

Santiago de Cali, Día de Mes del Año

Doctor

Diego Luis Linares O.

Director Maestría en Ciencia de Datos

Facultad de Ingeniería y Ciencias

Pontificia Universidad Javeriana de Cali

Asunto: Presentación para evaluación del proyecto aplicado

Cordial Saludo,

Con el fin de cumplir con los requisitos exigidos por la Universidad para optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto denominado "DESARROLLO DE MODELO DE CLASIFICACIÓN DE SUELO URBANO RECREATIVO BASADO EN DEEP LEARNING USANDO IMÁGENES SATELITALES", el cual fue realizado por el (los) estudiante (s) Nicolás Castaño Cardona con código 8986387, Jairo David Chia Bejarano con código 8986172 pertenecientes a la Maestría en Ciencia de Datos, bajo la dirección de Omar Andrés Castaño Idarraga identificado con C.C. 1214722914.

El suscrito director del Proyecto Aplicado autoriza para que se proceda a hacer la evaluación de este proyecto, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado y sustentado oficialmente.

Atentamente,

Nicolás Castaño Cardona



Nicolás Castaño Cardona
C.C. 1098817224 de
Bucaramanga

Jairo David Chia Bejarano
C.C. 80250210 de Bogotá

Omar Andrés Castaño Idarraga
C.C. 1214722914 de Medellín

Documentación anexa:

Resumen del Proyecto Aplicado en formato digital (máximo 1 página).

Una copia digital (PDF) del documento del proyecto aplicado

FICHA RESUMEN

POSIBLE TÍTULO: DESARROLLO DE MODELO DE CLASIFICACIÓN DE SUELO URBANO RECREATIVO BASADO EN DEEP LEARNING USANDO IMÁGENES SATELITALES

1. ÁREA DE TRABAJO: Planeación Urbana, Redes Neuronales Convolucionales
2. TIPO DE PROYECTO (Aplicado, Innovación, Investigación): Aplicado
3. ESTUDIANTE(S): Nicolas Castaño Cardona, Jairo David Chía Bejarano
4. CORREO ELECTRÓNICO: nicolascastacardona@gmail.com, elchivet@javerianacali.edu.co
5. DIRECCIÓN Y TELEFONO: Av. calle 161 # 12b – 30 (Bogotá, Colombia), 3148071501, Av. Calle 26 Nro. 73-64 (Bogotá, Colombia), 3156971120
6. DIRECTOR: Omar Castaño
7. VINCULACIÓN DEL DIRECTOR: Universidad Pontificia Javeriana (Sede Cali)
8. CORREO ELECTRÓNICO DEL DIRECTOR: omar.castano@udea.edu.co
9. CO-DIRECTOR (Si aplica):
10. GRUPO O EMPRESA QUE LO AVALA (Si aplica):
11. OTROS GRUPOS O EMPRESAS:
12. PALABRAS CLAVE (al menos 5): Clasificación, Imágenes, Satelitales, Aprendizaje, Profundo, Deep, Learning
13. FECHA DE INICIO: 28 de Enero de 2024
14. DURACIÓN ESTIMADA (En meses): 10
15. RESUMEN:

El presente documento se centra en el planteamiento de un proyecto que se enfoca en el fenómeno global del rápido crecimiento urbano, donde la expansión de las ciudades surge de manera acelerada y con ella la necesidad e importancia de tener zonas verdes y espacios recreativos dentro de ellas. La falta de información precisa sobre la ubicación y extensión de estos espacios ha llevado a deficiencias en la planificación urbana; en respuesta a esta problemática, se propone la implementación de un modelo automático basado en Redes Neuronales Convolucionales (CNN) para la detección y clasificación de parques y canchas deportivas en imágenes satelitales urbanas. El objetivo fundamental es estimar la cantidad de áreas recreativas en diferentes zonas urbanas, proporcionando datos cruciales para respaldar políticas públicas de expansión y mejorar la calidad de vida en entornos urbanos. Con el fin de llevar a cabo este propósito, inicialmente se planteó el diseño y entrenamiento de un modelo de CNN que clasifica

y cuenta estas áreas y espacios recreativos, para posteriormente evaluar y comparar la efectividad del modelo propuesto con modelos de clasificación. Los resultados esperados abarcan la exitosa implementación del modelo en un repositorio público, la creación de una documentación detallada del proceso, y el desarrollo de scripts que faciliten la replicabilidad del enfoque propuesto y de manera propositiva, la creación de una herramienta interactiva que facilite la implementación del modelo y los resultados obtenidos.



Pontificia Universidad
JAVERIANA
Cali

**DESARROLLO DE MODELO DE CLASIFICACIÓN DE SUELO URBANO RECREATIVO
BASADO EN DEEP LEARNING USANDO IMÁGENES SATELITALES**

*Jairo David Chía Bejarano
Nicolás Castaño Cardona*

*Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos*

Director(a)
Omar Andrés Castaño Idarraga

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, DICIEMBRE 1 DE 2024

TABLA DE CONTENIDO

INTRODUCCIÓN	7
1. DEFINICIÓN DEL PROBLEMA	8
PLANTEAMIENTO DEL PROBLEMA	8
FORMULACIÓN DEL PROBLEMA.....	8
2. OBJETIVOS DEL PROYECTO	10
2.1. OBJETIVO GENERAL	10
2.2. OBJETIVOS ESPECÍFICOS	10
3. MARCO TEÓRICO Y ANTECEDENTES.....	11
3.1. MARCO TEÓRICO	11
3.2. ANTECEDENTES	15
4. IMPLEMENTACION DEL MODELO CLASIFICACIÓN DE REDES NEURONALES CONVOLUCIONALES E IMPLEMENTANCIÓN SORE ESPACIOS PÚBLICOS RECREATIVOS EN SUELO URBANO	19
4.1. Objetivo específico 1	19
4.2. Objetivo específico 2	27
4.3. Objetivo específico 3	30
5. CONCLUSIONES Y TRABAJOS FUTUROS.....	30
5.1. CONCLUSIONES	30
5.2. TRABAJOS FUTUROS.....	30
6. REFERENCIAS BIBLIOGRÁFICAS	31

LISTA DE FIGURAS

LISTA DE TABLAS

LISTA DE ANEXOS

INTRODUCCIÓN

El acelerado crecimiento urbano experimentado en América Latina en las últimas décadas ha ejercido gran presión sobre los espacios públicos destinados a áreas verdes y recreación como parques, plazas y canchas deportivas. La creciente densificación de las ciudades no se ha visto acompañada de una adecuada ampliación de estas zonas para dar cobertura a la mayor población, resultando frecuentemente insuficientes.

Esta situación se ve agravada por la falta de información precisa y actualizada sobre la localización y extensión territorial de los existentes espacios recreativos urbanos que permita una adecuada planificación. Los métodos tradicionales de machine learning en teledetección, como uso de índices de vegetación o clasificación digital mediante Random Forest, presentan limitaciones para identificación detallada en imágenes satelitales de alta resolución debido a la confusión espectral con otros objetos.

Ante esta problemática, el presente proyecto tiene como propósito implementar un modelo automático para la detección y clasificación de parques y canchas deportivas en imágenes satelitales de zonas urbanas mediante la aplicación de redes neuronales convolucionales y técnicas de aprendizaje profundo.

Específicamente, se ha implementado una Red Neuronal Convolucional (CNN) capaz de extraer características distintivas de los objetos de interés, buscando superar los métodos convencionales antes mencionados, evaluando su efectividad en términos de precisión y conteo. Esta CNN optimizada ha sido entrenada con un riguroso conjunto de datos etiquetados y que han sido dispuestos para el uso público.

Por medio de este proyecto aplicamos un modelo que permite estimar la cantidad de estas áreas recreativas en diferentes zonas de la ciudad, el cual puede servir como base para generar posteriormente datos actualizados de cobertura espacial y así detectar posibles déficits en algunas regiones, información que es valiosa para sustentar políticas públicas de expansión o construcción ante el crecimiento poblacional. Además, este proyecto puede orientar a posteriores usuarios de este en la aplicación de esta propuesta de modelo de manera diferente en otros aspectos urbanos de las ciudades.

Los scripts y la estructura del modelo están publicados en repositorios públicos para uso de la comunidad con su respectiva documentación, además de una guía detallada que permite orientar a posteriores usuarios del modelo en su aplicación. Así se espera proveer una solución complementaria para automatizar el monitoreo de zonas recreativas urbanas, fundamentales para la habitabilidad y salud pública en contextos de alta urbanización

1. DEFINICIÓN DEL PROBLEMA

PLANTEAMIENTO DEL PROBLEMA

El espacio público urbano, que abarca áreas como parques, plazas y escenarios culturales al aire libre, desempeña un papel crucial en la calidad y planificación urbana al facilitar el intercambio social y promover actividades recreativas. Aunque la evaluación tradicional se basa en indicadores cuantitativos y cualitativos, como los déficits de espacio público, estas métricas revelan limitaciones para comprender problemas urbanos complejos [1]. Investigaciones recientes destacan la necesidad de enfoques más avanzados para una evaluación precisa, resaltando la importancia de considerar detalladamente el espacio público en el contexto de la planificación urbana para lograr ciudades más habitables y funcionales.

El mapeo preciso de usos de suelo urbano mediante la identificación de áreas verdes y recreativas a partir de imágenes satelitales constituye un desafío crítico en la planificación y gestión urbanas. Los métodos tradicionales de clasificación automática, como Random Forest y SVM, han demostrado limitaciones en la detección precisa de ciertas clases de objetos y espacios [2]. La incapacidad de estos enfoques para lograr una clasificación de imágenes detallada impide la generación eficiente de inventarios georreferenciados de tales áreas urbanas.

La situación actual se ve agravada por la creciente necesidad de datos precisos para respaldar decisiones de planificación y políticas urbanas, subrayando la urgencia de avanzar en la capacidad de clasificación y segmentación de imágenes satelitales. Este problema, en el ámbito de la ciencia de datos, requiere una solución que vaya más allá de los métodos convencionales de machine learning. La relevancia de esta investigación radica en su potencial para mejorar significativamente la precisión del mapeo de usos de suelo urbano, especialmente en la identificación de espacios públicos, facilitando así una planificación urbana más informada y eficaz.

FORMULACIÓN DEL PROBLEMA

¿Cómo implementar un modelo de clasificación de imágenes a partir de Redes Neuronales Convolucionales (CNN) que permita clasificar, identificar y contar espacios públicos abiertos, como parques y canchas deportivas, a partir de imágenes satelitales?

En el contexto de la planificación urbana, la identificación precisa de usos de suelo y, en particular, la clasificación detallada de espacios públicos abiertos, como parques y canchas deportivas, es esencial para una gestión efectiva de las ciudades. Los métodos convencionales de clasificación automática de imágenes satelitales han demostrado limitaciones en la detección de estas áreas específicas, lo que destaca la necesidad de enfoques más avanzados.

Teniendo en cuenta esto, surgen las siguientes incógnitas: ¿Cómo podemos diseñar y entrenar un modelo de redes neuronales convolucionales que obtenga mejores resultados que los obtenidos por los métodos tradicionales de machine learning para la clasificación de estas áreas urbanas? ¿Cuál es la efectividad comparativa de este modelo frente a métodos tradicionales de machine learning, en términos de precisión y detección de espacios públicos? ¿Cómo se desempeña el modelo de Redes Neuronales Convolucionales (CNN) cuando se aplica en un entorno práctico utilizando imágenes satelitales actuales para la identificación y clasificación de espacios públicos recreativos en un área urbana específica?

Resolver estas preguntas orientaron el proyecto hacia el desarrollo de un enfoque innovador en ciencia de datos para el mapeo detallado de usos de suelo urbano, particularmente en la identificación precisa de espacios públicos recreativos, contribuyendo así a una planificación urbana más informada y eficaz.

2. OBJETIVOS DEL PROYECTO

2.1. OBJETIVO GENERAL

Implementar un modelo clasificación de imágenes a partir de Redes Neuronales Convolucionales (CNN) que identifique, clasifique y cuente espacios públicos recreativos en suelo urbano haciendo uso de imágenes satelitales.

2.2. OBJETIVOS ESPECÍFICOS

- Desarrollar el modelo de redes neuronales convolucionales incorporando técnicas avanzadas de Deep Learning para lograr una clasificación precisa de áreas recreativas en imágenes satelitales urbanas.
- Evaluar la efectividad del modelo propuesto mediante comparaciones rigurosas contra métodos tradicionales de clasificación automática, como por ejemplo Random Forest, en términos de identificación, precisión y conteo de espacios públicos recreativos urbanos.
- Aplicar el modelo de Redes Neuronales Convolucionales (CNN) desarrollado en un entorno práctico, utilizando imágenes satelitales actuales, y evaluar su desempeño en la identificación y clasificación de espacios públicos recreativos en un área urbana específica.

3. MARCO TEÓRICO Y ANTECEDENTES

3.1. MARCO TEÓRICO

En este capítulo se establece un fundamento teórico relacionado con la investigación sobre el mapeo detallado de usos de suelo urbano, enfocado en la identificación de espacios públicos abiertos. Se abordarán definiciones clave y desafíos en la planificación urbana, seguido de un análisis de métodos tradicionales y las limitaciones en su precisión. Se explorarán algunos casos en la aplicación exitosa de Deep Learning y Redes Neuronales Convolucionales en la clasificación de objetos en imágenes. La evaluación comparativa entre modelos convencionales y basados en Deep Learning, junto con estrategias para mejorar la robustez, completará esta sección.

3.1.1. Introducción al Mapeo de Usos de Suelo Urbano

La complejidad de las áreas urbanas contemporáneas demanda una comprensión minuciosa y actualizada de los usos de suelo que las componen. El mapeo de usos de suelo urbano emerge como un componente crucial en la planificación y gestión efectiva de ciudades en constante cambio y crecimiento. Este proceso no solo se limita a una mera catalogación de las diferentes funciones y actividades presentes en el entorno urbano, sino que representa una herramienta poderosa para abordar desafíos inherentes a la expansión urbana, la optimización de recursos y la creación de entornos urbanos habitables y sostenibles.

En el contexto de este proyecto, se da especial énfasis a la definición de usos de suelo urbano como un factor determinante para la toma de decisiones en el desarrollo urbano sostenible. La precisión en la delimitación de estos usos no solo es esencial para la adecuada planificación de infraestructuras y servicios, sino que también impacta directamente en la calidad de vida de los habitantes. En este sentido, uno de los elementos cruciales dentro del amplio espectro de usos de suelo es la identificación de espacios públicos abiertos.

La importancia de los espacios públicos abiertos en la configuración de entornos urbanos saludables y accesibles no puede ser subestimada. Más allá de su función estética, estos espacios sirven como puntos de encuentro, recreación y actividad física, contribuyendo a la cohesión social y al bienestar general de la población urbana. Los gobiernos, por medio de sus entidades públicas, establecen parámetros y herramientas que le permite a las ciudades poder desarrollarse conforme a estructuras ordenadas que permiten dar cabida a espacios que contribuyen a la realización de actividades de bienestar social; tal es el caso de los Planes de Ordenamiento Territorial en Colombia, cuyos principios establecen una directriz para la adecuada distribución del uso del suelo en el espacio dentro del ordenamiento territorial de municipios y distritos.

No obstante, los desafíos asociados con la identificación precisa y mapeo detallado de estos

espacios en el contexto urbano son significativos. La necesidad de superar estas dificultades se convierte en un catalizador para la investigación propuesta, donde se busca avanzar en la eficacia de la identificación de espacios públicos abiertos mediante enfoques innovadores basados en técnicas de aprendizaje profundo y Redes Neuronales Convolucionales.

3.1.2. Clasificación Automática de objetos y espacios en Imágenes Satelitales

La clasificación automática de objetos y espacios en imágenes satelitales es un tema de gran interés en la ciencia de datos. Los métodos tradicionales de machine learning, como Random Forest y SVM, han sido ampliamente utilizados para este propósito. Sin embargo, estos métodos tienen limitaciones en la precisión de la clasificación de objetos y espacios en imágenes satelitales. Algunas de las limitaciones y desafíos que se presentan en estos métodos tradicionales son la falta de capacidad para manejar grandes cantidades de datos, la necesidad de una gran cantidad de recursos computacionales y la dificultad para manejar datos no estructurados [4]. En los últimos años, se han desarrollado avances en técnicas de machine learning para la clasificación de objetos y espacios en imágenes satelitales. Estos avances incluyen el uso de Redes Neuronales Convolucionales (CNN), que han demostrado ser muy efectivas en la clasificación de objetos y espacios en imágenes satelitales [5]. Las CNN son capaces de aprender características de las imágenes satelitales a través de múltiples capas, lo que les permite identificar patrones complejos en los datos.

La clasificación detallada de objetos complejos en escenas urbanas es un desafío importante en la ciencia de datos. Aunque se han desarrollado técnicas avanzadas de machine learning para la clasificación de objetos y espacios en imágenes satelitales, la clasificación detallada de objetos complejos en escenas urbanas sigue siendo un problema difícil de resolver. Algunas de las limitaciones que se presentan en la clasificación detallada de objetos complejos en escenas urbanas son la falta de datos etiquetados, la variabilidad en la apariencia de los objetos y la complejidad de las escenas urbanas.

Para abordar estas limitaciones, se han propuesto diversas técnicas de machine learning, como el uso de CNN y el aprendizaje por transferencia. Sin embargo, estas técnicas aún no son capaces de clasificar detalladamente objetos complejos en escenas urbanas con la misma precisión que los humanos.

Algunos estudios han propuesto el uso de técnicas de aprendizaje profundo para la clasificación detallada de objetos complejos en escenas urbanas. Por ejemplo, en el artículo desarrollado por Rodríguez [6], se propone un enfoque basado en CNN para la clasificación detallada de objetos en escenas urbanas. Los resultados experimentales muestran que el enfoque propuesto supera a los métodos tradicionales de machine learning en términos de precisión.

3.1.3 Aplicación de Deep Learning y Redes Neuronales Convolucionales (CNN) en Visión

por Computadora aplicada en imágenes satelitales.

La visión por computadora es un campo de la inteligencia artificial que se enfoca en permitir a las computadoras interpretar y comprender el mundo visual. El Deep Learning es una rama del Machine Learning que se enfoca en el aprendizaje de representaciones de datos a través de múltiples capas de procesamiento. Las Redes Neuronales Convolucionales (CNN) son un tipo de red neuronal profunda que se utiliza comúnmente en aplicaciones de visión por computadora. Las CNN son capaces de aprender características de las imágenes a través de múltiples capas de convolución y Pooling, lo que les permite identificar patrones complejos en los datos.

En el contexto de visión por computadora, las CNN han demostrado ser muy efectivas en la clasificación de objetos en imágenes. Las CNN pueden detectar características y patrones de una imagen, lo que les permite reconocer objetos, clases y categorías. Las CNN se han utilizado en una amplia variedad de aplicaciones de visión por computadora, incluyendo la detección de objetos, la segmentación de imágenes, la clasificación de imágenes y la generación de imágenes. En la detección de objetos, las CNN se utilizan para identificar la presencia y ubicación de objetos en una imagen. En la segmentación de imágenes, las CNN se utilizan para dividir una imagen en regiones semánticas. En la clasificación de imágenes, las CNN se utilizan para asignar una etiqueta o categoría a una imagen. En la generación de imágenes, las CNN se utilizan para generar nuevas imágenes a partir de un conjunto de datos de entrenamiento.

Comparadas con métodos clásicos de machine learning, las CNNs presentan dos grandes diferencias: primero, aprenden estas representaciones directamente desde los datos sin requerir diseño manual de descriptores. Segundo, incorporan invariancia espacial al compartir pesos en las convoluciones, reduciendo el overfitting o sobre-ajuste [7].

Estas propiedades hacen que las CNNs sean particularmente útiles para clasificar escenas complejas en imágenes de satélite. Trabajos como el de Karypidis [8] demuestran desempeños muy superiores de CNNs frente a técnicas tradicionales de machine learning en tareas de clasificación de imágenes. El modelo basado en CNN demuestra mejores porcentajes de exactitud en cuanto a los resultados en varios casos de prueba.

Mediante operaciones de convolución, las CNNs incorporan nociones de jerarquía, compartiendo parámetros a través de las distintas capas para identificar patrones cada vez más abstractos y poder reducir el sobreajuste, logrando así una buena generalización a nuevos conjuntos de datos. Además, por su naturaleza diferenciable, las CNNs pueden entrenarse de manera supervisada a gran escala mediante el algoritmo de backpropagation y gradiente descendente para optimizar su rendimiento en la tarea objetivo. Estas bondades han impulsado la aplicación de CNNs en diversos estudios de percepción remota.

3.1.4 Evaluación de Modelos de Clasificación de Objetos en Imágenes Satelitales.

La clasificación de objetos en imágenes satelitales es una tarea importante en el procesamiento de imágenes y la teledetección. La evaluación de estos modelos es crucial para determinar la precisión y la eficacia de estos. En este sentido, se han propuesto varios métodos de evaluación de modelos de clasificación de objetos en imágenes satelitales, como la matriz de confusión, la curva ROC, entre otros [9].

Tradicionalmente, los clasificadores como Random Forests o SVM eran validados mediante particionamiento aleatorio en conjuntos de entrenamiento y test, y calculando métricas agregadas sobre dichas particiones como la exactitud (accuracy) global, matrices de confusión o curvas ROC. No obstante, en escenarios complejos con múltiples clases no balanceadas, se requieren métricas más granulares como la precisión y exhaustividad por clase.

La evaluación rigurosa de los modelos de clasificación de imágenes constituye una parte fundamental en su desarrollo, pues permite cuantificar y comparar su capacidad predictiva antes de ser implementados en aplicaciones del mundo real. En el contexto de la percepción remota y la clasificación de objetos urbanos en imágenes satelitales, han surgido nuevos desafíos evaluativos con la introducción de métodos de deep learning.

Con la adopción de técnicas más complejas como las CNNs que pueden sobre-ajustarse fácilmente, cobra importancia el uso de conjuntos independientes de validación, y métricas como el error en ese conjunto. La generalización del modelo a nuevos lugares resulta un factor de evaluación crítico, por lo que se analiza su desempeño frente a imágenes de áreas geográficas distintas.

En cuanto a la comparación entre modelos convencionales y deep learning para la clasificación de objetos en imágenes satelitales, se ha demostrado que los modelos de deep learning superan a los modelos convencionales en términos de precisión y eficacia. Los modelos de deep learning, como las redes neuronales convolucionales (CNN), son capaces de aprender características complejas de las imágenes satelitales y, por lo tanto, son más precisos en la clasificación de objetos en imágenes satelitales [10].

Los factores clave en la evaluación de modelos para la clasificación de objetos en imágenes satelitales incluyen la selección de características, la selección de algoritmos de clasificación, la selección de conjuntos de datos de entrenamiento y prueba, la selección de parámetros de modelos, la validación cruzada, entre otros. Estos factores permiten realizar comparativas sistemáticas entre distintos algoritmos bajo métricas diversas, utilizando particionamientos adecuados de los datos y evaluando la robustez del modelo, lo cual resulta ser indispensable para garantizar su capacidad predictiva antes de implementar clasificadores de objetos en imágenes satelitales.

3.2. ANTECEDENTES

El mapeo y clasificación de áreas urbanas a partir de imágenes de satélite es un campo de investigación activo. Los métodos tradicionales de aprendizaje automático como Random Forest, SVM se han utilizado comúnmente para la clasificación de usos del suelo, pero presentan limitaciones en la identificación precisa de ciertos tipos de objetos [11].

Los métodos de aprendizaje profundo (deep learning) han emergido recientemente como potentes herramientas para el análisis y clasificación de imágenes de percepción remota [12]. Comparados con enfoques estándar de aprendizaje automático, los modelos de deep learning como las redes neuronales convolucionales (CNNs) han demostrado obtener mayor precisión en tareas complejas de segmentación semántica y detección de objetos.

Hemos revisado varios artículos e investigaciones relacionadas con la implementación de Deep Learning para la clasificación de imágenes a partir de imágenes satélites y consideramos que los ejemplos descritos a continuación constituyeron una base sobre la cual logramos desarrollar nuestro proyecto.

3.2.1 Revisión sistemática de técnicas de deep learning para la clasificación de imágenes de percepción remota.

Sobre esto, Ying Li y otros presentan, en su artículo titulado “Deep learning for remote sensing image classification: A survey” [13], una revisión sistemática de técnicas de deep learning para la clasificación de imágenes de percepción remota. Inicia con una descripción de modelos como redes neuronales convolucionales (CNN), autoencoders apilados (SAE) y redes de creencia profunda (DBN), que permiten extraer representaciones abstractas directamente desde los datos.

Luego, el artículo se enfoca en dos ejes: clasificación a nivel de píxel de imágenes hiperespectrales, y clasificación de escenas en imágenes aéreas/satelitales ópticas. En relación con el primer eje, se revisan investigaciones recientes sobre uso de CNNs, SAEs y DBNs para extraer características espectrales, espaciales y espectral-espaciales en forma supervisada y no supervisada. Se discuten experimentos donde los métodos basados en información espectral-espacial conjunta suelen tener mejor precisión. Respecto a clasificación de escenas, el artículo describe trabajos que usan CNNs pre-entrenadas en otros conjuntos de datos, CNNs que se entrenan desde cero en los nuevos datos, y CNNs donde se hace fine-tuning. También cubre métodos no supervisados como autoencoders apilados. Los experimentos revelan que las CNN pre-entrenadas superan ampliamente a técnicas tradicionales con características manuales.

En términos de aportes a nuestra propuesta, este artículo nos permitió contextualizar el problema de clasificación de imágenes dentro del campo de deep learning aplicado a percepción remota.

Revisa avances recientes en detección de áreas verdes que pueden servir como punto de partida. Además, provee conjuntos de datos y métricas de evaluación de modelo que le servirán como referencia.

Sin embargo, el proyecto que estamos desarrollando tiene un enfoque más aplicado y específico en detección de canchas deportivas y parques en entornos urbanos usando imágenes satelitales. La investigación de Ying Li y su equipo, tiene un alcance más amplio, analizando también sensores hiperspectrales y radar. Además, nuestra investigación buscará desarrollar un modelo de CNN optimizado para esta tarea, mientras la investigación de Li sólo revisa aplicaciones previas de modelos genéricos. Nos ayudó a enmarcar nuestra investigación, pero con un objetivo más ampliado.

3.2.2 Detección basada en aprendizaje profundo de cambios en la cubierta forestal urbana junto con cambios urbanos generales utilizando imágenes satelitales de muy alta resolución.

En nuestra búsqueda, encontramos otro artículo llamado “Deep Learning-Based Detection of Urban Forest Cover Change along with Overall Urban Changes Using Very-High-Resolution Satellite Images” [14], el cual propone una técnica semántica de detección de cambios enfocada en los cambios en la cubierta forestal urbana, junto con otros cambios urbanos. Utiliza dos redes neuronales: Deeplabv3+ para la generación de máscaras binarias de bosques urbanos, y Deeply Supervised Image Fusion Network (DSIFN) para la detección binaria de cambios. Ambas redes se entrenan de forma independiente en conjuntos de datos públicos. Luego se realiza transfer learning con un conjunto de datos generado a partir de imágenes satelitales bitemporales de muy alta resolución (VHR) de tres ciudades urbanas, adquiridas con diferentes sensores satelitales.

Se plantea una metodología dividida en tres partes: generación de máscaras binarias de bosques urbanos mediante Deeplabv3+, generación de máscara binaria de cambios mediante DSIFN y monitoreo de cambios en cubierta forestal. En el primer paso, Esta red neuronal convolucional permite realizar una segmentación semántica precisa a nivel de píxel. Utiliza un codificador basado en ResNet-50 que extrae características de alto nivel, y un decodificador que recupera la resolución espacial mediante upsampling. Además, emplea un módulo de Atrous Spatial Pyramid Pooling (ASPP) para incorporar contexto multi-escala. Las imágenes pre y post cambio se procesan de forma independiente en Deeplabv3+ para obtener máscaras binarias de las áreas forestales urbanas. Luego se aplica un umbral para binarizar los resultados.

En el segundo paso, esta red neuronal implementa un esquema de fusión profunda de imágenes bitemporales para detección de cambios. Consta de dos flujos que extraen características profundas de las imágenes pre y post cambio de forma independiente. Luego, estas características se concatenan para alimentar una red discriminadora de diferencias. Adicionalmente, se incorporan características basadas en diferencia de imágenes. También se aplican módulos de atención espacial y canal para refinar las características. Finalmente, se genera una máscara binaria de cambios mediante upsampling.

Por último, en el monitoreo de cambios en cubierta forestal, Las máscaras binarias generadas

previamente se combinan para crear un resultado de cambio semántico enfocado en las áreas forestales. Esto se logra extrayendo las regiones de cambio forestal de las máscaras pre y post cambio usando la máscara de detección de cambios. Luego se concatenan con esta última para el resultado final. De esta forma se pueden detectar disminuciones y aumentos en las áreas forestales urbanas con respecto a las demás transformaciones urbanas. Así se logra un monitoreo conjunto de la cubierta forestal y los cambios globales en las escenas analizadas.

Esta metodología presentada en el artículo fue utilizada como una base valiosa para desarrollar el modelo de red neuronal convolucional (CNN) que buscamos implementar para la identificación de espacios públicos urbanos en imágenes satelitales. Específicamente, la aproximación en tres pasos facilita incorporar tanto la extracción de las áreas de interés a través de segmentación semántica, como su análisis multitemporal para evaluar cambios mediante detección de diferencias en las características extraídas. La arquitectura Deeplab v3+ demostró precisión en la detección de cubiertas forestales urbanas, por lo que una adaptación supervisada de este modelo mediante transfer learning podría ser efectiva para extraer parques y canchas deportivas.

Más allá de Deeplab v3+ y DSIFN, existen arquitecturas recientes de redes neuronales convolucionales que podrían aportar más valor para nuestra propuesta. Por ejemplo, redes como Mask R-CNN han demostrado altas precisiones en tareas combinadas de segmentación semántica y detección de objetos. Otra alternativa es emplear técnicas de Vision Transformers, las cuales aplican mecanismos de atención para modelar relaciones en toda la imagen, logrando una visión más holística de la escena. Esto puede facilitar la identificación conjunta de diferentes elementos urbanos.

Este artículo nos enseñó el potencial de combinar segmentación semántica y detección de cambios mediante deep learning para mapeo urbano. Sobre esa base, investigamos arquitecturas avanzadas de CNN orientadas a optimizar el desempeño en nuestra aplicación específica de clasificación y monitoreo de parques y canchas.

Sin embargo, el artículo tiene un enfoque más amplio en distintos tipos de cambios urbanos, mientras que lo que buscábamos se centraba específicamente en canchas deportivas y parques. Además, nuestro objetivo es desarrollar un modelo CNN optimizado para esta tarea, mientras que el artículo utiliza modelos genéricos. Finalmente, el artículo trabaja con distintos sensores, y nosotros queremos enfocarnos en imágenes ópticas para un análisis más detallado.

3.2.3 Deep Learning aplicado a imágenes satelitales como herramienta de detección de viviendas sin servicio de energía en el caserío Media Luna – Uribia – Guajira

Finalmente, encontramos un trabajo de investigación titulado “Deep Learning aplicado a imágenes satelitales como herramienta de detección de viviendas sin servicio de energía en el caserío Media Luna – Uribia – Guajira” [15]. Esta investigación aborda la problemática de estimar la cantidad de viviendas sin servicio en zonas rurales a partir de imágenes satelitales, debido a que el acceso y censo en estas áreas resulta complejo. Para automatizar esta tarea, se propone

crear un modelo de red neuronal convolucional (CNN) capaz de detectar techos en dichas imágenes.

El objetivo de la investigación es desarrollar un modelo de red neuronal convolucional (CNN) capaz de detectar automáticamente techos de viviendas sin servicio en imágenes satelitales de zonas rurales, para estimar la cantidad de hogares en esas áreas. Se comparan dos arquitecturas: una CNN personalizada de 6 capas, y la red VGG16 pre-entrenada en millones de imágenes. El conjunto de datos contiene 2000 imágenes de 60x60 píxeles con techos y lotes vacíos. Se dividió en 75% entrenamiento y 25% validación para optimizar los modelos.

Se realizó una etapa de preprocesamiento, donde se reescalaron los píxeles de 0-255 a 0-1 para acotar la dispersión. También se aplicó zoom, rotación y volteos aleatorios para hacer al modelo más robusto frente a variaciones. Las imágenes se agruparon en batches de 32 muestras para entrenar por lotes. La CNN aplica dos capas convolucionales apiladas extrayendo 32 y 64 filtros respectivamente, los cuales detectan bordes y texturas. Se intercalan capas de Max Pooling que reducen la dimensión manteniendo características distintivas. Luego se aplanan la salida y se conecta a dos capas totalmente conectadas de 256 y 2 neuronas, esta última con función softmax para estimar probabilidades de cada clase. En total hay 60 mil parámetros entrenables en esta CNN compacta.

Se entrenó durante 8 épocas con 100 pasos por época, probando optimizadores Adam, Adamax y Nadam. Adamax logró la mayor precisión de validación: 99.79%, indicando un modelo bien entrenado. VGG16 consiste en 5 bloques apilados de CNN con 16 capas, aplicando múltiples convoluciones seguidas de Max Pooling para comprimir la representación en cada etapa. Pero dado el pequeño tamaño de las imágenes de entrada, la dimensionalidad se reduce muy rápido, por lo que luego de 4 épocas su precisión sólo alcanzó 60%. Para la detección en nuevas imágenes, éstas se dividieron en parches de 60x60 píxeles que se clasificaron por separado. Desplazando la cuadrícula de parches se hicieron 25 pasadas en total. El modelo CNN detectó entre 37 y 41 viviendas, muy cercano a los 42 hogares contados manualmente.

En términos de aportes, la investigación evidencia la efectividad de CNNs poco profundas para extraer características distintivas de objetos simples como techos de vivienda en imágenes de baja resolución. Además, destaca la importancia de tunear apropiadamente los hiperparámetros según la complejidad de las imágenes.

Sin embargo, esta investigación tiene un enfoque más específico en la detección de techos. Como ya lo hemos mencionado antes, nuestra investigación apunta a identificar parques y canchas deportivas, para lo cual se requirió plantear una CNN optimizada para estas nuevas clases. Aun así, consideramos que este trabajo se acerca bastante a lo que queremos implementar y fue utilizada a modo de guía útil para nuestra investigación aplicada a la detección de áreas recreativas. Podríamos considerarlo como punto de partida para aplicar transfer learning y desarrollar un este modelo especializado en detectar canchas deportivas y parques a partir de imágenes satelitales para mapeo urbano.

4. IMPLEMENTACION DEL MODELO CLASIFICACIÓN DE REDES NEURONALES CONVOLUCIONALES E IMPLEMENTACIÓN SORE ESPACIOS PÚBLICOS RECREATIVOS EN SUELO URBANO

4.1. Objetivo específico 1

Diseño y Entrenamiento de Redes Neuronales Convolucionales

En primera instancia se avanzó específicamente en el objetivo específico 1, en este se llevó a cabo la etapa de entrenamiento del modelo y definición de la arquitectura, se procedieron a cargar las imágenes incluidos en carpetas con su respectiva etiqueta, estas imágenes que en total fueron 454 para iniciar la prueba de entrenamiento contenían datos de diferentes etiquetas que en este caso fueron codificadas para poder ser leídas por el modelo, estas fueron codificadas de la siguiente manera:

0: Diamante de baseball (Baseballdiamond)

1: Campo de football (footballfield)

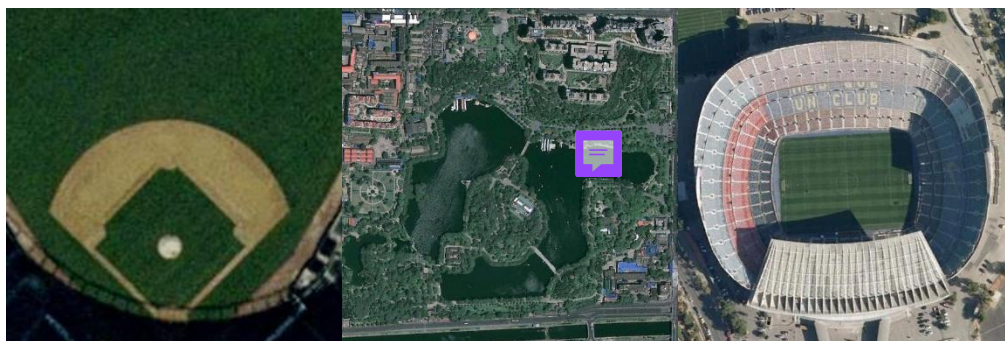
2: Campo de golf (Golfcourse)

3: Parque en la ciudad (Park)

4: Estanques (Pond)

5: Cancha de tennis (tenniscourt)

A continuación, se adjunta un ejemplo de cada clase a predecir de las etiquetas e imágenes cargadas:





Se detectó inicialmente que las imágenes iniciales tenían diferentes tamaños, razón por la cual al incluirlas en cualquier modelo este resultaría en error, por lo cual se redimensionan todas las imágenes, los tamaños encontrados fueron los siguientes:

Diferentes tamaños de imágenes encontrados:

(256, 255)

(256, 252)

(255, 256)

(251, 256)

(253, 256)

(600, 600)

(256, 247)

(256, 250)

(256, 256)

Por esta razón se procedió a redimensionar todas las imágenes a 256 X 256 (ancho por alto). Adicionalmente mediante el uso de Numpy se transformaron las imágenes en arreglos que describían cada uno de los píxeles dentro de la imagen, esto convirtió cada imagen en un vector que permitía ver en 3 canales de color (RGB) con una dimensión de 256 x 256 x 3 canales.

Posteriormente en función de procesar los datos y encontrar la mejor predicción posible para el entrenamiento de prueba preliminar a tener todas las imágenes, se normalizaron los datos de cada imagen en arreglos dividiendo cada arreglo en 255 que es el máximo valor RGB y permitiendo que cada vector se situó con una media igual a 0, esto con el fin de minimizar el ruido que se podría obtener al tener imágenes con valores muy diferentes.

Arquitectura Customizada

Una vez realizado este proceso de transformación, redimensionamiento, normalización, codificación de las etiquetas se procedió a realizar la división en datos de prueba y entrenamiento (20% y 80% respectivamente). Utilizando Keras para organizar y moldear la arquitectura del modelo con que se entrenará, en este caso un modelo de redes neuronales convolucionales, se usaron los siguientes parámetros:

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_1 (Conv2D)	(None, 125, 125, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 64)	0
flatten (Flatten)	(None, 246016)	0
dense (Dense)	(None, 64)	15,745,088
dense_1 (Dense)	(None, 6)	390

Total params: 15,764,870 (60.14 MB)

Trainable params: 15,764,870 (60.14 MB)

Non-trainable params: 0 (0.00 B)

1. **Conv2D (Capa convolucional):**

- Esta capa realiza la convolución de la imagen de entrada con un conjunto de filtros, que son matrices de pesos aprendidos durante el entrenamiento.
- El primer parámetro especifica el número de filtros que se aplicarán en paralelo. En este caso, hay 32 filtros.
- El segundo parámetro especifica el tamaño del filtro. Aquí se usa un tamaño de 3x3 píxeles.
- La función de activación ReLU (Rectified Linear Unit) se aplica a la salida de esta capa, lo que introduce no linealidad en la red.
- Esta capa tiene 896 parámetros entrenables, que son los pesos del filtro más un sesgo por cada filtro.

2. **MaxPooling2D (Capa de agrupación máxima):**

- Esta capa reduce la dimensionalidad espacial de la salida de la capa convolucional.
- Utiliza el método de agrupación máxima para tomar el valor máximo dentro de una región de la imagen.
- En este caso, se utiliza un tamaño de agrupación de 2x2 píxeles, lo que reduce la altura y la anchura de la imagen a la mitad.
- No tiene parámetros entrenables.

3. **Segunda Conv2D (Capa convolucional):**

- Similar a la primera capa convolucional, esta capa aplica otro conjunto de 64 filtros a la salida de la primera capa de agrupación.
- También utiliza un filtro de tamaño 3x3 y la función de activación ReLU.
- Tiene 18,496 parámetros entrenables.

4. **Segunda MaxPooling2D (Capa de agrupación máxima):**

- Al igual que la primera capa de agrupación, esta capa reduce aún más la dimensionalidad espacial de la salida de la segunda capa convolucional.
- También utiliza un tamaño de agrupación de 2x2 píxeles.
- No tiene parámetros entrenables.

5. **Flatten (Capa de aplanamiento):**

- Esta capa convierte la salida tridimensional de la última capa de agrupación en un vector unidimensional que puede ser utilizado por las capas totalmente conectadas.
- No tiene parámetros entrenables.

6. **Dense (Capa totalmente conectada):**


- Esta capa es una capa neuronal estándar donde cada neurona está conectada a todas las neuronas de la capa anterior.
- Se utiliza la función de activación ReLU.
- Tiene 15,745,088 parámetros entrenables.

7. **Capa de salida Dense:**


- Esta es la capa final de la red neuronal.
- Tiene 6 neuronas, una para cada clase en el problema de clasificación.
- Se utiliza la función de activación softmax para obtener la probabilidad de que la entrada pertenezca a cada clase.
- Tiene 390 parámetros entrenables.

Una vez creado el modelo secuencial con sus respectivas capas convolucionales y entrenamiento de la red neuronal se procedió a realizar el entrenamiento del modelo usando las diferentes épocas (epochs), en este caso 10, mediante las cuales la red neuronal entenderá como modificar los respectivos pesos de cada capa de la red neuronal con el fin de obtener los datos óptimos de accuracy y de los, en este caso un mayor accuracy y un menor valor de Loss, los resultados obtenidos en cada época fueron los siguientes:


Epoch 1/10

12/12  **8s** 554ms/step - accuracy: 0.2309 - loss: 2.6069 - val_accuracy: 0.4490 - val_loss: 1.4471


Epoch 2/10

12/12  **6s** 507ms/step - accuracy: 0.6285 - loss: 1.2499 - val_accuracy: 0.6832 - val_loss: 0.7761


Epoch 3/10

12/12  **6s** 509ms/step - accuracy: 0.7674 - loss: 0.6433 - val_accuracy: 0.8733 - val_loss: 0.3700


Epoch 4/10

12/12  **6s** 525ms/step - accuracy: 0.8563 - loss: 0.4159 - val_accuracy: 0.9256 - val_loss: 0.2624

Epoch 5/10

12/12  **6s** 503ms/step - accuracy: 0.9200 - loss: 0.2591 - val_accuracy: 0.9614 - val_loss: 0.1385

Epoch 6/10

12/12  **6s** 508ms/step - accuracy: 0.9801 - loss: 0.1080 - val_accuracy: 0.9862 - val_loss: 0.0666

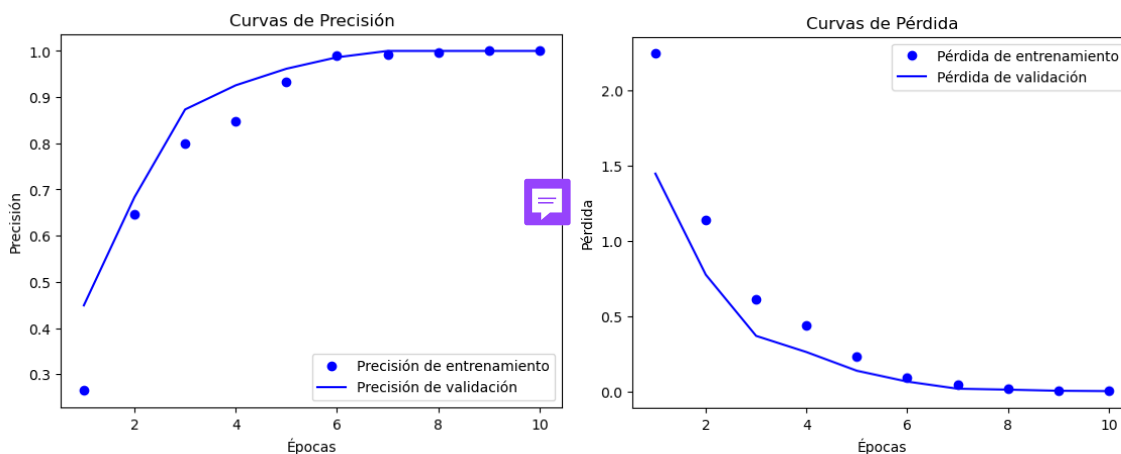
Epoch 7/10

```

12/12 ----- 6s 498ms/step - accuracy:
0.9860 - loss: 0.0569 - val_accuracy: 1.0000 - val_loss: 0.0194
Epoch 8/10
12/12 ----- 6s 498ms/step - accuracy:
0.9967 - loss: 0.0200 - val_accuracy: 1.0000 - val_loss: 0.0131
Epoch 9/10
12/12 ----- 6s 507ms/step - accuracy:
1.0000 - loss: 0.0096 - val_accuracy: 1.0000 - val_loss: 0.0057
Epoch 10/10
12/12 ----- 6s 514ms/step - accuracy:
1.0000 - loss: 0.0060 - val_accuracy: 1.0000 - val_loss: 0.0034

```

En este se puede observar cómo al final del experimento de obtuvo un accuracy o precisión de 1, lo que puede indicar que el modelo se sobre ajustó a los datos y es importante validar si es el caso, sin embargo, al probarlo con los datos de prueba, datos que no había visto nunca el modelo, se obtuvo un valor de 0.835 de accuracy, lo que indica que las predicciones respecto a cada clase son muy buenas, A continuación se puede ver en las gráficas como a medida que se fue entrenando el modelo se optimizaron las métricas sugeridas:



Al ser la primera prueba de inicio con el fin de comprobar los pasos y parámetros a usar dentro del modelo, y al obtener una buena precisión en primera instancia, se procedió a usar el modelo y probarse usando datos que nunca ha visto, particularmente se evaluó con una imagen de una cancha de tenis, a continuación, se expone la imagen:



Esta imagen etiquetada como Cancha de tenis, se procedió a incluirla en el modelo entrenado como un parámetro de entrada y se le hicieron los mismos procesos de transformación para que fuera ingresado al modelo y obtener una predicción. En este caso se obtuvo un valor de 5, lo cual es correcto.

En el proceso de creación y revisión de la etapa anteriormente descrita, se pudo observar que se trabajan con tipologías de datos muy similares a las que se obtienen en países de la región anglosajona o del continente Europeo, en ese caso, es importante destacar que en Colombia se manejan tipologías de espacios recreativos muy diferentes a las mostradas y utilizadas para entrenar este modelos, por esta razón se contempla la idea de crear un dataset etiquetado basado en imágenes de ciudades colombianas que permitan obtener un modelo más preciso y ajustado al contexto del país donde se aplicarán los resultados del entrenamiento.

Arquitectura VGG16

En cuanto al desarrollo del modelo utilizando la arquitectura preentrenada VGG16, esta red neuronal convolucional profunda fue creada por el grupo de investigación Visual Geometry Group (VGG) de la Universidad de Oxford. Lanzada en 2014, es conocida por su sencillez y efectividad, empleando únicamente capas convolucionales y de pooling en su diseño. VGG16 consta de 16 capas con pesos entrenables, de las cuales 13 son convolucionales y 3 están completamente conectadas. En este caso, se decidió congelar las capas del modelo base para acelerar el entrenamiento. El preentrenamiento se realizó utilizando el conjunto de datos "Imagenet", que conserva las características mencionadas. Las capas superiores de la red neuronal se dejaron libres para adaptarse al problema específico, que en este caso tiene 6 etiquetas de salida y, por lo tanto, 6 neuronas de salida.

Una vez establecida la arquitectura mediante el uso de la librería keras, se procedió a realizar el entrenamiento tratando de maximizar el accuracy y de minimizar la pérdida. en este caso se entrenó con 10 épocas al igual que el modelo anterior consiguiendo los siguientes resultados:

```
Epoch 1/10
16/16 ██████████ 40s 2s/step - accuracy: 0.3889
- loss: 2.6943 - val_accuracy: 0.8571 - val_loss: 0.5659
Epoch 2/10
```



```

16/16 ----- 42s 3s/step - accuracy: 0.7993
- loss: 0.5179 - val_accuracy: 0.8901 - val_loss: 0.4478
Epoch 3/10
16/16 ----- 46s 3s/step - accuracy: 0.8867
- loss: 0.3108 - val_accuracy: 0.8901 - val_loss: 0.2730
Epoch 4/10
16/16 ----- 49s 3s/step - accuracy: 0.9185
- loss: 0.2166 - val_accuracy: 0.9121 - val_loss: 0.2592
Epoch 5/10
16/16 ----- 50s 3s/step - accuracy: 0.9218
- loss: 0.2011 - val_accuracy: 0.9011 - val_loss: 0.2853
Epoch 6/10
16/16 ----- 50s 3s/step - accuracy: 0.9360
- loss: 0.1724 - val_accuracy: 0.9341 - val_loss: 0.2206
Epoch 7/10
16/16 ----- 49s 3s/step - accuracy: 0.9551
- loss: 0.1465 - val_accuracy: 0.9011 - val_loss: 0.3062
Epoch 8/10
16/16 ----- 49s 3s/step - accuracy: 0.9667
- loss: 0.1012 - val_accuracy: 0.9121 - val_loss: 0.2468
Epoch 9/10
16/16 ----- 49s 3s/step - accuracy: 0.9695
- loss: 0.0867 - val_accuracy: 0.9121 - val_loss: 0.2773
Epoch 10/10
16/16 ----- 51s 3s/step - accuracy: 0.9757
- loss: 0.0692 - val_accuracy: 0.9231 - val_loss: 0.2555

```

en comparación con el modelo anterior consiguió una mejora representativa de al menos 0.9 puntos en el nivel de predictibilidad, sin embargo, para obtener datos más precisos y con el fin de poder realizar la mejor comparativa entre ambas arquitecturas se realizó el siguiente cuadro donde se presenta el rendimiento de los modelos con diferentes métricas representativas:

Arquitectura	Pre-Trained	Accuracy	Precision	Recall	F1-Score
Custom	FALSO	0.83510	0.84050	0.83000	0.83600
VGG16	VERDADERO	0.92300	0.92960	0.92300	0.92360

Como se puede apreciar, aunque ambos resultados son buenos y representativos sobre el conjunto de prueba, la arquitectura pre entrenada de VGG16 presenta mejores resultados en todas las métricas, esto no implica que durante el desarrollo del proyecto nos decantemos por utilizar únicamente esta arquitectura, se seguirán probando diferentes arquitecturas como las Resnet18 que permitan obtener nuevas panorámicas de entrenamiento.

Bases de datos adicionales

Como parte del proceso de investigación, se realizó la consulta en diferentes bases de datos estudiadas y

sugeridas en artículos científicos, estas bases de datos han sido utilizadas para procesos de entrenamiento y vienen debidamente etiquetadas con el fin de ser usadas para modelos predictivos, cada una de estas bases contenía diferentes clases y número de imágenes, lo cual puede resultar beneficioso ya que podría servir para investigar de qué manera funcionan los modelos y arquitecturas con diferentes cantidades de imágenes, y adicionalmente, se pueden apreciar otras categorías como canchas de Basketball y GroundTrack field, esto es particularmente útil en la investigación debido a que lo que se busca es realizar un modelo de clasificación de suelo urbano recreativo, y en ese sentido en las ciudades colombianas es fácil encontrar canchas de Basketball en los espacio recreativos de las ciudad. A continuación, una muestra de las dos nuevas categorías:



Haciendo una descripción más precisa de las nuevas bases de datos, inicialmente se consiguió la base de datos etiquetada PatternNET con un total de 4,000 imágenes y 5 categorías que se usarán en el modelo todas sus imágenes, tanto de esta base como de las siguientes son del tamaño 256 x 256 píxeles de ancho y alto. La siguiente base recibe el nombre de Optimal 31 y esta cuenta con 240 imágenes y 4 categorías, finalmente, se obtuvo la base de datos MLRSNet que contiene un total de 16,556 imágenes y 7 categorías. A continuación, se presenta un resumen de la información encontrada y la cantidad de imágenes por categoría:

PatternNET	Optimal_31	MLRSNet
BaseballField: 800 Basketball Court: 800 Football Field: 800 Golf Course: 800 Tennis Court: 800	Baseballdiamond:60 Basketball Court:60 Golfcourse: 60 GraudTrack_Field:60	BaseballField: 2002 Basketball Court: 2895 Park: 1682 Stadium: 2462 Golf Course: 2515 GrondTrack Field: 2500 Tennis Court: 2500

Es importante resaltar, que los procedimientos anteriores se realizan para cada una de estas bases de datos, con el fin de hacer pruebas empíricas que permitan encontrar bajo qué condiciones y tipos de imágenes se consigue comportar mejor el modelo y así encontrar los mejores resultados dentro de las métricas entrenadas, esto hace parte de las siguientes actividades y avances a planear con el fin de desarrollar el proyecto acá expuesto.

4.2. Objetivo específico 2

Desarrollo de modelo de Machine Learning para evaluación de efectividad vs Deep Learning:

Se desarrolló un modelo en base a keras para el procesamiento de imágenes satelitales, apoyado con librerías de OS, matplotlib y PIL para su procesamiento y sklearn para el uso de los métodos ML.

Se cargaron tres conjuntos de imágenes de prueba correspondientes a tres categorías de espacios urbanos, cada uno con 100 imágenes. Estas se normalizaron y se redimensionaron para estandarizar la decodificación y optimizar el entrenamiento. Se subdivieron en 240 imágenes de entrenamiento y 60 de prueba.

Los resultados obtenidos en cuanto a Accuracy para cada uno de los métodos son:

- Random Forest: 0.583
- SVM: 0.584
- K-vecinos: 0.411

Adicionalmente, se ejecutaron pruebas que demostraron imprecisiones, las cuales asociamos a la falta de conjuntos de datos más “reales” a los escenarios urbanos del territorio nacional, por lo cual hemos decidido armar un plan para construir un conjunto de datos de imágenes satelitales del territorio colombiano, el cual estaremos desarrollando de manera progresiva.

Estos son los resultados de las pruebas de predicción de imágenes utilizando los modelos tradicionales de Machine Learning:

Imagen	Random Forest	SVM	K-Vecinos
Campo de Beisbol - Dota 1.5	baseballdiamond	baseballdiamond	baseballdiamond
Cancha de Futbol	tenniscourt	tenniscourt	golfcourse
Cancha de Tenis	tenniscourt	tenniscourt	baseballdiamond
Campo de futbol y canchas multiples	tenniscourt	baseballdiamond	baseballdiamond

3.3 Prueba

En esta prueba, utilizaremos una imagen de un campo de béisbol del mismo dataset de imágenes

```
img_path = "G:/Py Drive/Universidad/tesis/tesis/imagenes/campo de beisbol.tif" # Cambia esto a la ruta de la imagen nueva
predicir_imgen(img_path, rf_model, svm_model, knn_model, decodificacion)
```

Predicción del modelo Random Forest: baseballfield
 Predicción del modelo SVM: baseballfield
 Predicción del modelo K-Vecinos: baseballfield

En la siguiente prueba, utilizaremos una imagen de google maps de una cancha de futbol

```
img_path = "G:/Py Drive/Universidad/tesis/tesis/imagenes/imga_test_3.jpg" # Cambia esto a la ruta de la imagen nueva
predicir_imgen(img_path, rf_model, svm_model, knn_model, decodificacion)
```

Predicción del modelo Random Forest: tenniscourt
 Predicción del modelo SVM: tenniscourt
 Predicción del modelo K-Vecinos: golfcourse

En la siguiente prueba, utilizaremos una imagen de un conjunto de canchas de tenis

```
img_path = "G:/Py Drive/Universidad/tesis/tesis/imagenes/imga_test_2.jpg" # Cambia esto a la ruta de la imagen nueva
predicir_imgen(img_path, rf_model, svm_model, knn_model, decodificacion)
```

Predicción del modelo Random Forest: tenniscourt
 Predicción del modelo SVM: tenniscourt
 Predicción del modelo K-Vecinos: baseballfield

En la siguiente, probaremos una imagen que contiene un campo grande que parece ser un campo de futbol y un conjunto de canchas múltiples.

```
img_path = "G:/Py Drive/Universidad/tesis/tesis/imagenes/test_3.tif" # Cambia esto a la ruta de la imagen nueva
predicir_imgen(img_path, rf_model, svm_model, knn_model, decodificacion)
```

Predicción del modelo Random Forest: tenniscourt
 Predicción del modelo SVM: baseballfield
 Predicción del modelo K-Vecinos: baseballfield

Encuentramos entonces que el modelo de K-Vecinos no es tan preciso como se espera para los resultados esperados.

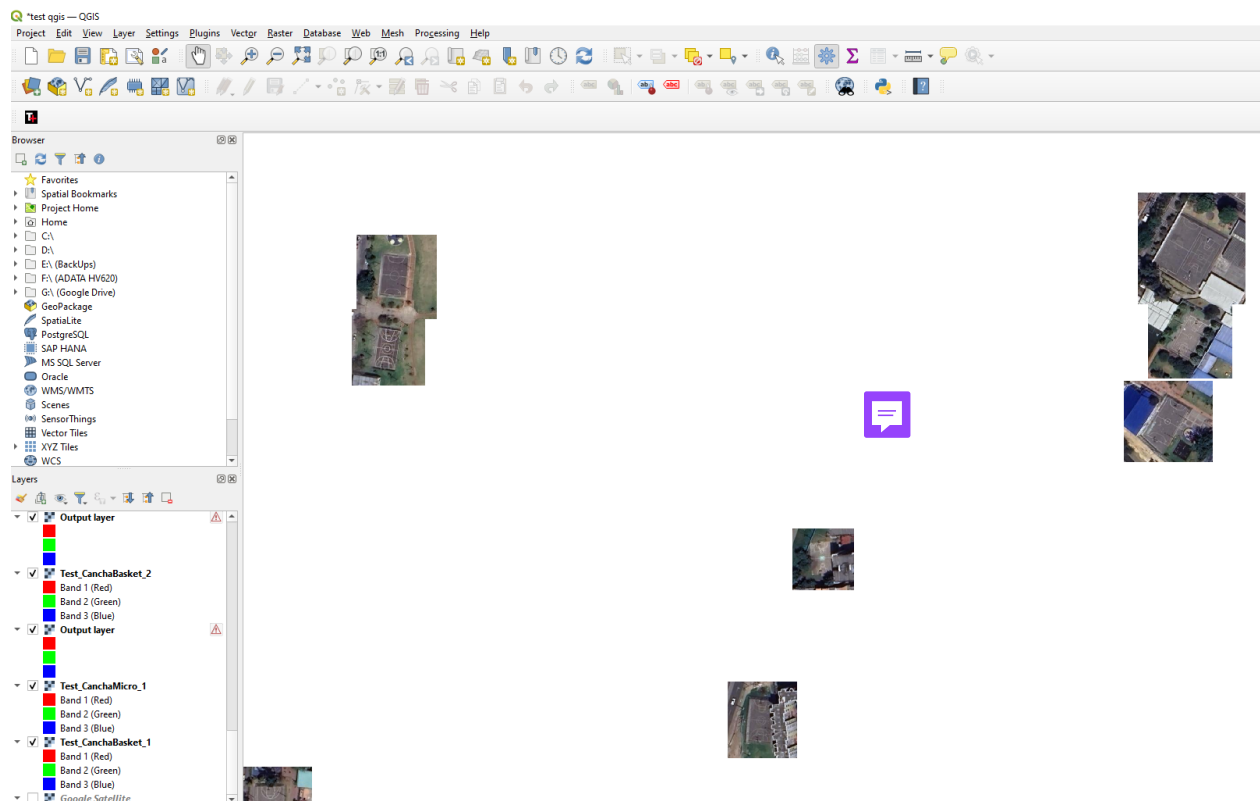
El nivel de precisión de resultados no es el esperado, ya que en algunos casos se confunden los tipos de escenarios, por ejemplo, el campo de futbol con el campo de tenis. Sin embargo, de los modelos tradicionales de Machine Learning, Random Forest y SVM son los que cuentan con mejor tipo de resultados. Podríamos descartar a K-Venicos como un modelo tradicional que se adapte a los requerimientos del análisis de imágenes satelitales.

Conjunto de Imágenes Nacionales

Se realizó una investigación para obtener un método de etiquetado de imágenes que sirviera para aportar nuestro proyecto, sin embargo, nos fue imposible encontrar en conjuntos de datos abiertos algunos datasets con imágenes de espacios urbanos recreativos del territorio nacional. Entonces, decidimos buscar herramientas que nos permitirán y facilitarán la construcción de un conjunto de datos.

Probamos con un proyecto llamado Iris, el cual utiliza métodos de deep learning para el etiquetado de imágenes. Sin embargo, su proceso se hace mediante la segmentación de imágenes, por lo cual no es útil para nuestro proceso. Buscamos más proyectos que pudieran ser útiles, sin embargo, algunos no estaban disponibles de manera pública.

Recurrimos a un método más operativo, pero efectivo, y es el de la obtención de imágenes raster a partir del software QGIS, el cual nos permite recortar de una manera más precisa y al tamaño que requerimos imágenes obtenidas a partir de diferentes tecnologías satelitales, como Google Maps o ESRI. Dedicamos un tiempo de nuestro proyecto a esta actividad para poder generar un conjunto de datos más “real” y acorde a lo que requerimos, con imágenes de escenarios deportivos y recreativos implementados en las ciudades colombianas.



Decidimos construir un dataset con 8 clases de imágenes satelitales correspondientes a los siguientes escenarios deportivos que comúnmente se encuentran disponibles en el territorio nacional:

- Cancha de Basquetbol
- Campo de Beisbol
- Campo de Futbol
- Cancha de Microfutbol
- Cancha Múltiple
- Pista de Patinaje
- Cancha de Tenis
- Parque barrial / zonal

Cada una de estas clases fue alimentada con al menos 100 imágenes originales descargadas del software QGIS. Adicionalmente, hemos utilizado *Image Data Generator de Keras* para aumentar el grupo de imágenes a 150 por clase para los procesos de entrenamiento y testeo.

La calidad de imágenes satelitales del territorio colombiano dista mucho de tener calidades comparables con los datasets públicos que encontramos para diseñar nuestro modelo, los cuales estaban basados en imágenes satelitales del territorio europeo. Sin embargo, consideramos que era pertinente utilizar este tipo de imágenes de baja calidad puesto que la idea es que nuestro modelo pueda ser útil con este tipo de imágenes (de Google Maps) las que prácticamente cualquier entidad podría estar en capacidad de obtener mediante capturas de pantalla u otros métodos y aplicar nuestro modelo para hacer sus respectivos

análisis.

4.3. Objetivo específico 3

5. CONCLUSIONES Y TRABAJOS FUTUROS

5.1. CONCLUSIONES

5.2. TRABAJOS FUTUROS

6. REFERENCIAS BIBLIOGRÁFICAS

- [1] J. E. Jiménez Caldera and G. Y. Durango Severiche, "Diagnóstico y planificación del espacio público urbano. La participación de los ciudadanos usuario," Universidad Nacional de Colombia, 2020. [Online]. Available: <https://www.redalyc.org/journal/748/74866570019/html/#fn1EWKL>
- [2] J. Ding, N. Xue, G.-S. Xia, X. Bai, W. Yang, M. Y. Yang, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges," 2021.
- [3] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, and Liangpei Zhang, "AID: A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification," 2016.
- [4] J. A. Valero Medina and B. E. Alzate Atehortúa, "Comparison of maximum likelihood, support vector machines, and random forest techniques in satellite images classification," *Tecnura*, vol. 23, no. 59, pp. 13–26, Jan. 2019.
- [5] Ingeoexpert, "Clasificaciones de imágenes de satélite", Nov 13, 2020, <https://ingeoexpert.com/articulo/clasificaciones-de-imagenes-de-satelite/>
- [6] M. Rodríguez and M. Moctezuma, "Análisis Bayesiano y Fusión de Datos Para La Clasificación de Escenas Urbanas del Distrito Federal," *Ingeniería, Investigación y Tecnología VII*, México D.F., vol. 1, pp. 17-28, 2006.
- [7] X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, & F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8-36, 2017. <https://ieeexplore.ieee.org/document/8113128>
- [8] E. Karypidis, S. G. Mouslech, K. Skoulariki, and A. Gazis, "Comparison Analysis of Traditional Machine Learning and Deep Learning Techniques for Data and Image Classification," 2022. <https://doi.org/10.48550/arXiv.2204.05983>
- [9] S. K. Agrawal, "Metrics to Evaluate your Classification Model to take the right decisions," Sept 29, 2023. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/>

- [10] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, D. Tiede, & J. Aryal, "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sensing*, vol. 11, no. 2, p. 196, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/2/196>
- [11] A. Asokan and J. Anitha, "Machine Learning based Image Processing Techniques for Satellite Image Analysis -A Survey," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 119-124.
- [12] L. Ma et al., "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019. <https://www.sciencedirect.com/science/article/pii/S0924271619301108?via%3Dihub#ab005>
- [13] L. Zhang et al., "Deep Learning for Remote Sensing Image Classification: A Survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 6, e1264, 2018. <https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1264>
- [14] A. Javed, T. Kim, C. Lee, J. Oh, and Y. Han, "Deep Learning-Based Detection of Urban Forest Cover Change along with Overall Urban Changes Using Very-High-Resolution Satellite Images," *Remote Sens.*, vol. 15, p. 4285, 2023. <https://doi.org/10.3390/rs15174285>
- [15] L. Valdés Ávila and J. Baquero Vanegas, "Deep Learning aplicado a imágenes satelitales como herramienta de detección de viviendas sin servicio de energía en el caserío Media Luna – Uribia – Guajira." Jul 16, 2019.

Se deben presentar de forma rigurosa y completa las referencias bibliográficas utilizadas en el documento (No incluir bibliografía que no haya sido referenciada en el documento). Se debe seguir una sola norma de referenciación en todo el documento, que es la norma IEEE.

Utilizar en lo posible bibliografía reciente de fuentes confiables (libros, artículos científicos, etc.). Evitar utilizar fuentes no confiables como blogs, Wikipedia, o documentos sin autor.