Bogotá y las proyecciones de población del DANE	37

Lista de abreviaturas y símbolos

Lista de abreviaturas y acrónimos

Abreviatura Término

API	<i>Application Programming Interface</i>
CNN	<i>Convolutional Neural Network</i>
CNPV2018	Censo Nacional de Población y Vivienda del 2018, Colombia
DANE	Departamento Administrativo Nacional de Estadística
DL	<i>Deep Learning</i>
GEE	Google Earth Engine
INE	Institución Nacional de Estadística
Eurostat	Oficina Europea de Estadística
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MedianAE	<i>Median Absolute Error</i>
PPS	<i>Probability Proportional to Size sampling</i>
RSD	<i>Remote Sensing Data</i>
RMSE	<i>Root Mean Square Error</i>
%RMSE	<i>Relative Root Mean Square Error</i>
VGG	<i>Visual Geometry Group</i>

Lista de símbolos

Símbolo Término

<i>hab</i>	Número de habitantes
<i>km</i>	Kilómetro(s)
r^2	Coefficiente de determinación

1. Introducción

El presente trabajo ha sido motivado por la exploración de alternativas al modelado para predecir el número de habitantes en diferentes áreas de Colombia. Una de estas alternativas consiste en utilizar fuentes de información gratuitas que tomen registro de la actividad humana, como lo son algunos proyectos estatales de captura de Datos de Sensores Remotos (RSD, por su sigla en inglés); Landsat [7, 8] y Sentinel [9] son ejemplos de estos proyectos.

Existen dos problemas que se han abordado tradicionalmente desde la Demografía: saber cuántas personas viven en un área determinada [4] y saber dónde habitan específicamente esas personas [4, 10]. El uso de imágenes aéreas ha permitido predecir la población, resolviendo esas dos preguntas con un mismo enfoque [4].

Frente a esta necesidad de información sobre la distribución territorial de la población, las Instituciones Nacionales de Estadística (INEs), tales como el DANE, ofrecen públicamente las proyecciones de población ¹ desagregadas de acuerdo a las divisiones político-administrativas del respectivo país (Figura 1-1). Sin embargo, el modelado demográfico utilizado para este propósito no tiene el alcance ni el detalle geográfico de un censo poblacional [11], pues solo proporciona recuentos de población en áreas amplias [4].

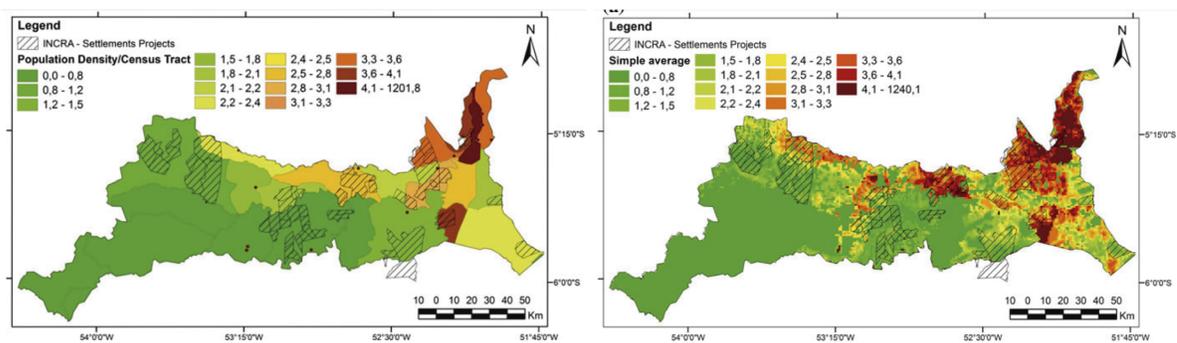


Figura 1-1.: Desagregación de proyecciones en superficies de densidad de población [1]

Durante estos últimos años se ha explorado la posibilidad de realizar regresión de la población con el uso de arquitecturas de Redes Neuronales Convolucionales (CNN, por su sigla en

¹<https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>

inglés) a partir de imágenes de satélite en áreas de estudio específicas. Doupe et al. [10] así como Robinson et al. [4] abordaron estas problemáticas de manera exploratoria, adaptando la estrategia de ajustar las arquitecturas VGG (*Visual Geometry Group*) para que predigan la población que se encuentra habitando en áreas donde se capturaron y preprocesaron mosaicos multi-espectrales construidos con RSD.

Los dos trabajos anteriormente mencionados motivaron a continuar la investigación sobre la aplicación de esta estrategia de predicción, en donde se aprovechan los datos censales sobre la ubicación de la población con los RSD que coincidan espacial y temporalmente. En la sección 2 se muestran las generalidades de estas investigaciones, así como la motivación de estos trabajos en aplicar técnicas de *Deep Learning* (DL) en los RSD.

En la sección 3 se definen los objetivos que fueron previamente establecidos y que guiaron la presente investigación.

La sección 4 mostrará una descripción de los datos y sus respectivas fuentes utilizadas como insumo para construir el conjunto de datos con el que se entrena y evalúa el modelo de predicción implementado: la malla de población del Censo Nacional de Población y Vivienda del 2018 (CNPV2018), así como los conjuntos de datos preprocesados de Landsat 8.

El desarrollo del presente trabajo de investigación se reseña detalladamente en la sección 5, iniciando por el manejo de los datos espaciales de la malla de población, la recolección de las imágenes preprocesadas vía aplicación web, el preprocesamiento y la construcción del conjunto de datos de esta investigación.

Por último, en la sección 6 se muestran y comparan los resultados obtenidos. Asimismo, se discuten las conclusiones y recomendaciones que se generaron luego de obtener los resultados cuantificados del modelado y de todo el proceso de investigación.

2. Antecedentes

Los primeros acercamientos al uso de imágenes aéreas para la predicción del número de habitantes en zonas particulares, utilizaron el registro fotográfico para contar el número de construcciones y así inferir el total de habitantes habiendo estimado el promedio de habitantes por edificación [12, 13], incluso utilizando estratificación [14]. Con la disposición a diferentes capas de información espacial y RSDs con escalas mucho más pequeñas, se exploró el uso de métodos de regresión sobre descriptores geográficos generados a partir de aquellas capas que están ligadas al uso del suelo y la actividad humana [15]. En la última década, se ha optado por la generación de mapas de densidad de población a partir de modelos dasimétricos que redistribuyen las proyecciones de población y que originalmente se encuentran a escala político-administrativa [1, 16].

Los modelos dasimétricos aprovechan como ventajas el escalamiento de la información en áreas particulares y, también, abren la posibilidad de utilizar múltiples fuentes de información auxiliar sobre la distribución espacial de la población [16, 17]. Sin embargo, la validación de los modelos dasimétricos se convierte en una tarea fuera del alcance de sus metodologías, dado que su ajuste es ad-hoc y netamente matemático, además, los datos reales sobre la verdad fundamental no existen [4]. Frente a esta situación, se ha explorado recientemente el uso de regresiones con CNNs que permitan realizar la validación estadística de las predicciones obtenidas [10, 4]. El presente trabajo busca explorar la aplicación de esta estrategia en el ámbito nacional y local.

A continuación se describen brevemente las capas de información telecapturadas que han sido utilizadas como información auxiliar para realizar el modelado de la cantidad de habitantes, la descripción de las fuentes de datos sobre población que han sido relacionadas a la actividad humana evidenciada en los RSDs y, finalmente, se justifica la relevancia de utilizar técnicas de DL y más específicamente las CNN.

2.1. Predicción de población a partir de Datos de Sensores Remotos

En relación a la estrategia de conteo de número de edificaciones como parte de la información que permite estimar la población, se compiló la Tabla **2-1**, en donde se relacionan diferentes

estudios, encontrados mediante revisión literaria, con la respectiva resolución espacial de cuadrícula de población obtenida, la plataforma fuente de las imágenes satelitales, la franja del espectro electromagnético que fue utilizado y el modelado empleado.

Esta tabla permite observar la evolución a lo largo del tiempo de las aplicaciones encontradas en la revisión literaria que soporta esta investigación. Se observa como la resolución de las imágenes utilizadas se hace cada vez más fina.

Tabla 2-1.: Evolución de las características de las imágenes aéreas

Autores	Año de estudio	Resolución espacial	Plataforma	Espectro	Modelo
Horton [18]	1974	1:50000	Fotografía aérea	Luz visible	Conteo
Curis y Barret [19]	1986	1:25000	Fotografía aérea	Luz visible	Densidad
Taragi, Bisht y Sokhi [13]	1988	1:10000	Fotografía aérea	Luz visible	Conteo
Zhang [15]	2003	1:25000	Fotografía aérea	Infrarrojo	Regresión
Amaral, Gavlak y Escada [1]	2012	1km x 1km	Landsat	Infrarrojo, Luz visible	<i>Fuzzy inference</i>
Karume et al. [14]	2017	200m x 200m	GeoEye	Luz visible	Conteo
Ayila, Babatunde y Yohanna [20]	2018	0.6m x 0.6m	QuickBird	Infrarrojo, Luz visible	Conteo

En cuanto a las capas de información auxiliar que evidencian las actividades humanas, Wang et al. [12] resaltan en las dos últimas décadas el uso de capas de información geográfica sobre topografía, altimetría, luz nocturna, temperatura y datos de cobertura del suelo. Anotan que el modelamiento de la población con imágenes de sensores remotos aprovecha al máximo su variedad y la resolución espacial que las fuentes ofrecen.

En cuanto a la variedad, son referidos los diferentes proyectos de acceso libre y licenciados, con sensores pasivos y activos, resoluciones espectrales propias del espectro infrarrojo o de la luz visible correspondientes a los sensores pasivos, de los LiDAR y los RADAR. Zhang [15] afirma que el espectro infrarrojo permite distinguir entre edificios residenciales y no residenciales. Esto le permitió segmentar automáticamente los polígonos de los edificios residenciales según su número de pisos en la ciudad de Tianjin en el año 1990, finalmente, construyó un modelo de regresión con descriptores geográficos por unidades de área. Amaral, Gavlak y Escada [1] usaron el método dasimétrico junto con inferencia difusa, presentando una solución en la que llevan los datos de la población de Maraná a una resolución más fina proporcionada por fuentes de datos espaciales auxiliares que dan cuenta de la distribución de esta población. Por lo cual, las imágenes de sensores remotos son una clara opción de información auxiliar [16, 17]. Sorichetta et al. [16] entrenan un modelo dasimétrico con *Random Forest* (bosque aleatorio) a nivel de unidades administrativas con diferentes capas sobre descriptores de la población, así como información del clima y del uso del suelo para luego utilizar la granularidad de la resolución de estas capas de covariables y redistribuyendo la población de las mismas unidades administrativas.

2.2. Población representada en geometría de puntos y mallas

Una malla de población, de acuerdo a la Oficina Europea de Estadística (Eurostat), es una rejilla o cuadrícula (también llamada retícula) compuesta por celdas que se superponen a un territorio que recopila datos del número de habitantes. Es una herramienta para analizar las interrelaciones entre las actividades humanas y el medio ambiente a partir de la distribución espacial de la población [2] y analizar fenómenos de forma independiente a los límites administrativos [3].

Un modelado demográfico arroja cifras de población por unidades administrativas, lo que significa que este modelado no tiene el alcance ni el detalle geográfico de un censo poblacional [11], pues solo proporciona recuentos de población en áreas amplias [4].

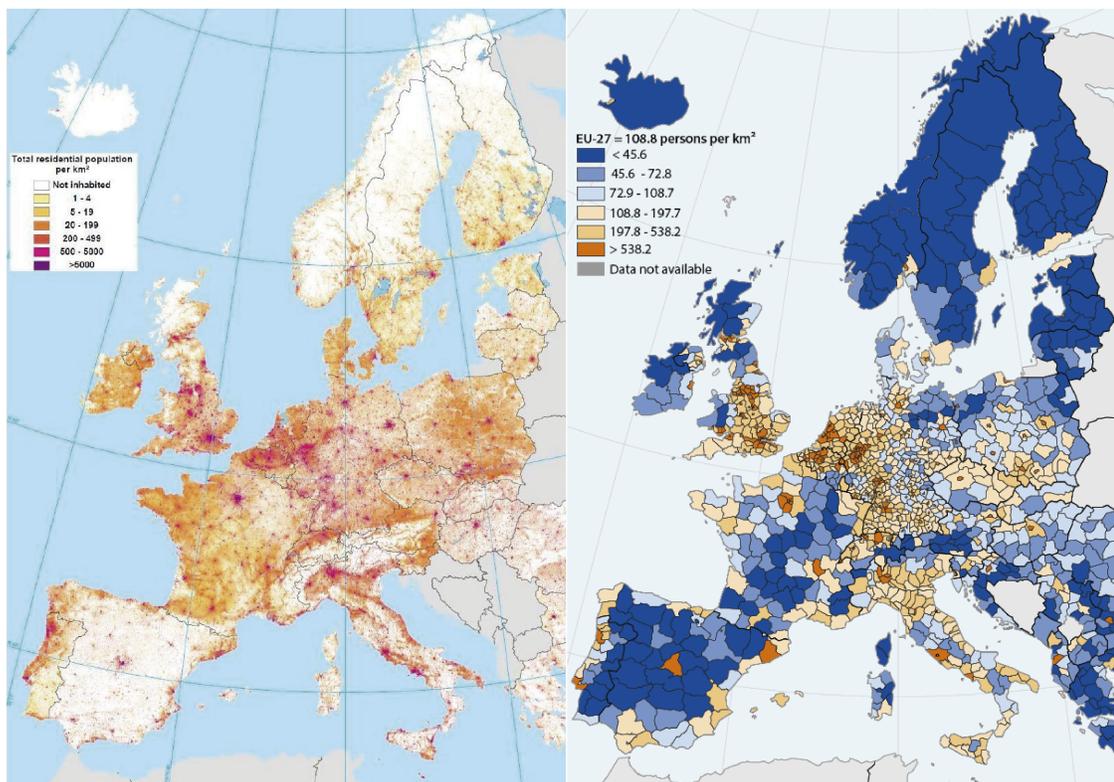


Figura 2-1.: Densidad de población en Europa, representada en malla (izquierda) y en división político-administrativa [2, 3]

Tal y como se citó en la introducción, se recalca la importancia de saber, al mayor detalle posible, dónde vive la población y el número de habitantes por área, siendo esto un componente fundamental de muchos procesos de toma de decisiones tales como desarrollo urbano,

contención de enfermedades infecciosas, planificación de evacuaciones, gestión de riesgos y planificación de conservación del ambiente [4]. Las estimaciones precisas de una población en riesgo de enfermedad requieren datos poblacionales precisos y geográficamente explícitos [10]. El tamaño y la distribución de la población sirven como determinantes clave de la asignación de recursos para el estado y gobiernos locales, por lo que las estimaciones de población son fundamentales en las decisiones de cuándo y dónde construir instalaciones públicas, incluidas escuelas, bibliotecas, plantas de tratamiento de aguas residuales, hospitales e infraestructura de transporte [12]. Dichos conjuntos de datos de población son vitales para medir los impactos del crecimiento de la población, monitorear el cambio y planificar intervenciones [16, 17, 21], tener una conciencia más completa del medio ambiente global y como cambia [22]. Los datos a nivel nacional siempre oscurecen una importante heterogeneidad sub-nacional, por lo cual, los mapas de densidad y las mallas de población de alta resolución son un componente crítico para los esfuerzos de desarrollo global, incluida la prestación de servicios, la asignación de recursos y la respuesta a desastres [10].

En la figura **2-1** se aprecia, en forma gráfica, la diferencia entre los mapas generados a partir de cifras agregadas por unidades político-administrativas y la desagregación sobre una malla cuadrículada de población en el continente europeo [2].

2.3. Deep Learning y Redes Neuronales Convolucionales

Para mostrar la pertinencia del uso de las CNN para realizar procesamiento de datos de sensores remotos, Yuan et al. [23] explican que durante la última década, el marco del DL ha superado a los modelos tradicionales con una mejora considerable en el rendimiento. Se ha observado un progreso sustancial en el desarrollo de una metodología de aprendizaje automatizado para una variedad de aplicaciones de las ciencias de la Tierra. Muestran su potencial en cartografía de la cobertura terrestre, recuperación de parámetros ambientales, la fusión y reducción de datos, y la reconstrucción y predicción de la información. El advenimiento de las imágenes de satélite a escala global nos proporciona una oportunidad sin precedentes de crear mapas de población precisos, homogéneos y rápidamente actualizados [10].

Al no haber limitaciones en la obtención de los datos, como es el caso de la teledetección, la atención se centra en la capacidad de extraer información valiosa de estos en forma automática [24]. Las técnicas de aprendizaje automático son una herramienta eficaz para extraer información de datos intrínsecamente desestructurados como lo son los datos de detección remota, proporcionando soluciones efectivas a problemas sociales [4]. Un caso como el de la aplicación de las arquitecturas CaffeNet y GoogleNet que realiza Castelluccio et al. [24] sobre el conjunto de datos de UC Merced y las precisiones mayores del 90 % obtenidas en la clasificación del uso del suelo a partir de imágenes aéreas, indican el buen desempeño del

modelado de imágenes que tienen las CNN. Las características susceptibles a ser detectables por una CNN dentro de una imagen, como lo son las líneas rectas de las redes de transporte y los edificios, están asociadas con una mayor densidad de población. Las regiones rurales como las montañas tienen una estructura espacial menos rígida y generalmente están asociadas con una menor densidad de población [10]. De acuerdo a la evidencia del trabajo con imágenes de gran escala, hecho por Simonyan y Zisserman [25], se puede mejorar el desempeño predictivo de las CNN utilizando filtros pequeños, es decir, de 3×3 , junto con muestreos o pooling de tamaño 2×2 , procurando aumentar en número de capas convolucionales en la arquitectura. Se menciona además que otras alternativas pueden ser aumentar la resolución de las imágenes de entrada e incluir más capas densas dentro de las arquitecturas. Simonyan y Zisserman mencionan que el poder de la arquitectura VGG se encuentra en su simplicidad, utiliza filtros con un campo receptivo pequeño de 3×3 , en lugar de utilizar filtros más grandes (5×5 o 7×7 , como en Alexnet) y porque tiene el mismo tamaño en mapa de características y número de filtros en cada capa convolucional del mismo bloque, aumentando el tamaño de las entidades en las capas más profundas, duplicándolas aproximadamente, después de cada capa de agrupación máxima.

Basado en las observaciones sobre la experimentación que realizaron los dos autores anteriormente señalados, es importante mencionar que para esta investigación se optó por empezar a experimentar con arquitecturas que tuviesen tamaños de filtro grande en relación a lo que ellos mencionan. lo cual se describirá con mayor detalle en la sección Metodología y su subsección Modelado.

2.3.1. Predicción de población con Redes Neuronales Convolucionales

Previo a esta investigación, la respectiva revisión bibliográfica arrojó dos publicaciones en las que se basó el modelado que cumple con el propósito y la promesa de valor de este trabajo.

Doupe et al. [10] utilizaron datos de los censos de población de 2002 en Tanzania y de 2009 en Kenia junto con composiciones de imágenes de Landsat 7, donde se logró construir un conjunto de datos de 145.649 observaciones para entrenar y validar la aplicación de una arquitectura VGG modificada. La figura **2-2** muestra la modificación de la arquitectura VGG aplicada, en donde se eliminaron los dos últimos bloques convolucionales, modificando también las capas de entrada y de salida para adecuar la arquitectura para realizar regresión, donde la entrada son las imágenes satelitales y la salida es el número de habitantes relacionada a esa imagen.

Por su parte, Robinson et al. [4] utilizaron imágenes de Landsat 7 coincidentes con las mallas de población tomadas a partir de diferentes encuestas, censos y proyecciones de población

entre los años 2000 y 2010, donde además se muestreó espacialmente sobre una malla de 2.599×5.769 celdas sobre el territorio continental de Estados Unidos. La modificación de la arquitectura VGG aplicada se modificó únicamente las capas de entrada y de salida del mismo modo que el caso anterior, pero se resalta que estos autores modelaron con la salida categorizada de acuerdo a la cantidad de habitantes correspondiente a la respectiva imagen satelital (figura 2-3).

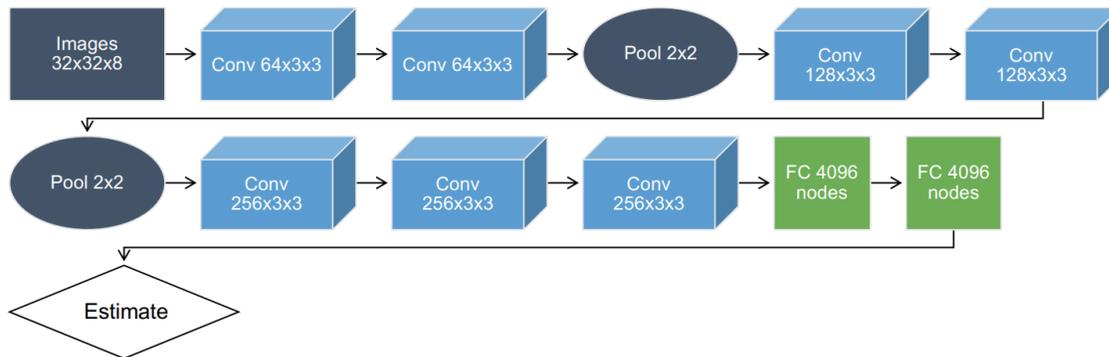


Figura 2-2.: Arquitectura VGG modificada por Doupe et al. [4]

Los resultados del presente trabajo son comparados a través de las métricas propuestas por los trabajos de Doupe et al. y Robinson et al., en la sección Resultados es su subsección Comparativa del desempeño de predicción.

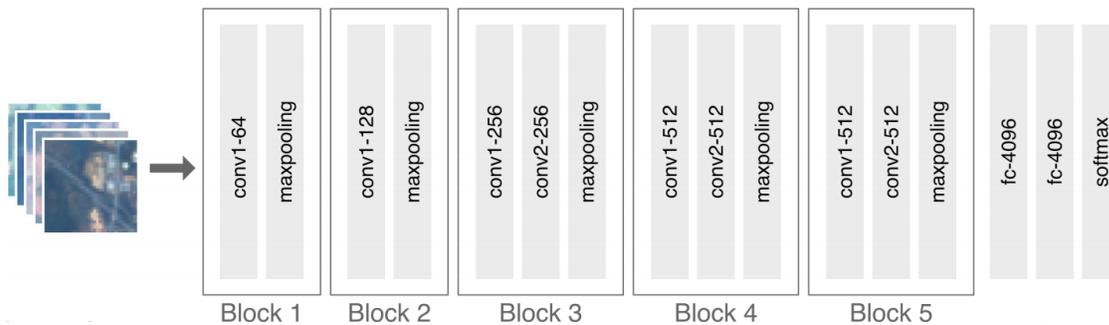


Figura 2-3.: Arquitectura VGG modificada por Robinson et al. [4]

3. Objetivos

Basándose en los trabajos de Doupe et al. y Robinson et al. y ante la pregunta de investigación planteada: *¿Cómo pueden las CNN predecir mallas de población en el contexto colombiano a partir de RSD?*, se plantearon los siguientes objetivos para la investigación.

3.1. Objetivo general

Aplicar una red neuronal convolucional que permita predecir mallas de población para área metropolitana de Bogotá, entrenada a partir de datos censales de población e imágenes de sensores remotos de acceso libre del 2018.

3.2. Objetivos específicos

1. Construir un conjunto de datos con composiciones de mosaicos de imágenes de sensores remotos para una malla de población de 2018 que sirvan para el entrenamiento de las arquitecturas de red neuronal convolucional.
2. Implementar arquitecturas de red neuronal convolucional en entornos de desarrollo y paquetes de aprendizaje profundo de uso libre para realizar predicciones de los datos de la malla de población respectiva.
3. Evaluar métricas de desempeño de los modelos obtenidos mediante validación estadística, haciendo la valoración de su aplicabilidad a diferentes niveles geográficos.

4. Fuentes de datos

Esta investigación utilizó dos fuentes de información secundaria. La primera fue la malla de población en cuadrícula de un kilómetro (1km) de resolución, generada a partir de los datos georreferenciados del CNPV2018 del DANE, los cuales se encuentran libremente disponibles en su plataforma del Geoportal¹; La segunda fue la plataforma de cómputo en la nube de GEE² donde se encuentra el acceso libre a las imágenes con corrección radiométrica y atmosférica para la superficie terrestre colombiana a partir del año 2013 correspondientes al proyecto Landsat 8³. A continuación, se describen en detalle estas dos fuentes de datos.

4.1. Malla de población del censo de población del 2018 en Colombia

Este primer conjunto de datos representa la información sobre la cantidad de habitantes y su respectiva ubicación relacionada a una malla cuadrículada que se extiende sobre todo el territorio terrestre colombiano en el año 2018.

Esta malla de población es una agregación del número de habitantes en celdas cuadradas de (1) kilómetro cuadrado. Los datos sobre número de viviendas, hogares y personas se encuentran originalmente georreferenciada en geometría de puntos, puesto que los datos se capturan durante el operativo censal a través de un Dispositivo Móviles de Captura (DMC) del que toma, además, la posición puntual de la vivienda donde se llena el formulario. El DANE ha dispuesto estos datos para el uso del público general agregados al nivel de la malla de población. Esta agregación la realiza el DANE con el propósito de anonimizar la ubicación de los habitantes con baja densidad de población [26].

Los datos se descargan en un archivo en formato *shapefile* (.shp), los cuales se encuentran estructurados en una tabla, los registros representan las celdas de la malla. De esta forma, la tabla descargada contiene dos elementos relevantes: la geometría georreferenciada de las

¹<https://geoportal.dane.gov.co/servicios/descarga-y-metadatos/visor-descarga-geovisores/>

²<https://earthengine.google.com/>

³https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1_TOA

celdas de la malla y el número de habitantes asociado a esa celda.

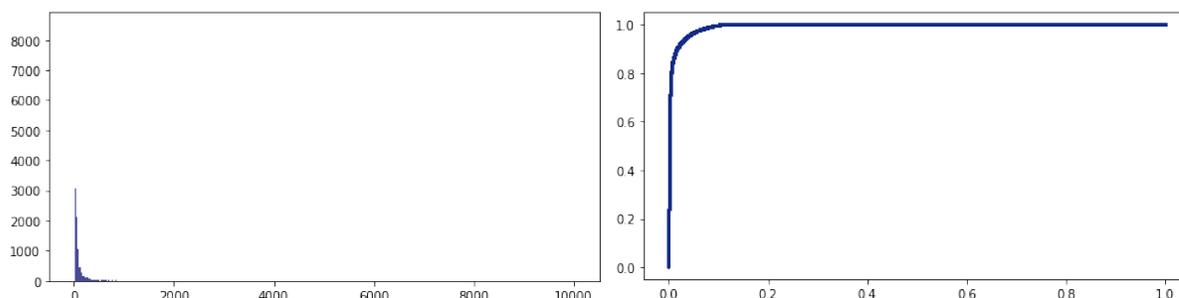


Figura 4-1.: Histograma y diagrama de Pareto de la distribución de la población del CNPV2018 en las celdas de la malla del DANE

La malla de población divide, a manera de partición, el territorio terrestre colombiano. el territorio queda representado en 1'147.103 celdas cuadradas de $1km \times 1km$ sobre la que se distribuyen 43'408.725 habitantes.

La figura 4-1 muestra dos gráficas; la primera es el histograma truncado del número de habitantes por celda; la segunda muestra un diagrama de Pareto, donde se compara el porcentaje acumulado de número de habitantes con el porcentaje acumulado de celdas ordenadas de mayor a menor número de habitantes. Estas dos representaciones muestran que la relación entre número de celdas y número de habitantes sigue una ley potencial, es decir, un alto porcentaje de la población se encuentra concentrada en un porcentaje muy bajo de celdas.

A partir de los datos del diagrama de Pareto se pudo observar que el primer 1% de las celdas contienen el 86,6% de los habitantes censados en 2018.

4.2. Google Earth Engine

GEE es una plataforma de análisis geoespacial para RSD, donde se disponen libremente los datos de diferentes proyectos satelitales como Landsat, Sentinel y MODIS, entre otros. Se disponen de conjuntos de datos listos para ser analizados puesto que ya se les ha aplicado las respectivas correcciones radiométricas, atmosféricas y de georreferenciación [27]. la plataforma dispone de su propio Sistema de Información Geográfica (GIS) que permite construir composiciones de imágenes, eliminar nubosidades, recortar mosaicos a partir de geometrías suministradas por el usuario; el procesamiento es realizado en la nube y suministrado por la misma plataforma desde su aplicación web programada a través del lenguaje *Javascript*.

4.2.1. Proyecto Landsat 8

Landsat 8 es un satélite operado conjuntamente por el Servicio Geológico de los Estados Unidos (USGS) y la Administración Nacional de Aeronáutica y el Espacio (NASA). Landsat 8 observa la Tierra continuamente desde el 2013 hasta la actualidad. Obtiene imágenes de toda la superficie de la Tierra con una resolución de aproximadamente 30 metros y una vez cada 2 semanas, en donde se incluyen datos multi-espectrales a 7 bandas más 2 bandas de datos térmicos [7].

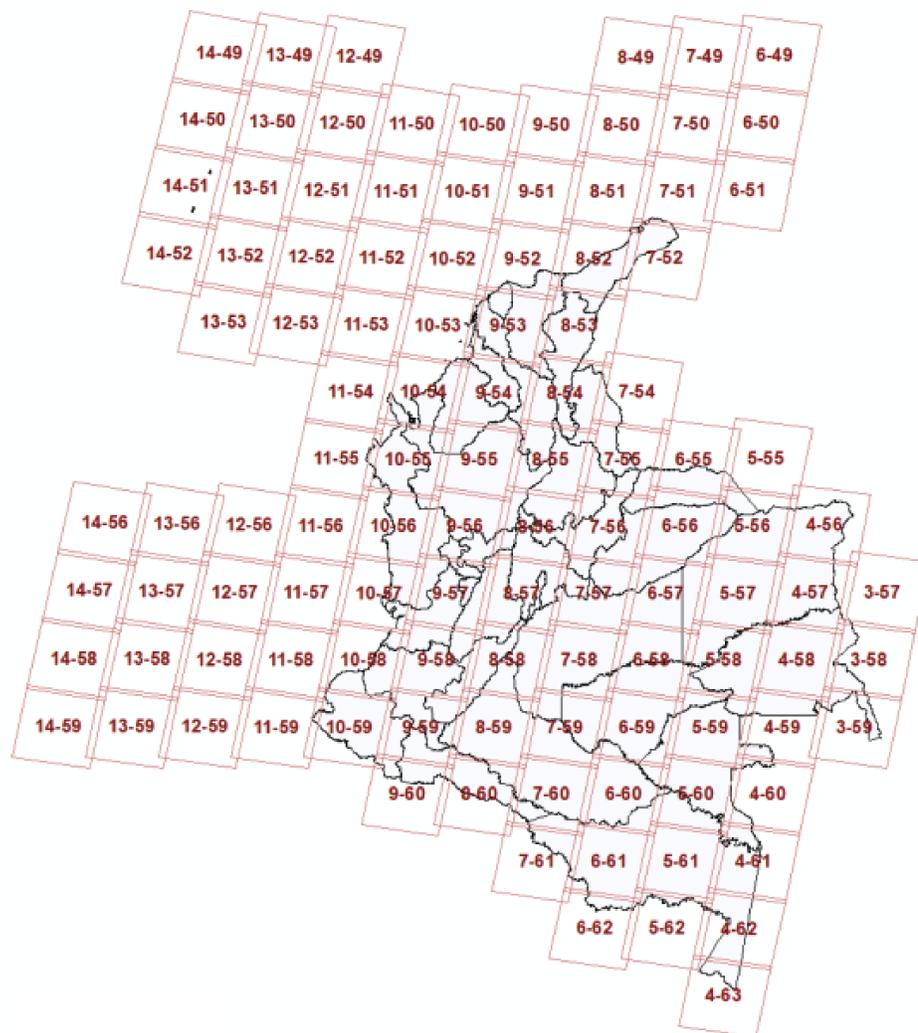


Figura 4-2.: Cuadrícula de Landsat 8 en Colombia. Tomada de <http://arcgis.com/0vWm5K>

En la figura 4-2 se encuentra la identificación de las cuadrículas de Landsat 8 correspondientes al territorio colombiano. El satélite recolecta las imágenes con esa cuadrícula de referencia; estas imágenes, llamadas escenas, toman registro del suelo y de las nubes, por lo

que para obtener imágenes del suelo en un área específica, como lo es la geometría de una de las celdas de la malla de población, es necesario realizar composiciones de escenas teniendo en cuenta el porcentaje de nubosidad calculado por GEE y el periodo de estudio.

A pesar de que los proyectos MODIS y Sentinel ofrecen mayores resoluciones espectrales y espaciales, se optó por utilizar los datos del proyecto Landsat 8, teniendo en cuenta que la plataforma GEE ofrecía métodos ya implementados de composición de imágenes con datos sin nubosidad⁴. Cabe mencionar también que los trabajos de Doupe et. al y Robinson et al. optaron por utilizar datos del proyecto Landsat 7.

El tamaño de una escena de Landsat es de $170km$ de norte a sur por $183km$ de este a oeste, lo cual es considerablemente más grande que el tamaño de una celda de $1km \times 1km$. Dependiendo de la celda de la malla de población de interés, se suministra la geometría de la celda a GEE para que la plataforma entregue la respectiva imagen multi-espectral compuesta. En el siguiente capítulo se explica el proceso de recolección de las imágenes correspondientes a las celdas de interés en la malla.

⁴<https://developers.google.com/earth-engine/apidocs/ee-algorithms-landsat-simplecloudscore>

5. Metodología

Para la obtención de un modelo que permita predecir número de habitantes a partir de imágenes de satélite, se tomó una muestra de la malla de población que da prelación a celdas con mayor número de habitantes, con la que se extraen, en GGE, las respectivas imágenes para el conjunto de datos de entrenamiento, validación y prueba, junto con las respectivas imágenes de prueba y los datos de población del área metropolitana de Bogotá, para luego optar y aplicar a ajustar una CNN. A continuación se describe el proceso de aplicación de esta metodología.

5.1. Obtención de imágenes satelitales

Dado que la malla de población supera el millón de celdas y dada la cantidad de recursos tanto computacionales como de tiempo que implicaría recopilar una imagen por cada una de las celdas de la malla, se ha extraído una muestra de celdas de la malla, de forma que represente tanto las diferentes densidades de población en las unidades de malla, así como compilar las diversas características de las imágenes multi-espectrales de Landsat 8 que dan evidencia de la actividad humana relacionada a la cantidad de habitantes que se encuentran en las respectivas celdas.

5.1.1. Muestreo de celdas sobre la malla de población

Como se mencionó, se aplicó un muestreo en las celdas de la malla, principalmente por dos razones: ahorro en tiempo y recursos computacionales. En el trabajo de referencia de Doupe et al., las áreas de estudio fue compendiado en su totalidad con las imágenes de satélite, mientras que en el trabajo de Robinson et al. se aplicó un muestreo aleatorio simple sobre el área de estudio.

Dada la naturaleza exploratoria de esta investigación, se parte de la hipótesis de que el muestreo aplicado tiene un efecto directo en el desempeño del modelo dependiendo de las imágenes de los lugares con los que se entrene y con los que se evalúe.

En este estudio se utilizó un muestreo proporcional al tamaño [28] (PPS por la siglas en inglés de *Probability Proportional to Size sampling*) procurando incluir en la muestra, en la medida más balanceada posible, áreas con todos los niveles de densidad disponibles en la malla de población original.

Teniendo en cuenta el tamaño de almacenamiento de una imagen y asegurando el poder cargar el conjunto de datos aumentado en su totalidad sobre la memoria RAM disponible en la máquina empleada para esta investigación, se estableció usar un tamaño de muestra de 12.500 celdas.

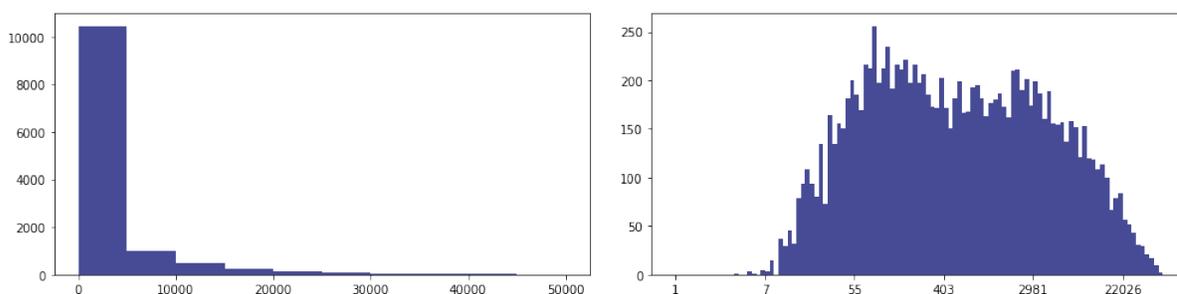


Figura 5-1.: Distribución muestral del número de celdas según su número de habitantes

Adicional al diseño muestral, y para asegurar una representación lo más alta posible de las celdas con más de 1.800 habitantes, se hizo inclusión forzosa sobre estas celdas en la muestra. Se resalta la importancia de incluir forzosamente estas celdas en la muestra, ya que representan el 3,2% del número de celdas de la muestra y a la vez el 75,0% de los habitantes que suman esta misma.

La figura 5-1 muestra la distribución del número de habitantes en las celdas incluidas en la muestra. Si se observa la gráfica de la izquierda, la mayor parte de la muestra incluye celdas con baja densidad de población, mientras que en la figura de la derecha se observa que el número de habitantes en escala logarítmica tiene un mejor balance de celdas en los diferentes niveles de densidad de población. Esto muestra que a pesar de aplicar el diseño muestral PPS y la inclusión forzosa de las celdas de mayor número de habitantes, la ley potencial que relaciona el número de habitantes y número de celdas sigue viéndose reflejada en la muestra.

La figura 5-2 muestra la ubicación de las celdas seleccionadas en la muestra sobre el territorio colombiano. En este mapa se puede apreciar en términos generales que la muestra está presente principalmente en áreas urbanas, puesto que resalta las áreas metropolitanas del país; por ejemplo: Pasto e Ipiales, valle del Cauca, valle del alto Magdalena en el Huila, Villavicencio, sabana de Bogotá, Eje Cafetero, valle de Aburrá, Bucaramanga, región Sabanas, costa Caribe y archipiélago de San Andrés y Providencia.

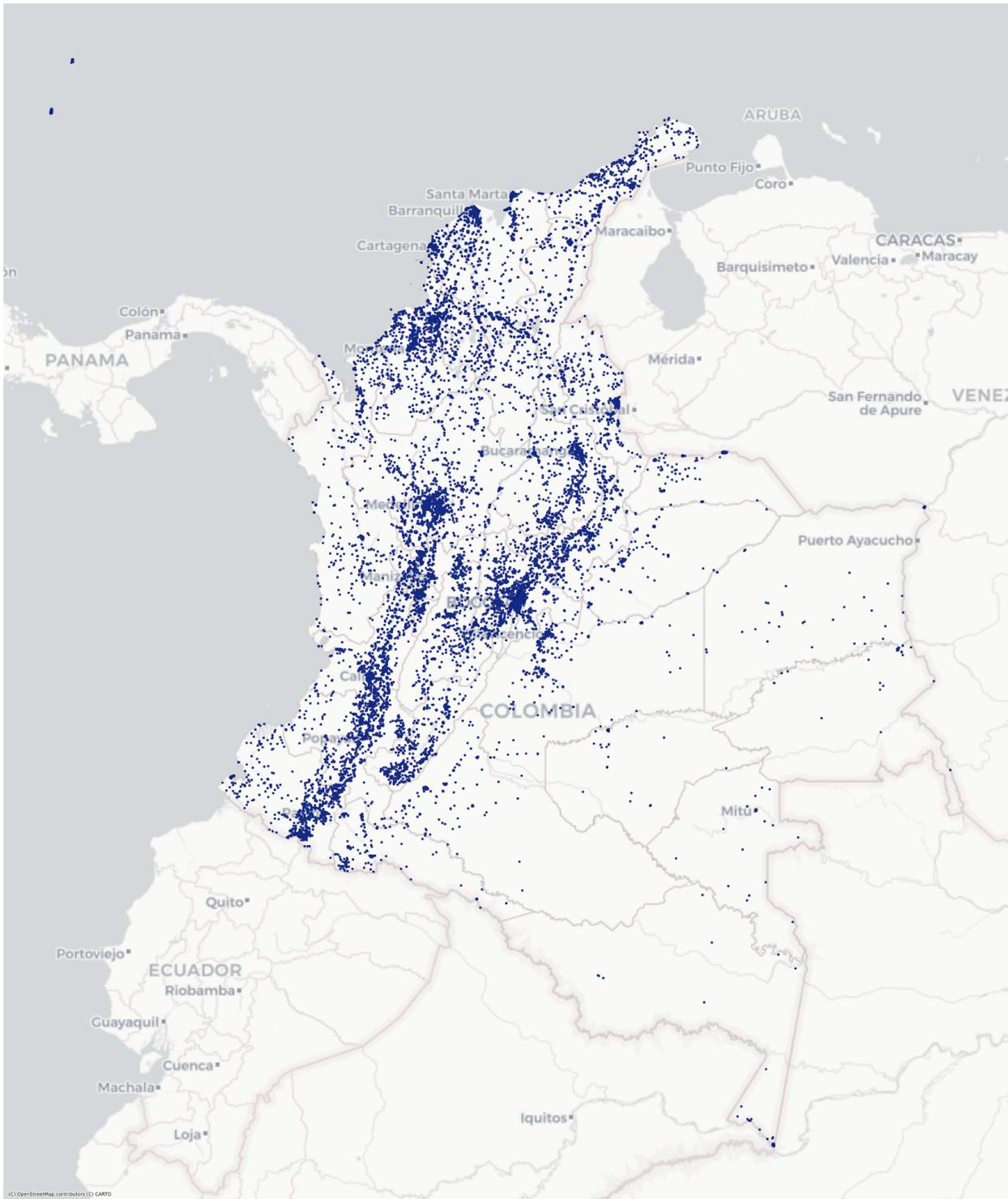


Figura 5-2.: Distribución espacial de la muestra de celdas sobre el territorio terrestre colombiano. Elaboración propia

5.1.2. Descarga de imágenes por celda en Google Earth Engine

Luego de tener preparada la muestra de celdas en un nuevo archivo *shapefile*, cada una con su respectivo conteo de habitantes de acuerdo al censo, se asoció una imagen satelital a cada una de las celdas de la muestra, para lo cual se elaboró un proceso para obtener estas imágenes en forma sistemática desde la aplicación web de GEE. Las subsecciones a continuación describen y justifican este proceso de captura de imágenes.

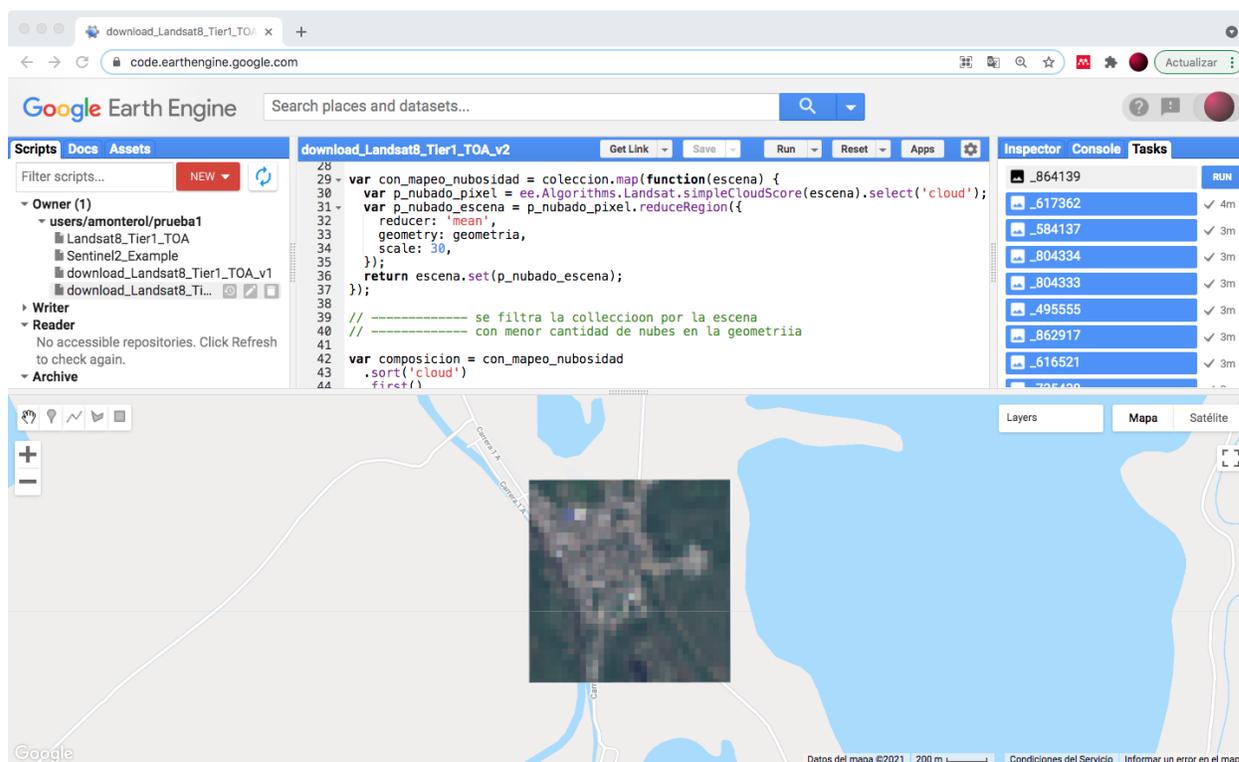


Figura 5-3.: Interfaz gráfica de la aplicación web de GEE

5.1.2.1. Método de composición de imágenes por celda

La obtención de una imagen satelital para cada celda de la muestra está sujeta a las siguientes condiciones:

- El conjunto de escenas tomadas sobre el área de la celda durante el periodo de estudio
- La intersección de cuadrículas sobre las que se toman las escenas, en caso en que el área de las celdas de interés quedan sobre los límites de las cuadrículas
- La nubosidad presente en las escenas sobre el área de las celdas

A través del editor de código de GEE, programable en lenguaje *Javascript*, se gestiona la extracción de imágenes individuales, libres de nubosidad en la mayor medida posible, compuestas de las escenas que intersectan con la geometría de la celda; las dimensiones de las imágenes resultantes tienen 32 píxeles x 32 píxeles, coincidentes con la resolución espacial y dimensiones de las celdas. Estas imágenes comprenden 9 de las once 11 bandas disponibles, correspondientes a las 7 bandas del espectro de luz más las 2 bandas del espectro térmico. En la parte inferior de la figura 5-3, sobre el panel del mapa, se muestra una imagen obtenida del procedimiento, construida con las tres bandas del espectro de luz visible. GEE permite llevar a almacenamiento imágenes a los directorios de la cuenta personal de Google Drive.

El método desarrollado en *Javascript*, junto la capa de celdas de la muestra de la malla de población, se encuentran dentro de los anexos a este documento, en donde se especifican los argumentos y el funcionamiento de las sentencias de este método.

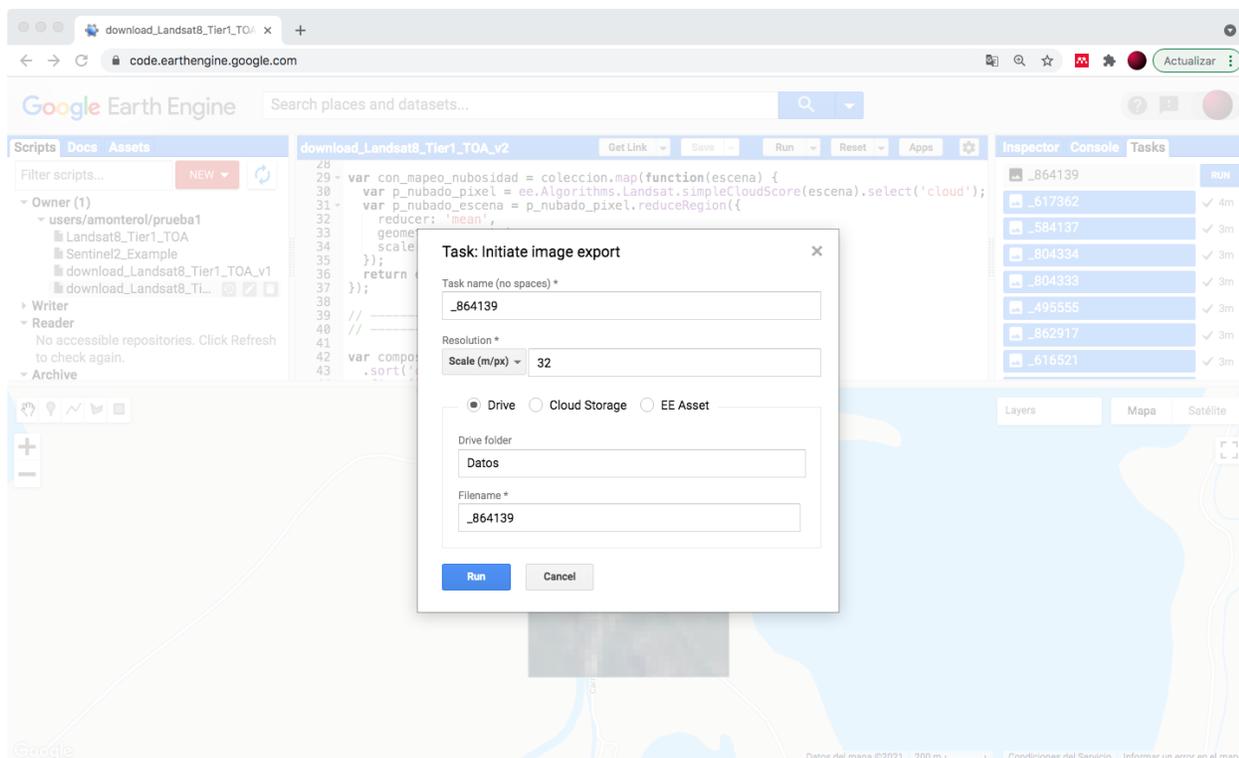


Figura 5-4.: Formulario emergente para confirmar la descarga de las imágenes individuales desde GEE

5.1.2.2. Automatización de descarga

GEE tiene una limitación: no permite realizar descargas de forma completamente automatizada, pues solicita la confirmación manual de las dimensiones de la imagen y el nombre

del archivo de cada una de las imágenes compuestas a descargar. Si se observa de nuevo la figura 5-3, se observa en el panel superior derecho en la pestaña *Tasks* que, para descargar la imagen, es necesario realizar clic sobre el botón *RUN*. En la figura 5-4 se muestra el formulario emergente que se despliega luego de ordenar el clic sobre el botón *RUN*, en el cual se diligencia, nuevamente y como se mencionó, los respectivos campos.

Para superar esta limitación de tener que descargar manualmente las 12.500 imágenes de la muestra mas las 9.537 imágenes del área metropolitana de Bogotá, y luego de ensayar herramientas para *web scraping* que pudiesen automatizar el manejo del explorador de internet, se generó un método en Python con el uso de su librería PyAutoGUI¹, que permitiese automatizar el control del teclado y el ratón de un computador personal, la cual hace las veces de robot para ejecutar las confirmaciones de descarga de cada imagen. Dentro de los anexos a este documento se encuentra el método desarrollado.

El tiempo de ejecución de la descarga de las 12.500 imágenes de la muestra mas las 9.537 imágenes del área metropolitana de Bogotá tomó alrededor de 3 y 4 días sin interrupción desde un computador regular de escritorio con conexión a internet.

5.1.2.3. Verificación de descarga

La obtención de las imágenes fue automatizada y además fue necesario verificar los siguientes aspectos en el conjunto de imágenes descargado, asegurando la consistencia del etiquetado de las imágenes con la cantidad de habitantes correspondiente:

- La descarga del conjunto completo de imágenes
- La coincidencia en ubicación de la celda que reportaba Google Maps con OpenStreet-Map
- La coincidencia entre la imagen reportada por la interfaz gráfica de GEE y la imagen descargada

Para esta tarea, se seleccionaron aleatoriamente 30 imágenes de la muestra y 30 imágenes del conjunto de imágenes del área metropolitana de Bogotá, para comprobar a través del identificador de la celda en el *shapefile* la coincidencia en los criterios mencionados. No se obtuvo ninguna inconsistencia en el conjunto de imágenes de la muestra, ni en el conjunto de imágenes del área metropolitana de Bogotá.

¹<https://pyautogui.readthedocs.io/en/latest/>



Figura 5-5.: Verificación de la coincidencia entre la visualización: con una de las bandas del espectro visible en Python, con la imagen desplegada en el panel de Google Maps y con la ubicación de la celda sobre una capa de información de OpenStreetMap

5.2. Preprocesamiento

Luego de recopilar las imágenes correspondientes a la muestra de de celdas de la mallas, se aplica la respectiva normalización a los valores de los píxeles en las imágenes para que la implementación de las CNN reciban entradas con la escala de valores adecuada. Se aplica un aumento de datos en las imágenes normalizadas y se construye el conjunto de datos que sirven para modelar y validar modelos de aprendizaje automático evaluados en la investigación. Las siguientes secciones describen el procedimiento aplicado en detalle.

5.2.1. Normalización

Luego de descargar los conjuntos de imágenes, se observó que las bandas de las imágenes contenían valores de píxeles con decimales y en diferentes rangos, con diferencias apreciables principalmente entre las bandas del espectro de luz y las del espectro térmico. Se resalta que los valores de píxeles encontrados no coincidían con los valores transformados desde la resolución radiométrica de 16 bits reportada en las especificaciones de la plataforma Landsat 8 debido a que estos valores fueron previamente corregidos [27].

Con el propósito de suministrar valores de píxeles adecuados a las capas de entrada de las CNNs que se implementaron en este trabajo, y debido a la marcada diferencia entre los valores de las bandas del espectro de luz y térmico, se aplicaron 9 normalizaciones a través de cada una de las 9 bandas de las imágenes.

Para precisar, la normalización consistió en tomar cada valor de píxel, restarle la media y dividir por la desviación estándar de los valores de los píxeles de la respectiva banda en el conjunto de datos de la muestra.

5.2.2. Aumento de datos

A partir de la muestra de 12.500 imágenes de muestras del territorio colombiano, se aumentó este conjunto de imágenes en ocho veces, aplicando espejo y giros de 90, 180 y 270 grados, tal y como se ilustra en la figura 5-6, de lo cual se obtuvo un conjunto de datos con 100.000 imágenes a partir de la muestra original.

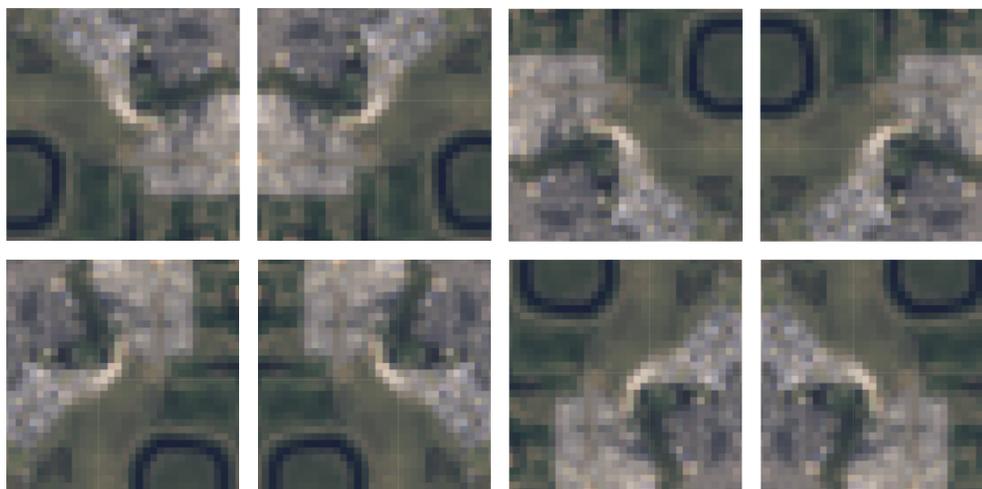


Figura 5-6.: Obtención de ocho imágenes a partir de una, aplicando espejos y giros de 90, 180 y 270 grados

El aumento de este conjunto de imágenes permite tener un mayor conjunto de datos de entrenamiento para las CNN, lo cual evita principalmente el sobreajuste de los modelos obtenidos [29].

5.2.3. Conjunto de datos obtenido

Finalmente, y luego de tener un conjunto de 100.000 imágenes de la muestra de celdas de la malla de población de 2018, así como tres conjuntos de 3.179 imágenes que cubren los municipios del área metropolitana de Bogotá para los años 2016, 2018 y 2020, se partió la muestra de imágenes de Colombia aleatoriamente en tres subconjuntos: entrenamiento, validación y prueba, cada uno con el 80, 10 y 10% del total de imágenes, respectivamente.

Los conjuntos de imágenes del área metropolitana de Bogotá fueron utilizados únicamente como prueba. La preparación del conjunto de datos se encuentra ilustrada en la figura 5-7.



Figura 5-7.: Separación del conjunto de datos de imágenes para realizar entrenamiento, validación y prueba

El tamaño total en disco de los conjuntos de datos mencionados en la figura 5-7 en formato de lectura .npy para la librería *numpy*² de Python fue de 3,9 Giga bytes. Anexo a este documento, y relacionado en la lista de anexos de este documentos, se encuentran los conjuntos de imágenes obtenidos.

La obtención de este resultado da cumplimiento al primer objetivo específico que fue trazado para esta investigación.

5.3. Modelado

El modelo a implementar tiene como propósito tomar imágenes satelitales de entrada, para que se obtenga, a modo de regresión, una predicción de número de habitantes como salida. El modelo deduciría, a partir de las características que este haya aprendido a detectar, el número de habitantes que residen en el área de la que se obtuvo la imagen.

Si se aprecian, con el ojo humano, las imágenes compuestas por las tres bandas correspondientes al espectro visible, se pueden distinguir, como se aprecia en las figuras 5-3, 5-6 y 6-3, por ejemplo: edificaciones, zonas verdes, vías vehiculares, suelo desnudo, ríos, mar, lagunas, cultivos, montañas, bosques, incluso nubes que no fueron eliminadas en la composición, entre otros. Estos elementos son además presenciados en mayor o menor número, tamaño y

²<https://numpy.org/>

densidad. Se espera, por lo tanto, que el modelo distinga de esas características para inferir el número de habitantes del área comprendida por la imagen.

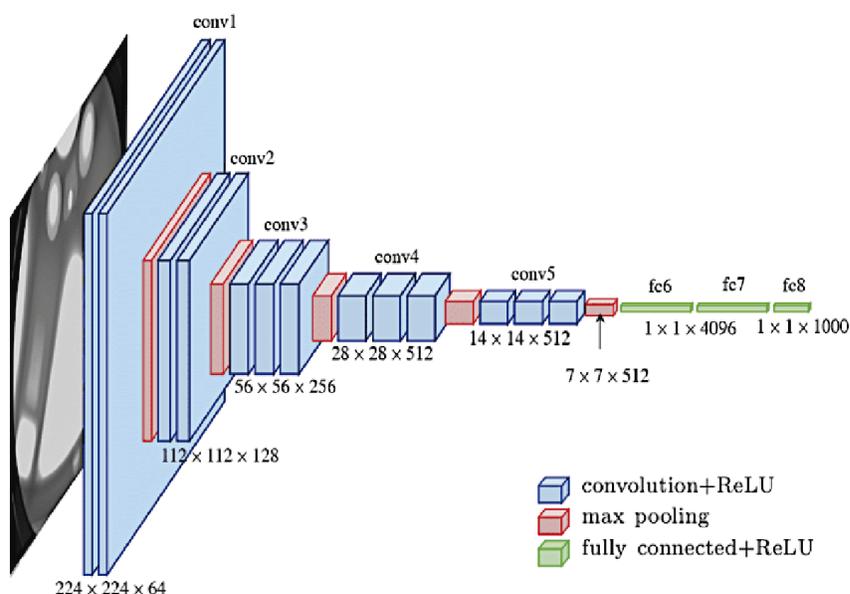


Figura 5-8.: Arquitectura VGG-A [5]

En los trabajos de referencia de Doupe et al. y Robinson et al. se implementaron modificaciones al modelo de CNN con arquitectura VGG-A (Figura 5-8), originalmente desarrollada por Simonyan y Zisserman para el reconocimiento y clasificación en imágenes [25], recordando que se justifican el uso de filtros pequeños (3×3), muestreos o *poolings* de tamaño 2×2 y el aumento en número de capas, de forma de que se mejore el desempeño predictivo de las clasificaciones.

Si se comparan las figuras 2-3 y 5-8, se observa que las únicas modificaciones incluidas por Robinson et al. fueron las capas de entrada y salida, de forma que la arquitectura sirviese para realizar regresión; mientras que Doupe et al. simplificaron la arquitectura original eliminando los últimos dos bloques convolucionales. Estas alteraciones a la arquitectura VGG-A motivaron a la experimentación de posibles arquitecturas más simplificadas que tuviesen, al menos, el mismo desempeño predictivo obtenido en esas dos investigaciones de referencia. A continuación se describe el proceso de experimentación que se siguió en esta investigación para el ajuste de los hiperparámetros de la arquitectura final obtenida.

5.3.1. Experimentación

El principal indicador con el que se evaluaron el desempeño predictivo del modelado en los trabajos Doupe et al. y Robinson et al. fue mediante la métrica de exactitud Error Medio Absoluto (MAE, por su sigla en inglés), siendo además la función de costo que se evaluó en el entrenamiento de los modelos.

Para valorar el desempeño de las modificaciones a la afinación de hiperparámetros en las CNN implementadas, se tomaron los siguientes valores como referencia de partida, para que fueran comparados con el MAE obtenido sobre el conjunto de validación:

- El número de habitantes promedio en una celda de la muestra de la malla de población: 2.860,5 hab.
- El mejor MAE reportado por Doupe et al.: 1.037,8 hab.
- El mejor MedianAE reportado por Robinson et al.: 6.357,0 hab.

Para la primera implementación realizada en esta investigación, se construyó una arquitectura VGG-A en semejanza a la utilizada por Robinson et al., en donde se modificaron únicamente las capas de salida y de entrada, de forma que se obtuviese el modelo de regresión con imágenes. Luego de implementar y entrenar esta arquitectura modificada durante 50 épocas, se obtuvo un MAE de 1490. En la segunda implementación se eliminaron los dos bloques convolucionales tal y como lo realizaron Doupe et al., de donde se obtuvo un MAE de 1.351, luego de entrenar también durante 50 épocas.

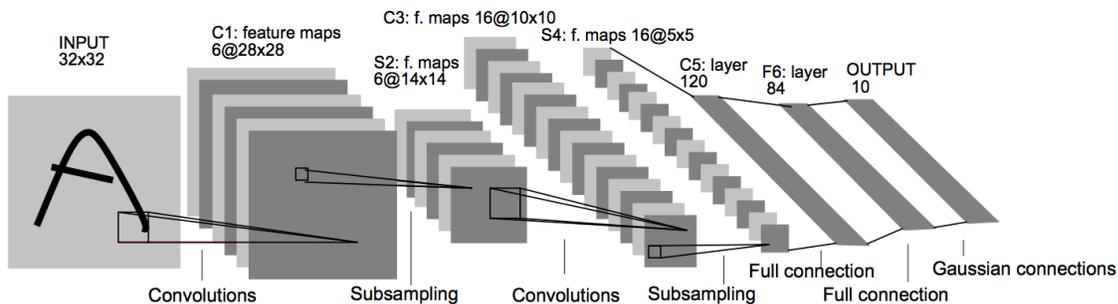


Figura 5-9.: Arquitectura LeNet-5 [6]

Dado que la arquitectura VGG-A no fue diseñada con el propósito de hacer regresión, y luego de haber observado esta mejora apreciable al reducir en dos bloques correspondientes a cuatro capas convolucionales, se decidió experimentar entonces con una arquitectura de CNN con una menor cantidad mucho menor de bloques, esperando obtener el mismo efecto

de mejoría mencionado anteriormente, prescindiendo también de la complejidad de la arquitectura VGG que se consideró inicialmente.

Luego de una revisión bibliográfica centrada en este particular, se encontró el trabajo investigativo de Yann LeCun [6], en el que se propuso una arquitectura de solo 2 bloques convolucionales de solo una capa de convolución en cada bloque, llamada LeNet-5. Esta arquitectura se puede apreciar en la figura 5-9. La red neuronal LeNet-5 fue originalmente diseñada para la clasificación de caracteres escritos a mano. En la tabla 5-1 se muestra una comparación de las arquitecturas mencionadas ³.

Al implementar la arquitectura LeNet-5, y luego de modificar sus capas de entrada y salida para que realizara la respectiva regresión, el MAE que se estaba obteniendo se redujo hasta por debajo de 1.000. Debido a que se logró superar en desempeño predictivo a los trabajos de referencia, se optó por terminar de calibrar los hiperparámetros de la CNN a partir de esta arquitectura, lo cual únicamente consistió en cambiar la función de activación por la ReLU, tomar un tamaño de lote de 128 que permitiese acelerar el proceso de entrenamiento, reducir la tasa de aprendizaje y extender el número de épocas.

Tabla 5-1.: Comparación de las arquitecturas CNN utilizadas

	VGG-A Simonyan y Zisserman (2015)	LeNet-5 LeCun, Bottou, Bengio y Haffner (1998)
Propósito	Reconocimiento en imágenes	Reconocimiento de caracteres escritos a mano
Capas ocultas entrenables	11	5
Bloques	5	2
Capas de convolución	8	2
Función de activación	ReLU	Sigmoid, Tanh
Parámetros entrenables (impl.)	32 Mill.	179 mil

5.3.2. Implementación final

En la gráfica 5-10 se muestra la evolución de la función de costo en los conjuntos de entrenamiento y validación de acuerdo al número de épocas ejecutadas en el proceso de entrenamiento de la CNN. El tiempo transcurrido durante las 100 épocas fue de aproximadamente

³El numero de parámetros entrenables reportado fue el obtenido durante la implementación y no el reportado por los trabajos originales [25, 6]

7 minutos. A continuación se muestra la configuración de la red LeNet-5 modificada para realizar regresión con el conjunto de imágenes de satélite:

Parámetros entrenables: 179.439

Tamaño de lote: 128

Épocas: 100

Taza de aprendizaje: 0,001

Función de activación: ReLU

Optimizador: adam

Función de costo: *Mean Absolute Error* (MSE)

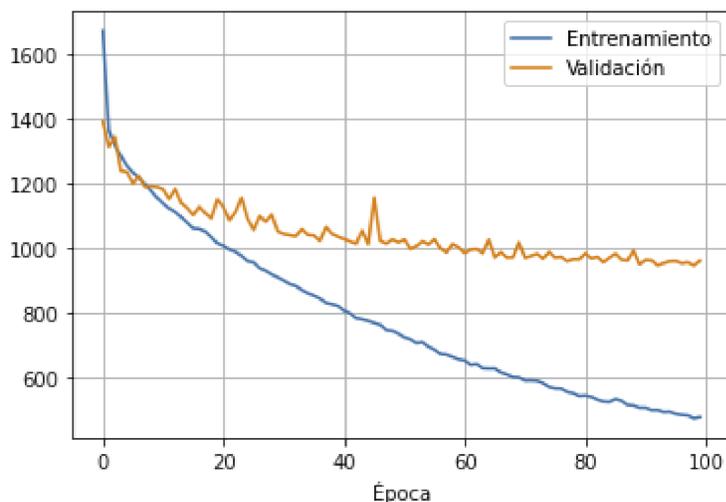


Figura 5-10.: Evolución de la función de costo MAE durante el entrenamiento de la red LeNet-5 luego de la afinación de los hiperparámetros

Aunque la gráfica 5-10 muestra que posiblemente se hubiesen podido ejecutar más épocas para reducir aún más la función de costo, se decidió por detener el entrenamiento, puesto que este conjunto de validación ya mostraba mayor desempeño predictivo respecto al MAE en comparación con los dos trabajos de referencia, previniendo, además, posibles sobreajustes, sobre todo, en conjuntos de datos más alejados del contexto de la muestra de entrenamiento, tal y como lo son los conjuntos de datos para el área metropolitana de Bogotá.

5.3.3. Validación del modelo final

Como se apreció en esta descripción de la experimentación, no se utilizó ningún criterio de búsqueda que fuese automatizado ni tampoco validado estadísticamente para conducir la afinación de hiperparámetros de la CNN obtenida. La afinación de hiperparámetros se consistió en mejorar, en la mayor medida posible, la exactitud sobre la parte del conjunto de datos de

validación, por medio de la configuración y experimentación manual con los hiperparámetros de las CNN.

Cabe también mencionar, por un lado, que en el trabajo de Doupe et al. se valida a partir de las imágenes de un territorio diferente al del que se tomaron las imágenes de entrenamiento, es decir, se entrenó con imágenes de Kenia para validar con las imágenes de Tanzania. De manera similar, para este trabajo, uno de los dos conjuntos de validación, el de el área metropolitana de Bogotá, es territorialmente diferente a la muestra de imágenes satelitales de entrenamiento de Colombia.

Por otra parte, en el trabajo de Robinson et al. se toma una submuestra del conjunto de datos de imágenes tomado del territorio continental de Estados Unidos, lo cual hace analogía al subconjunto de validación de la muestra de imágenes de Colombia para este estudio. Se resalta entonces, que el método de validación expuesto en este trabajo sigue los mismos esquemas que los dos trabajos de referencia, de forma tal que los modelados sean comparables al utilizar las mismas métricas expuestas en esos dos trabajos.

5.3.3.1. Herramientas tecnológicas utilizadas

A continuación se especifican las características de las herramientas computacionales que fueron utilizadas para el entrenamiento del modelo de predicción:

Software: Python 3.7, Keras 2.5.0

Sistema operativo: Ubuntu 18.04 (Google Cloud Computing)

CPU: Intel Xeon CPU 2.20GHz

RAM: 13Gb

GPU: NVIDIA Tesla T4

Para el desarrollo del software, se utilizó la aplicación web Google Colab, en donde se dispone de una interfaz de programación (API, por su sigla en inglés) mediante *notebooks* bajo el lenguaje Python 3, y en la que se puede además ejecutar estos métodos sin necesidad de hacer mayores configuraciones o alistamiento del entorno para realizar tareas de Ciencia de Datos [30].

Para manejo de la capas de información espacial se utilizó la librería GeoPandas [31], la cual utiliza las mismas estructuras de datos que la librería Pandas [32], permitiendo así manipular y graficar los datos de las geometrías a partir del uso de tablas del mismo modo en que se realiza con la API en Python de Pandas, e integrando sistemas de georreferenciación para la localización de las geometrías en el área de estudio. Para la verificación visual de los cruces espaciales a partir de los datos fuente se utilizó el software QGIS [33].

La implementación de la CNN de este trabajo se realizó mediante la interfaz del software TensorFlow [34] llamada Keras [35] y que funciona sobre la misma interfaz de Python 3. Keras permite implementar las arquitecturas de CNN a partir de sentencias que hacen las veces de las capas de la arquitectura que se está implementando, ya sean convolucionales, de *pooling*, capas densas, funciones de activación, entre otras; permitiendo la configuración sencilla de los hiperparámetros y de la arquitectura en sí.

La elaboración de los mapas y de los gráficos estadísticos de este trabajo fueron desarrollados mediante la API en Python de la librería matplotlib [36].

En cuanto al funcionamiento de los métodos, se anexan a este documento, los *notebooks* que explican el paso a paso que se siguió, los métodos son:

- Obtención del conjunto de imágenes
- El método de pre-procesamiento
- El método de modelado
- El método para validar el modelado
- obtención del archivo del modelo luego de su entrenamiento, en extensión .h5

La obtención de esta implementación da cumplimiento al segundo objetivo específico que fue trazado para esta investigación.

6. Resultados

Los resultados han sido apreciados desde el alcance en el desempeño predictivo que logró la implementación de la arquitectura modificada LeNet-5, luego de utilizar el diseño muestral PPS sobre el conjunto original sobre datos de población. El desempeño predictivo del modelo obtenido fue evaluado sobre los conjuntos de datos de prueba y fueron comparado a través de las estadísticas calculadas por los trabajos de referencia de Doupe et al. y Robinson et al. Posteriormente, se realizó una exploración descriptiva de los residuales de las unidades de predicción sobre el conjunto de pruebas del área metropolitana de Bogotá de 2018 que diera mayor explicación sobre el desempeño predictivo obtenido. Para finalizar, se hizo una exploración sobre la predicción agregada de los municipios que componen el área metropolitana de Bogotá para los periodos 2016, 2018 y 2020 comparada a las proyecciones de población del DANE, para mostrar la relevancia alternativa que propone este modelado frente a las cifras obtenidas mediante modelado demográfico.

6.1. Comparativa en el desempeño de predicción

Este trabajo de investigación ha sido motivado, en su medida, por tener un modelado que permitiese replicar los trabajos de referencia de Doupe et al. y Robinson et al. en el contexto local y nacional, por lo cual es pertinente la comparativa de los resultados del desempeño obtenido en este modelado frente a estos dos antecedentes. Aunque los dos trabajos solamente coinciden en la medición del MAE en sus conjuntos de datos de validación, en el presente trabajo se evaluaron todas las métricas reportadas en cada uno de los dos trabajos.

En la tabla **6-1** se observa que Doupe et al. reportaron el MAE, el error cuadrático medio o *Root Mean Square Error* (RMSE) y el error cuadrático medio relativo o *Relative Root Mean Square Error* (RMSE); se observa también que Robinson et al. reportaron el MAE, el error mediano absoluto o *Median Absolute Error* (MedianAE) y el error porcentual absoluto medio o *Mean Absolute Percentage Error* (MAPE).

La tabla **6-1** muestra como las estadísticas de variación absoluta (MAE, MedianAE y RMSE) tuvieron un desempeño similar al modelo desarrollado por Doupe et al., mientras que las estadísticas de variación relativa (%RMSE y MAPE) tuvieron un bajo desempeño respecto a las dos investigaciones de referencia. El desempeño respecto al r^2 , que muestra el porcentaje de variación explicada por el modelo, tiene un desempeño similar al del trabajo de Robinson

et al.

Tabla 6-1.: Comparativa de las métricas de desempeño predictivo del modelo obtenido contra los reportes de los trabajos de referencia

		Cobertura		MAE	MedianAE	RMSE	%RMSE	MAPE	r ²
*	PPS + LeNet-5	Colombia	2018	947,8	252,3	2.184,9	697,8	13.773,9	0,841
		A. M. Bogotá	2018	1.172,9	267,4	2.653,2	5.066,7	16.421,2	0,970
	Doupe et al. [10]	Tanzania	2002	1037,8	-	2.102,4	51,5	-	-
	Robinson et al. [4]	US. Continental	2010	23.005,0	6.357,0	-	-	73,8	0,910

Como se mencionó en la descripción de la fuente de datos de la malla de población del CNPV2018, se hizo importante tener en cuenta el balanceo del conjunto de datos de entrenamiento, de forma que se incluyera, en menor proporción, imágenes a las que les correspondiera un número bajo de habitantes, pues podrían afectar negativamente el desempeño del modelo en la tarea de predecir en áreas con alta densidad de población; si se consulta la distribución del número de habitantes en los conjuntos de datos de los dos trabajos de referencia, esta consideración no fue tenida en cuenta, lo cual sugiere, en una primera exploración de estos resultados, que las diferencias antes mencionadas acerca de las métricas se deben hipotéticamente a que: imágenes con población muy densa en modelos entrenados con mayoría de datos de entrenamiento de baja densidad afectan negativamente el desempeño frente a las métricas de variación absoluta.

Para corroborar la hipótesis antes mencionada, se procedió a subdividir el subconjunto de datos de validación de la muestra de Colombia en 2018 por rangos de población y a medir las mismas métricas de desempeño antes evaluadas, de forma que se pueda observar el ajuste del modelo dependiendo de la densidad de población correspondiente a las imágenes.

Tabla 6-2.: Comparativa de las métricas de desempeño predictivo del modelo obtenido por rangos en el subconjunto de validación para Colombia en 2018

Núm. hab.	Núm. imágenes	MAE	MedianAE	RMSE	%RMSE	MAPE	r ²
(0,47801]	10.000	947,8	252,3	2.184,9	697,8	13.773,9	0,841
(0,54]	1.446	126,4	35,2	493,3	1.762,2	7.283,1	-1.512,6
(54,403]	3.272	200,2	89,0	507,1	330,1	4.151,3	-26,1
(403,2977]	2.915	669,6	464,9	1.015,8	77,1	1.583,1	-0,911
(2977,47801]	2.367	2.825,6	1.834,8	4.288,8	44,8	756,3	0,689

En la tabla **6-2** se puede apreciar efectivamente que, mientras haya una mayor densidad de población en las imágenes evaluadas, mejor es el desempeño en las métricas de variación relativa, peor es el desempeño en las métricas de variación absoluta, y viceversa.

A continuación se muestra el comportamiento de los residuales que justifica los valores altos en las estadísticas de variación relativa.

6.2. Análisis exploratorio de los residuales

Como se mostró en la sección sobre la descripción de la malla de población, el mayor porcentaje de celdas tiene valores de número de habitantes entre 0 y 54, si se observa la gráfica **6-1** en la parte superior izquierda se encuentran residuales en el orden de los 5.000 y hasta 10.000 habitantes, estos son los valores que afectan negativamente el desempeño del modelo frente a las estadísticas de variación relativa y, por tanto, son las celdas que están afectando negativamente el modelo obtenido.

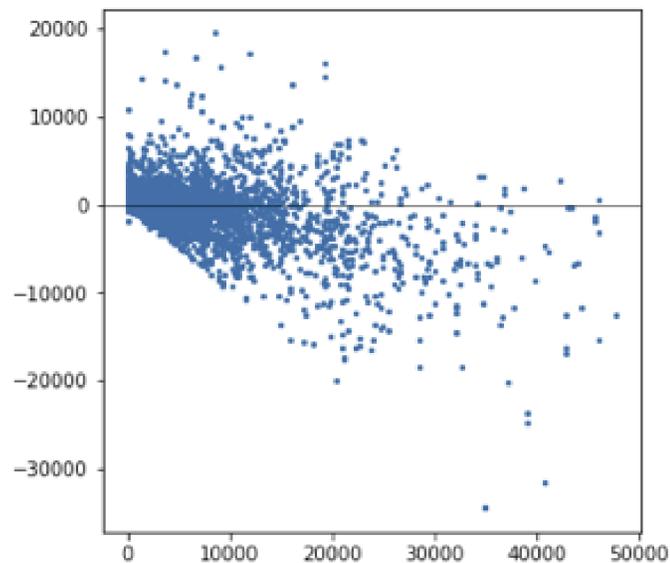


Figura 6-1.: Diagrama de dispersión entre número de habitantes y el residual obtenido, evaluado en el conjunto de prueba del área metropolitana de Bogotá del 2018

6.2.1. Mapeo de los residuales para el área metropolitana de Bogotá en 2018

Debido al carácter espacial de los datos de población, es imprescindible explorar el desempeño del modelo sobre mapas de referencia para las áreas de estudio, de forma que se puedan detectar posibles relaciones entre el buen o mal desempeño del modelo en ciertas ubicaciones del territorio y que den más indicaciones para la mejora del modelado.

Se tomaron entonces los residuales obtenidos para el área metropolitana de Bogotá y se ilustraron sobre un mapa de calor que solapara un mapa de fondo con capas de información generadas por OpenStreetMap. Se observa así en el mapa de la figura 6-2 que las áreas rurales muestran menor variación que las áreas urbanas. Se observa también cómo en zonas del central y norte de Bogotá existen celdas con notable sobre estimación, así como zonas del occidente y suroccidente de Bogotá con una notable sub-estimación. Sin embargo, se aprecia que ni la sub-estimación ni la sobre-estimación estén agrupadas en lugares específicos de esta área de estudio

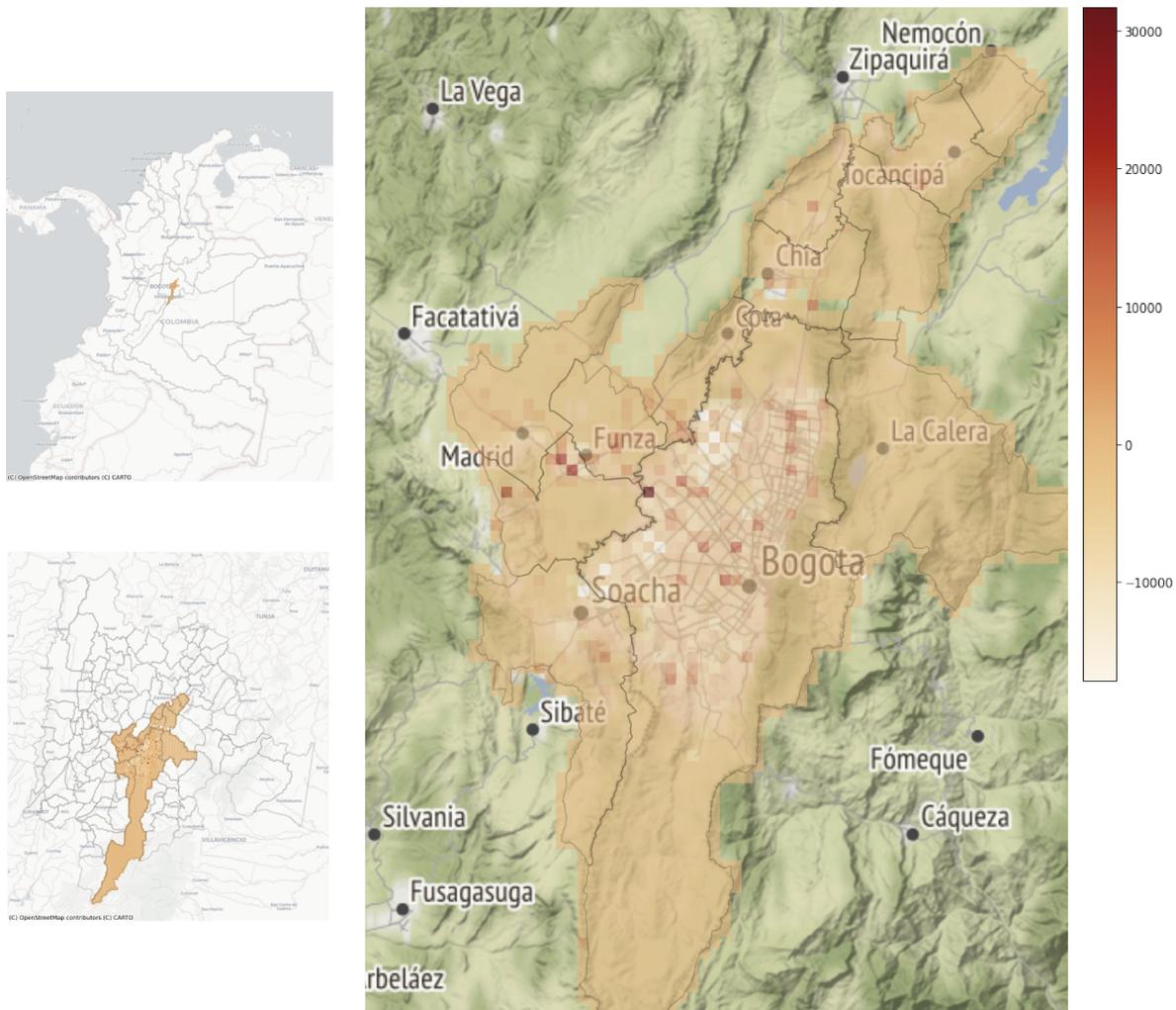


Figura 6-2.: Mapeo de los residuales sobre el área metropolitana de Bogotá de 2018

También se observa en la figura 6-2 la notable diferencia en variación de los residuales entre las zonas rurales, que es baja, y las zonas urbanas, la cual es alta. Se aprecia que las zonas rurales tienen una mayor extensión territorial y, por tanto, representan una mayor cantidad

de los datos evaluados. Se observa que el área del municipio de Bogotá es notablemente más extensa que los demás municipios, por lo que la comparación entre estos se vea afectada en este aspecto.

En la subsección titulada Predicción de población en el área metropolitana de Bogotá se exploraron las diferencias en el desempeño del modelo, separando la evaluación entre municipios. En el mapa de la figura **6-2** se aprecia que algunos municipios como Tocancipá y La Calera muestran ser relativamente más rurales que, por ejemplo, Funza, Chía y Soacha. Lo cual es tenido en cuenta para la interpretación de sus métricas de desempeño.

6.2.2. Exploración de la predicción en casos particulares de imágenes

Se realizó un chequeo visual en algunas las imágenes satelitales del área metropolitana de Bogotá en 2018, en donde se pudiese apreciar unos residuales cercanos a cero, casos con sobre estimación y casos con sub-estimación, de manera que se aprecie la manera en que la CNN está reconociendo características en las imágenes que permitan dar testimonio de la cantidad de población que pueda estar residiendo en esa área.

Para el chequeo visual, se construyó la figura **6-3** donde se construyó una matriz con doce (12) imágenes. Las filas de la matriz están organizadas de menor a mayor número de habitantes. La primera columna muestra predicciones acertadas, la segunda columna presenta casos con sobre-estimación y la tercera columna presenta casos con sub-estimación.

Un vistazo general de la figura **6-3** muestra que el modelo tiende a entregar estimaciones con valores bajos para imágenes con porciones de color verde y que, además, tiende a entregar estimaciones con valores altos en imágenes con porciones de tonos brillantes.

En la matriz de imágenes también se pueden apreciar algunos de los posibles malos entendidos que realiza el modelo. Por ejemplo, la imagen de la primera fila y la segunda columna indica que posiblemente el modelo interpretó, como áreas urbanas, la nubosidad presente en una imagen mal lograda en la composición. También, se observa en la última columna, que las imágenes que combinan áreas sin construcción y áreas densamente edificadas son susceptibles a ser sub-estimadas. En las dos figuras del centro se observan construcciones de mayor extensión, algunas por ejemplo en color azul, por lo que el modelo lo podría estar confundiendo con áreas de población más densas. En la imagen central de la última fila, aunque se observa una imagen en la que la extensión se encuentra construida en su mayoría, existen también lugares con suelo desnudo y construcciones de baja densidad de población que se pueden estar confundiendo con características de imágenes con mayor densidad.

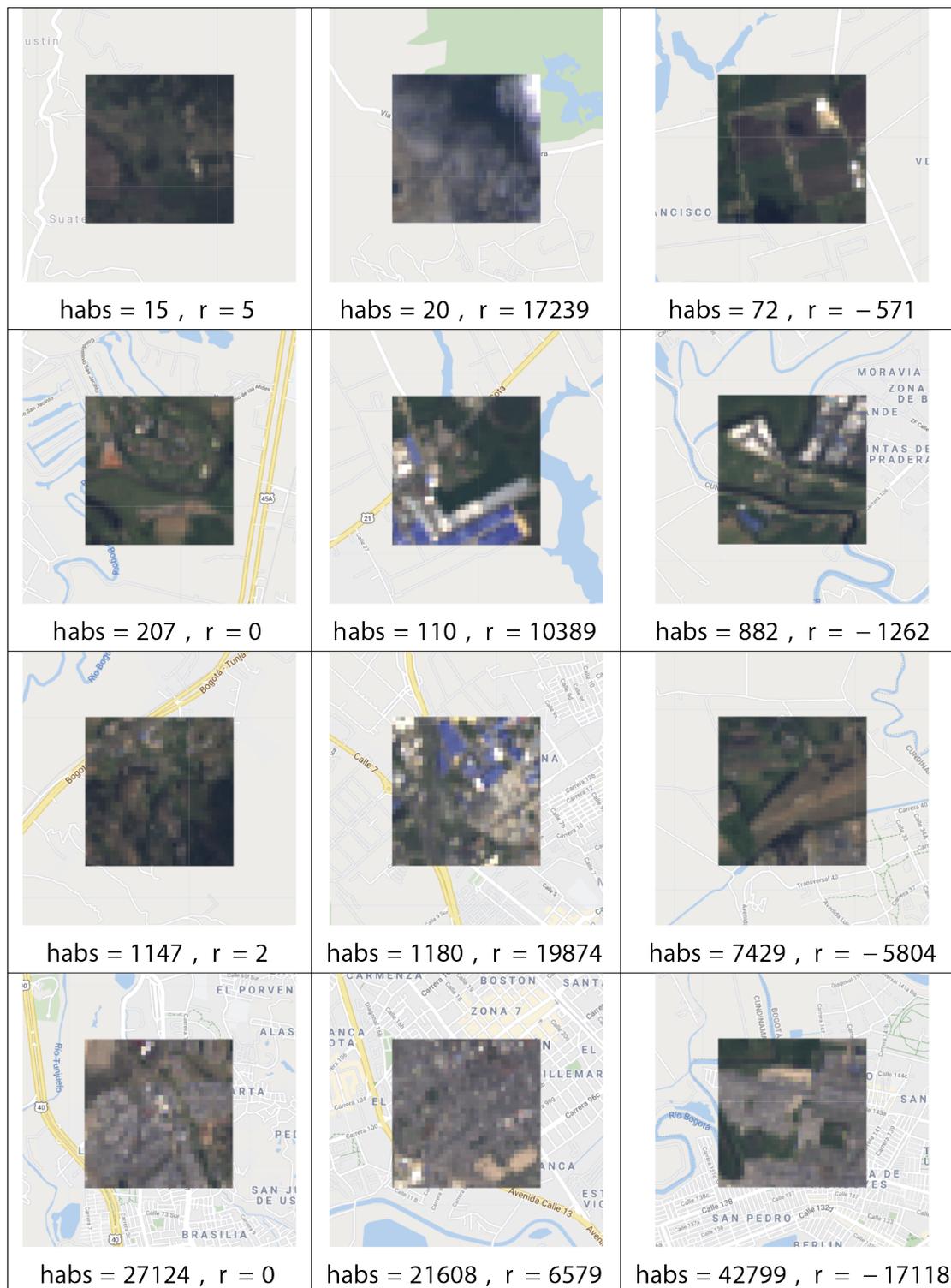


Figura 6-3.: Ejemplos de predicción en casos con residuales pequeños, casos con sobre-estimación y casos con sub-estimación

6.3. Predicción de población en el área metropolitana de Bogotá

Dado el conjunto de datos de prueba del 2018 para los municipios del área metropolitana de Bogotá, se realizó una comparación entre la densidad de número de habitantes por kilómetro cuadrado (es decir, el área que comprende una imagen de los conjuntos de datos) y el rendimiento, de forma que se aprecie si existe una posible correlación entre el desempeño del modelo respecto al MAE y la densidad de población de las imágenes con las que se evalúa el modelo. En la tabla 6-3 se observa la evaluación del desempeño predictivo agregado por municipios en el área metropolitana de Bogotá de 2018.

Tabla 6-3.: MAE para el área metropolitana de Bogotá en 2018, desagregado por municipio

	Núm. Celdas	Densidad habs. 2018 (km ²)	MAE
Área Metropolitana	3.179	2.758,2	694,7
Bogotá, D.C.	1.830	4.050,6	736,8
Soacha	190	3.474,6	846,3
Mosquera	106	1.228,5	1.553,0
Chía	82	1.612,0	846,7
Cajicá	45	1.827,6	618,1
Madrid	156	719,6	815,5
Cota	46	710,7	1.079,4
Funza	73	1.276,1	1.491,3
Tocancipá	104	384,6	654,1
Gachancipá	49	347,5	293,3
Sopó	123	209,6	248,5
La Calera	375	79,6	103,0

Se verificó por medio del cálculo de un coeficiente de correlación de Pearson que la densidad de habitantes por kilómetro cuadrado no estaba linealmente relacionado al MAE ni al área en kilómetros cuadrados (que coincide con el número de celdas).

Se observa también en la tabla 6-3 que los MAE pueden variar considerablemente en relación al municipio sobre el que se aplique el modelo y que si se compara esta tabla con las áreas del mapa de la figura 6-2, se aprecia que los municipios con poca área urbana como lo son Madrid, La Calera, Cajicá, Tocancipá, Sopó, Tocancipá y Gachancipá son los que coincidentalmente tienen un mejor MAE. Esto último coincide, además, con los resultados

expuestos en la tabla **6-2**, pues se obtuvieron mejores MAE en las imágenes de la muestra de Colombia en 2018 con menor número de habitantes.

Las evaluaciones de los resultados que han sido expuestas hasta el momento han mostrado la aplicabilidad del modelado obtenido bajo un mismo periodo de estudio, en este caso, el 2018, la siguiente subsección muestra que es también factible usar el modelado en otros periodos de estudio.

6.3.1. Comparación de las predicciones con las series de proyecciones de población del DANE

Teniendo a disposición los conjuntos de datos de prueba para el área metropolitana de Bogotá en los periodos 2016, 2018 y 2020, se pudo explorar la similitud de las cifras de población arrojadas por las fuentes oficiales del DANE ¹ y las predicciones agregadas por municipio que arroja el modelo obtenido. Cabe anotar que esta comparación es meramente exploratoria, pues las cifras de proyección se encuentran, por lo general, por encima de las cifras de los conteos censales sobre los que se basa este estudio, puesto que el modelado demográfico para realizar proyecciones de población incluye el ajuste por áreas de la omisión censal ², Sin embargo, esta comparación sopesa la relevancia que tiene el modelo para realizar predicciones de población. En las tablas **6-4** y **6-5** se comparan las series de predicciones y las proyecciones de población para los periodos 2016, 2018 y 2020 en cada uno de los municipios del área metropolitana de Bogotá.

En la tabla **6-5** se observa que en los municipios de carácter urbano como Bogotá y Soacha tienden a ser sub-estimados, mientras que el resto muestra una sobre-estimación. Estas diferencias se ven mucho más marcadas en los periodos 2016 y 2020 que son los no coincidentes con el periodo del conjunto de datos de entrenamiento. Tanto en el área metropolitana de Bogotá, como en el municipio de Bogotá, se observa especialmente una notable sub-estimación para el periodo 2020. Cabe reconocer que si se observa el mapa de la figura **6-2**, el municipio de Bogotá es más extenso en área que el resto, por lo que en este mismo municipio se podrían presentar más particularidades de las que esta tabla ofrece. A diferencia de las proyecciones, las predicciones no siguen necesariamente una tendencia creciente, tal y como se observa en la tabla **6-4**. Esta es una condición a la que se encuentran sujetas las proyecciones demográficas. Además, las diferencias tienen a incrementarse en los periodos 2016 y 2020 que son aquellos con los que no se entrenó el modelo. Solamente los municipios de Chía, Cajicá y Madrid muestran la tendencia de crecimiento.

¹<https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>

²<https://www.dane.gov.co/files/censo2018/informacion-tecnica/CNPV-2018-Poblacion-Ajustada-por-Cobertura.xls>

Tabla 6-4.: Predicciones por municipio y periodo en el área metropolitana de Bogotá comparadas con las proyecciones de población del DANE

	Proyecciones DANE			Predicción * PPS+LeNet-5		
	2016	2018	2020	2016	2018	2020
Área Metropolitana	8.503.304	8.768.162	9.286.927	9.127.831	8.851.215	7.587.058
Bogotá, D.C.	7.300.918	7.412.566	7.743.955	7.321.118	7.114.533	5.719.849
Soacha	583.007	660.179	753.548	567.985	560.026	434.707
Mosquera	114.348	130.221	150.665	221.487	293.515	279.519
Chía	118.119	132.181	149.570	158.254	164.666	187.880
Cajicá	73.490	82.244	92.967	79.314	80.195	98.833
Madrid	99.781	112.254	127.138	185.092	207.468	275.241
Cota	29.221	32.691	36.992	91.602	72.438	109.824
Funza	83.139	93.154	105.086	187.311	143.197	208.501
Tocancipá	35.666	39.996	45.714	112.753	98.627	107.172
Gachancipá	15.147	17.026	19.376	29.495	21.013	18.161
Sopó	23.146	25.782	28.999	50.994	46.540	63.574
La Calera	27.322	29.868	32.917	122.425	48.998	83.796

Tabla 6-5.: Diferencia entre las predicciones por municipio y período en el área metropolitana de Bogotá y las proyecciones de población del DANE

	2016	2018	2020
Área Metropolitana	624.527	83.053	-1.699.869
Bogotá, D.C.	20.200	-298.033	-2.024.106
Soacha	-15.022	-100.153	-318.841
Mosquera	107.139	163.294	128.854
Chía	40.135	32.485	38.310
Cajicá	5.824	-2.049	5.866
Madrid	85.311	95.214	148.103
Cota	62.381	39.747	72.832
Funza	104.172	50.043	103.415
Tocancipá	77.087	58.631	61.458
Gachancipá	14.348	3.987	-1.215
Sopó	27.848	20.758	34.575
La Calera	95.103	19.130	50.879

La obtención de estos resúmenes de las predicciones por municipio en el área metropolitana de Bogotá, dan cumplimiento al tercer objetivo específico que fue trazado para esta investigación.

7. Conclusiones y recomendaciones

7.1. Conclusiones

Luego de replicar las investigaciones de Doupe et al. y de Robinson et al. con los RSD y la malla de población del CNPV2018 para Colombia, se ha podido apreciar que los resultados en relación a la métrica MAE llegan a ser comparables en el caso del área metropolitana de Bogotá de 2018 e incluso superiores en el caso de la muestra de celdas del 2018 en Colombia.

Se logra destacar que la regresión con imágenes de satélite sí ofrece una alternativa para la predicción de población, particularmente, en áreas pequeñas como las que se pueden delimitar con ayuda de una malla de población, pues el modelado funciona y puede seguir mejorando con una investigación más extendida con otras fuentes de RSD y de mallas de población, así como con una experimentación más exhaustiva con la afinación de hiperparámetros de las CNN.

En cuanto a la afirmación anterior, se pudo apreciar que las arquitecturas alternativas a la VGG-A con menos capas son una opción para realizar regresión con imágenes de satélite; si bien las arquitecturas VGG son las que han mostrado el mejor desempeño predictivo para el reconocimiento y clasificación en imágenes [25], se pudo observar que esto no es extendible en la misma manera para tareas de regresión. Particularmente, se obtuvo que la arquitectura modificada de la LeNet-5 fue capaz de aprender características que determinan el número de habitantes a partir de las imágenes de Landsat 8 en Colombia.

Si se compara esta investigación con los trabajos de Doupe et al. y de Robinson et al., se aprecia que el diseño muestral en el área de estudio afecta directamente en las métricas de desempeño, bien si son estadísticas que miden la variación absoluta o la variación relativa de los residuales.

A partir de la revisión bibliográfica que antecedió esta investigación, se puede concluir también que la investigación en el uso de técnicas de DL con RSD es incipiente puesto que, al momento, se ha optado por utilizar modificaciones de las arquitecturas de CNN con mejor desempeño predictivo en conjuntos de datos que no se relacionan a los RSD. Así mismo, la investigación en la aplicación de CNNs para realizar regresión sigue siendo incipiente, pues no existen referencias de arquitecturas específicamente diseñadas para realizar regresión. Si

se menciona además la estrategia utilizada por Robinson et al. de categorizar la variable de salida para aplicar la regresión, el desempeño de las estadísticas de variación absoluta no están mostrando un buen desempeño en relación al trabajo de Doupe et al. y el presente trabajo.

7.2. Recomendaciones

El presente trabajo deja algunas indicaciones a seguir para investigaciones que se encaminen a realizar regresión con CNNs a partir de RSD.

En casos en los que se gestione manualmente la captura de información, es pertinente establecer e implementar procesos de verificación para la correcta captura de estos, pues el factor humano incluido en este proceso puede afectar negativamente la calidad de los datos. Cabe mencionar que durante el proceso de captura de datos de esta investigación, se cometieron errores de etiquetado en los que se obtuvieron conjuntos de imágenes inservibles y que no había forma de recuperar.

Luego de observar la figura **6-2** del mapa de residuales del área metropolitana de Bogotá, se aprecia lo importante que es especificar y delimitar el área de estudio al mejor detalle posible, de forma que se reduzca la variabilidad en la densidad de la población de las celdas. Se sugiere, por ejemplo, modelar por aparte, las zonas rurales y las zonas urbanas. Esto también incluye utilizar estrategias de muestreo, tales como incrementar el tamaño de muestra, calibrar o replantear el diseño muestral y experimentar con técnicas de aumento de datos permitiría mejorar la predicción en aquellas áreas que afectan negativamente las métricas de desempeño del modelo, en particular, las áreas con densidades de población muy baja y muy alta que se observan en la figura **6-2** del mapa de residuales del área metropolitana de Bogotá.

Queda abierta la posibilidad de incluir otras capas ráster que sirvan de datos auxiliares sobre la distribución espacial de la población. Por ejemplo, los modelos digitales de elevación (DEM, por su sigla en inglés), modelos topográficos o también imágenes satelitales de luz nocturna, dado el manejo flexible que tienen las CNNs respecto a las imágenes compuestas por múltiples canales.

Se abre además la posibilidad de estudiar conjuntos de datos de otros sensores remotos de libre acceso y de explorar el modelado con otras resoluciones radiométricas, espectrales y espaciales. Tanto la inclusión de diferentes capas de información así como datos de otros sensores remotos diferentes a los del proyecto Landsat, deja al pendiente verificar la pertinencia y el aporte de información de las bandas y las capas a incluir dentro del modelado que se realice y, particularmente, para la predicción de número de habitantes.

El modelado planteado en este trabajo puede servir también para realizar predicciones en factores asociados al número de habitantes y que llegan a ser igualmente importantes para la gestión pública como lo son: número de edificaciones, número de viviendas o número de hogares. Como muestra el trabajo de Doupe et al., puede ser incluso pertinente el incluir, en el entrenamiento, datos de mallas de población y de RSD para otras países, abriendo también la posibilidad de realizar *Transfer Learning*.

A. Anexo: Productos generados

Producto	Vínculo
Método de descarga de composiciones de Landsat 8 en GEE	https://code.earthengine.google.com/?scriptPath=users%2Famonterol%2Fprueba1%3Adownload_Landsat8_Tier1_TOA_v2
Robot para la descarga automatizada de imágenes desde GEE	https://drive.google.com/file/d/1v88r-9u-uVeEcLIICHXHaDd-UxCOZuAm/view?usp=sharing
Método de preprocesamiento	https://drive.google.com/drive/folders/1KSPamCKuenyprhIaMK9rk7Rm_S6Yvrs-?usp=sharing
Método de modelado	https://drive.google.com/drive/folders/1kb81cWqR3jMLUZjMMzpTgqkJ10UZIRLv?usp=sharing
Modelo LeNet-5 obtenido	https://drive.google.com/file/d/1MNV_e2FA9-t4lgYmQ10y-7U7E8GFo4Mj/view?usp=sharing
Conjuntos de datos e imágenes	https://drive.google.com/drive/folders/1T9emWTPa1C07idkivC2P_7gin80Uwlf8?usp=sharing

Bibliografía

- [1] S. Amaral, A. A. Gavlak, M. I. S. Escada, and A. M. V. Monteiro, “Using remote sensing and census tract data to improve representation of population spatial distribution: Case studies in the Brazilian Amazon,” *Population and Environment*, vol. 34, no. 1, pp. 142–170, sep 2012. [Online]. Available: <https://link-springer-com.ezproxy.unal.edu.co/article/10.1007/s11111-012-0168-2>
- [2] Eurostat, “Glossary:Population grid cell - Statistics Explained.” [Online]. Available: <https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Population{-}grid{-}cell>
- [3] —, “Population grids - Statistics Explained.” [Online]. Available: <https://ec.europa.eu/eurostat/statistics-explained/index.php/Population{-}grids>
- [4] C. Robinson, F. Hohman, and B. Dilkina, “A deep learning approach for population estimation from satellite imagery,” in *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities, GeoHumanities 2017*. Association for Computing Machinery, Inc, nov 2017, pp. 47–54.
- [5] M. Ferguson, R. Ak, Y.-T. Lee, and K. Law, “Automatic localization of casting defects with convolutional neural networks,” 2017, pp. 1726–1735.
- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-Based Learning Applied to Document Recognition,” in *Proceedings of the IEEE*, vol. 86, no. 11, 1998, pp. 2278–2324. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.7665>
- [7] N. M. Short, U. S. N. Aeronautics, S. A. Scientific, T. I. Branch, E. R. R. S. A. C. (U.S.), and G. S. F. Center, *The Landsat Tutorial Workbook: Basics of Satellite Remote Sensing*, ser. NASA reference publication. National Aeronautics and Space Administration, Scientific and Technical Information Branch, 1982. [Online]. Available: <https://books.google.com.co/books?id=9RsrAAAAIAAJ>
- [8] M. A. Wulder, T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Allen, M. C. Anderson, A. S. Belward, W. B. Cohen, J. Dwyer, A. Erb, F. Gao, P. Griffiths, D. Helder, T. Hermosilla, J. D. Hipple, P. Hostert, M. J. Hughes, J. Huntington, D. M. Johnson, R. Kennedy, A. Kilic, Z. Li, L. Lyburner, J. McCorkel,

- N. Pahlevan, T. A. Scambos, C. Schaaf, J. R. Schott, Y. Sheng, J. Storey, E. Vermote, J. Vogelmann, J. C. White, R. H. Wynne, and Z. Zhu, “Current status of Landsat program, science, and applications,” *Remote Sensing of Environment*, vol. 225, pp. 127–147, may 2019.
- [9] E. European Space Agency, *Sentinel-2 User Handbook*. European Space Agency, 2015.
- [10] P. Doupe, E. Bruzelius, J. Faghmous, and S. G. Ruchman, “Equitable Development through Deep Learning: The Case of Sub-National Population Density Estimation,” in *Proceedings of the 7th Annual Symposium on Computing for Development*, ser. ACM DEV ’16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: <https://doi.org/10.1145/3001913.3001921>
- [11] F. Batista e Silva, J. Gallego, and C. Lavallo, “A high-resolution population grid map for Europe,” *Journal of Maps*, vol. 9, no. 1, pp. 16–28, 2013.
- [12] L. Wang and X. Li, “Population estimation with remote sensing,” in *Comprehensive Remote Sensing*. Elsevier, jan 2017, vol. 1-9, pp. 59–66.
- [13] R. C. Taragi, K. S. Bisht, and B. S. Sokhi, “Generating population census data through aerial remote sensing,” *Journal of the Indian Society of Remote Sensing*, vol. 22, no. 3, pp. 131–138, sep 1994. [Online]. Available: <https://link-springer-com.ezproxy.unal.edu.co/article/10.1007/BF03024774>
- [14] K. Karume, C. Schmidt, K. Kundert, M. E. Bagula, B. F. Safina, R. Schomacker, D. Ganza, O. Azanga, C. Nfundiko, N. Karume, and G. N. Mushagalusa, “Use of Remote Sensing for Population Number Determination,” *The Open Access Journal of Science and Technology*, vol. 05, no. 03, 2017.
- [15] B.-g. Zhang, “Application of remote sensing technology to population estimation,” *Chinese Geographical Science*, vol. 13, no. 3, pp. 267–271, sep 2003. [Online]. Available: <https://link-springer-com.ezproxy.unal.edu.co/article/10.1007/s11769-003-0029-0>
- [16] A. Sorichetta, G. M. Hornby, F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem, “High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020,” *Scientific Data*, vol. 2, sep 2015.
- [17] S. Zhao, Y. Liu, R. Zhang, and B. Fu, “China’s population spatialization based on three machine learning models,” *Journal of Cleaner Production*, vol. 256, p. 120644, may 2020.
- [18] F. HORTON, “Remote sensing techniques and urban acquisition,” *ESTESJ, SENGERL(eds.). Remote Sensing Techniques*, pp. 243–276, 1974.

- [19] E. C. B. L. F. Curtis, "Introduction to environmental remote sensing," *Chapman and Hall Ltd*, pp. 238–241, 1986.
- [20] A. E. Ayila, B. Paul Babatunde, and J. P. Yohanna, "Population estimation and census track demarcation in Hwolshe, Plateau State, Nigeria: A geospatial approach," *Remote Sensing Applications: Society and Environment*, vol. 10, pp. 183–189, apr 2018.
- [21] C. T. Lloyd, A. Sorichetta, and A. J. Tatem, "Data Descriptor: High resolution global gridded data for use in population studies," *Scientific Data*, vol. 4, no. 1, pp. 1–17, jan 2017. [Online]. Available: www.nature.com/sdata/
- [22] A. Dmowska and T. F. Stepinski, "A high resolution population grid for the conterminous United States: The 2010 edition," *Computers, Environment and Urban Systems*, vol. 61, pp. 13–23, jan 2017.
- [23] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, J. Gao, and L. Zhang, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sensing of Environment*, vol. 241, p. 111716, may 2020.
- [24] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land Use Classification in Remote Sensing Images by Convolutional Neural Networks," aug 2015. [Online]. Available: <http://arxiv.org/abs/1508.00092>
- [25] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for large-scale image recognition," *American Journal of Health-System Pharmacy*, vol. 75, no. 6, pp. 398–406, 2018.
- [26] DANE, "Guía para la anonimización de bases de datos en el Sistema Estadístico Nacional," Bogotá, D.C., p. 72, 2018.
- [27] G. Chander, B. L. Markham, and D. L. Helder, "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors," *Remote Sensing of Environment*, vol. 113, no. 5, pp. 893–903, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425709000169>
- [28] A. H. Bagour, "Probability Proportional to size Sampling," Ph.D. dissertation, Oklahoma State University, 2004.
- [29] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019. [Online]. Available: <https://doi.org/10.1186/s40537-019-0197-0>
- [30] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 2019.

-
- [31] K. Jordahl, J. V. den Bossche, M. Fleischmann, J. Wasserman, J. McBride, J. Gerard, J. Tratner, M. Perry, A. G. Badaracco, C. Farmer, G. A. Hjelle, A. D. Snow, M. Cochran, S. Gillies, L. Culbertson, M. Bartos, N. Eubank, Maxalbert, A. Bilogur, S. Rey, C. Ren, D. Arribas-Bel, L. Wasser, L. J. Wolf, M. Journois, J. Wilson, A. Greenhall, C. Holdgraf, Filipe, and F. Leblanc, “geopandas/geopandas: v0.8.1,” 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3946761>
- [32] T. Pandas development team, “pandas-dev/pandas: Pandas,” 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3509134>
- [33] QGIS Development Team, *QGIS Geographic Information System*, Open Source Geospatial Foundation, 2009. [Online]. Available: <http://qgis.org>
- [34] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [35] F. Chollet and Others. (2015) Keras. [Online]. Available: <https://github.com/fchollet/keras>
- [36] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.