

Santiago de Cali, 07 de Junio de 2024

Doctor:
Diego Luis Linares Ospina
Director Maestría en Ciencia de Datos
Facultad de Ingeniería y Ciencias
Pontificia Universidad Javeriana - Cali

Con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el Trabajo de Grado y posteriormente optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto de Trabajo de Grado denominado MODELO DE CLASIFICACIÓN DE TIPOLOGÍAS CONSTRUCTIVAS BASADO EN FOTOGRAFÍAS DE CAMPO Y EL MODELO LADM-COL, el cual será realizado por el (la) estudiante Jonathan Montoya Castellanos con código 8980505 perteneciente al énfasis en Sistemas y Computación, bajo la dirección del profesor Julián Gil Gonzales.

El suscrito director del Trabajo de Grado autoriza para que se proceda a hacer la evaluación de este Proyecto ante el Tribunal que para el efecto se designe, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado oficialmente.

Atentamente,



Firma
Jonathan Montoya Castellanos

C.C. 1073707557 de Soacha



Firma
Julián Gil González

C.C. 1088286439 de Pereira

ACTA DE ENTREGA

El Programa Nuestra Tierra Próspera, financiado por USAID, en adelante el PROGRAMA, hace entrega de:

- Archivos XTF del barrido predial masivo de Ataco, Cáceres (incluyendo corregimientos urbanos), San Jacinto, Fuente de Oro y Puerto Lleras.
- Repositorio de fotografías de construcciones de Ataco, Cáceres (incluyendo corregimientos urbanos), San Jacinto, Fuente de Oro y Puerto Lleras.

A Jonathan Montoya Castellanos, identificado con la cédula 1073707557, teniendo en cuenta las siguientes consideraciones:

1. Que la información compartida en virtud de la presente acta, en adelante **INFORMACIÓN CONFIDENCIAL**, pertenece a USAID y la misma es considerada sensible y de carácter restringido en su divulgación, manejo y utilización.
2. Que la **INFORMACIÓN CONFIDENCIAL** se encuentra bajo custodia del **PROGRAMA** y ha sido desarrollada en el marco de la ejecución de sus actividades.
3. Que la **INFORMACIÓN CONFIDENCIAL** se comparte en el marco del desarrollo del proyecto aplicado en Maestría en Ciencia de Datos de la Universidad Pontificia Javeriana de Cali, denominado, “MODELO DE CLASIFICACIÓN DE TIPOLOGÍAS CONSTRUCTIVAS BASADO EN FOTOGRAFÍAS DE CAMPO Y EL MODELO LADM-COL”, cuyo objeto exclusivo es “Desarrollar un modelo de clasificación de imágenes a partir de técnicas de aprendizaje automático con el fin de llevar a cabo tareas de tipificación de construcciones siguiendo el estándar del Modelo LADM-COL”.

En consecuencia, el estudiante Jonathan Montoya Castellanos, identificado con CC 1073707557, estudiante de la Maestría en Ciencia de datos de la Universidad Pontificia Javeriana de Cali, se compromete con las siguientes cláusulas para el uso de la información:

Primera. Objeto: La **INFORMACIÓN CONFIDENCIAL** será única y exclusivamente utilizada para desarrollar un modelo de clasificación de imágenes a partir de técnicas de aprendizaje automático con el fin de llevar a cabo tareas de tipificación de construcciones siguiendo el estándar del Modelo LADM-COL, en el marco del desarrollo del proyecto aplicado de la Maestría en Ciencia de Datos.

Segunda. Confidencialidad: En virtud de la presente acta el estudiante Jonathan Montoya Castellanos, se obliga a no divulgar directa, indirecta, próxima o remotamente, ni a través de ninguna otra persona o institución, la **INFORMACIÓN CONFIDENCIAL**, así como también a no utilizar dicha información con objetivos distintos a lo establecido en la primera cláusula de este documento.

Tercera. Obligaciones:

1. Mantener la **INFORMACIÓN CONFIDENCIAL** segura y usarla solamente para los propósitos establecidos. En caso de ser solicitada por el Programa o que ya no requieran la información, devolverla toda al Programa incluyendo cualquier tipo de copias que pudieran existir.
2. Proteger la **INFORMACIÓN CONFIDENCIAL**, sea verbal, escrita, visual, tangible, con acceso remoto, intangible o que por cualquier otro medio reciba, siendo legítimo poseedor de esta USAID, restringiendo su uso exclusivamente a las personas delegadas para tal fin.
3. Usar la **INFORMACIÓN CONFIDENCIAL** que se le entregue, únicamente para los efectos señalados en el objeto de la presente acta.
4. Mantener la **INFORMACIÓN CONFIDENCIAL** en reserva hasta tanto adquiera el carácter de pública si esto ocurre.
5. Responder por el mal uso que le den sus representantes a la **INFORMACIÓN CONFIDENCIAL**.
6. No transmitir, comunicar, revelar o de cualquier otra forma divulgar total o parcialmente, pública o privadamente, la **INFORMACIÓN CONFIDENCIAL** sin el previo consentimiento por escrito por parte del **PROGRAMA**.
7. Compartir la versión final de su tesis y material de apoyo con el PROGRAMA, así como realizar una presentación del Modelo de Clasificación de Tipologías.

Cuarta. Forma de entrega: La **INFORMACIÓN CONFIDENCIAL** se entrega en el siguiente enlace:

https://tetrattechinc-my.sharepoint.com/:f/g/personal/j_montoyacastellano_tetrattech_com/EvF1N805UNpCgVoqvufXmhbB5DeUe9jlsQf0WoDBt3499Q?e=iQLIOF

para que Jonathan Montoya Castellanos pueda descargarlas en un plazo de 30 días.

Quinta. Divulgación. Cualquier divulgación autorizada de la **INFORMACIÓN CONFIDENCIAL** a terceras personas estará sujeta a las mismas obligaciones de confidencialidad derivadas de la presente acta y Jonathan Montoya Castellanos deberá informar estas restricciones incluyendo la identificación de la información como confidencial.

Sexta. Exclusiones a la confidencialidad: Jonathan Montoya Castellanos en ningún caso queda eximido de la obligación de confidencialidad.

Séptima. Legislación aplicable: Esta acta por las leyes de la República de Colombia y se interpretará de acuerdo con las mismas.

Octava. Reconocimiento: Teniendo en cuenta que la **INFORMACIÓN CONFIDENCIAL** es propiedad de **USAID**, se deberá reconocer el apoyo de USAID en la elaboración de cualquier tipo de insumo o producto que resulten de su uso.

Novena. Incumplimiento: En caso de que Jonathan Montoya Castellanos incumpla parcial o totalmente con las obligaciones establecidas en la presente acta éste será responsable de los daños y perjuicios que dicho incumplimiento llegase a ocasionar al Programa y/o USAID.



USAID
DEL PUEBLO DE LOS ESTADOS
UNIDOS DE AMÉRICA

Programa
Nuestra Tierra
Próspera

Décima. Aceptación del acta de entrega: Jonathan Montoya Castellanos ha leído y estudiado de manera detenida los términos y el contenido de la presente acta y por tanto manifiestan estar conforme y acepta todas las condiciones.

Se entrega en Bogotá D.C., a los 16 días del mes de agosto de 2023.

ENTREGA

03162B9992B
F43D9A567B
E5689B741DF

Digitally signed by
03162B9992BF43D9
A567BE5689B741DF
Date: 2023.08.28
17:26:41 -05'00'

SERGIO ZEGARRA
Subdirector de operaciones
Programa Nuestra Tierra Próspera

RECIBE

Jonathan Montoya Castellanos
Estudiante de la UPJC

CC 1073707557



**Maestría en Ciencia de Datos
Facultad de Ingeniería y Ciencias**

**FICHA RESUMEN
TRABAJO DE GRADO DE MAESTRÍA**

TITULO: “MODELO DE CLASIFICACIÓN DE TIPOLOGÍAS CONSTRUCTIVAS BASADO EN FOTOGRAFÍAS DE CAMPO Y EL MODELO LADM-COL”

1. ÉNFASIS: Sistemas y Computación
2. TIPO DE PROYECTO: Aplicado
3. ÁREA DE TRABAJO: Catastro, Modelo LADM-COL, Ciencia de Datos
4. ESTUDIANTE (S): Jonathan Montoya Castellanos
5. CORREO ELECTRÓNICO: jmontoyac31ud@javerianacali.edu.co
6. DIRECCIÓN Y TELÉFONO: Carrera 3 30 130 Soacha, 3185277483
7. DIRECTOR: Dr. Julián Gil González
8. VINCULACIÓN DEL DIRECTOR (en la universidad): Cátedra
9. CORREO ELECTRÓNICO DEL DIRECTOR: julian.gil@javerianacali.edu.co
10. CO-DIRECTOR(ES) (Si aplica):
11. GRUPO O EMPRESA QUE LO AVALA (Si aplica):
12. OTROS GRUPOS O EMPRESAS: Tetra Tech ARD, Programa Nuestra Tierra Prospera
13. PALABRAS CLAVE (al menos 5): Clasificación de Imágenes, Catastro Colombiano, LADM-COL, Calificación de Construcciones, Reconocimiento predial, evaluación de calidad
14. ODS QUE APLICA EL PROYECTO (Agenda 2030): ODS 9: Industria, innovación e infraestructura
15. FECHA DE INICIO (Desarrollo del proyecto): 11/06/2023

16. RESUMEN (máximo 400 palabras).

El proyecto planteado, “Modelo de clasificación de tipologías constructivas basado en fotografías de campo y el modelo LADM-COL”, adquiere relevancia en el ámbito de implementación de procesos catastrales, barridos prediales masivos y evaluación de calidad de productos catastrales. Esto se debe a las necesidades técnicas que enfrenta el Catastro Multipropósito en el país.

La captura y almacenamiento de fotografías de construcciones es común en los procesos catastrales como soporte de campo a la información que reposa en las bases de datos. La cantidad de construcciones almacenadas en las bases de datos de un municipio representan un volumen importante de información que debe mostrar “completitud” y “coherencia” entre los soportes de campo y lo registrado en las bases de datos. Esta coherencia y completitud suele construirse y evaluarse de manera puntual, con visitas en campo o soportes fotográficos para determinar la calificación de construcciones asignada en las bases de datos, este proceso debe considerar profesionales especializados, tiempo y logística que se traducen en costos para las instituciones.

Dado lo anterior, se identifican los siguientes problemas: el tiempo de ejecución de los procesos, la disponibilidad de personal capacitado y la calidad de los productos generados por dichos profesionales, entre otros aspectos de los procesos catastrales que dificultan el cumplimiento de los objetivos y metas del catastro multipropósito.

El objetivo propuesto para este proyecto fue desarrollar un modelo de clasificación de imágenes utilizando técnicas de aprendizaje automático, con el fin de llevar a cabo tareas de tipificación de construcciones siguiendo el estándar del Modelo LADM-COL.

El resultado esperado para este proyecto fue el obtener un modelo de clasificación de imágenes que brinde un apoyo significativo al proceso de reconocimiento predial, al clasificar las construcciones de manera estandarizada y automatizada.



Pontificia Universidad
JAVERIANA
Cali

**MODELO DE CLASIFICACIÓN DE TIPOLOGÍAS CONSTRUCTIVAS BASADO
EN FOTOGRAFÍAS DE CAMPO Y EL MODELO LADM-COL**

Jonathan Montoya Castellanos
Código 898050

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

Director(a)
Dr. Julián Gil Gonzalez

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, 7 DE JUNIO DE 2024

TABLA DE CONTENIDO

INTRODUCCIÓN	5
1. DEFINICIÓN DEL PROBLEMA	6
1.1 PLANTEAMIENTO DEL PROBLEMA	6
1.2 FORMULACIÓN DEL PROBLEMA	7
2. OBJETIVOS DEL PROYECTO	8
2.1 OBJETIVO GENERAL	8
2.2 OBJETIVOS ESPECÍFICOS	8
3. MARCO DE REFERENCIA	8
3.1 MARCO TEÓRICO	8
3.2 ANTECEDENTES	12
4. DESARROLLO DEL PROYECTO	19
4.1 METODOLOGÍA	19
4.2 RESULTADOS	39
5. CONCLUSIONES Y TRABAJOS FUTUROS	50
5.1 CONCLUSIONES	50
5.2 TRABAJOS FUTUROS	50
6. ANEXO	51
7. REFERENCIAS BIBLIOGRÁFICAS	65

LISTA DE FIGURAS

Ilustración 1. Representación Red Neuronal Convolucional (CNN) Fuente: García et al [7]	12
Ilustración 2. Sistema Difuso Metodología Fuente: Cortés [15, p. 62].....	15
Ilustración 3. Parámetros del Sistema Difuso Fuente: Cortés [15]	16
Ilustración 4. Diagrama de flujo metodológico Fuente: Venegas [15]	17
Ilustración 5. Modelo UML de Construcciones LADM-COL 1.0.....	19
Ilustración 6. Archivos XTF y Back up recopilados.....	20
Ilustración 7. Repositorios de fotografías	20
Ilustración 8. Análisis exploratorio del consolidado XTF	23
Ilustración 9.1. Fotografía 150 x 150	28
Ilustración 9.2. Fotografía 250 x 250.....	
Ilustración 10. Fotografías de Fachada Etiquetadas.....	29
Ilustración 11. Fotografías con error en etiqueta	30
Ilustración 12. Modelo Base, clasificador de fachadas.....	31
Ilustración 13. Modelo Convolucional clasificador de fachadas	31
Ilustración 14. Estrategia de recolección y procesamiento	32
Ilustración 15. Ejemplos de fotografías con etiqueta incorrecta	34
Ilustración 16. Modelo Base, clasificador de tipologías	36
Ilustración 17. Modelo convolucional, clasificador de tipologías Estructura 1	36
Ilustración 18. Modelo convolucional, clasificador de tipologías Estructura 2	36
Ilustración 19. Modelo convolucional, clasificador de tipologías Estructura 3	37
Ilustración 20. Modelo convolucional, clasificador de tipologías Estructura 4	37
Ilustración 21. Modelo convolucional, clasificador de tipologías Estructura 5	38
Ilustración 22. Modelo convolucional, clasificador de tipologías Estructura 6	38
Ilustración 23. Métricas de desempeño del modelo.....	39
Ilustración 24. Matriz de confusión del modelo	39
Ilustración 25. Métricas de desempeño globales modelo 8 y modelo 9.....	41
Ilustración 26. Matriz de confusión global modelo 8 y modelo 9	41
Ilustración 27. Modelo convolucional, estructura de pruebas.....	42
Ilustración 28. Métricas de desempeño, modelo de prueba 1	42
Ilustración 29. Matriz de confusión, modelo de prueba 1	42
Ilustración 30. Métricas de desempeño, modelo de prueba 2	43
Ilustración 31. Matriz de confusión, modelo de prueba 2.....	43
Ilustración 32. Métricas de desempeño, modelo de prueba 3	43
Ilustración 33. Matriz de confusión, modelo de prueba 3.....	44
Ilustración 34. Casos con difícil discriminación de tipología	44
Ilustración 35. Métricas de desempeño modelo SVM1 y modelo SVM2.....	45
Ilustración 36. Matriz de confusión modelo SVM1 y modelo SVM2	45
Ilustración 37. Métricas de desempeño modelo SVM3 y modelo SVM4.....	46
Ilustración 38. Matriz de confusión modelo SVM3 y modelo SVM4	46
Ilustración 39. Tipología 1 Modelo funcional 8	47
Ilustración 40. Tipología 2 Modelo funcional 8	48
Ilustración 41. Tipología 3 Modelo funcional 8	48
Ilustración 42. Errores de clasificación en tipología 1	49
Ilustración 43. Errores de clasificación en tipología 2.....	49
Ilustración 44. Errores de clasificación en tipología 3.....	49
Ilustración 45. Métricas de desempeño modelo 1	51
Ilustración 46. Matriz de confusión modelo 1	51
Ilustración 47. Curvas de pérdida y precisión modelo 1	51
Ilustración 48. Métricas de desempeño modelo 2.....	52
Ilustración 49. Matriz de confusión modelo 2	52
Ilustración 50. Curvas de pérdida y precisión modelo 2.....	52
Ilustración 51. Métricas de desempeño modelo 3.....	53
Ilustración 52. Matriz de confusión modelo 3	53

Ilustración 53. Curvas de pérdida y precisión modelo 3.....	53
Ilustración 54. Métricas de desempeño modelo 4.....	54
Ilustración 55. Matriz de confusión modelo 4.....	54
Ilustración 56. Curvas de pérdida y precisión modelo 4.....	54
Ilustración 57. Métricas de desempeño modelo 5.....	55
Ilustración 58. Matriz de confusión modelo 5.....	55
Ilustración 59. Curvas de pérdida y precisión modelo 5.....	55
Ilustración 60. Métricas de desempeño modelo 6.....	56
Ilustración 61. Matriz de confusión modelo 6.....	56
Ilustración 62. Curvas de pérdida y precisión modelo 6.....	56
Ilustración 63. Métricas de desempeño modelo 7.....	57
Ilustración 64. Matriz de confusión modelo 7.....	57
Ilustración 65. Curvas de pérdida y precisión modelo 7.....	57
Ilustración 66. Métricas de desempeño modelo 8.....	58
Ilustración 67. Matriz de confusión modelo 8.....	58
Ilustración 68. Curvas de pérdida y precisión modelo 8.....	58
Ilustración 69. Métricas de desempeño modelo 9.....	59
Ilustración 70. Matriz de confusión modelo 9.....	59
Ilustración 71. Curvas de pérdida y precisión modelo 9.....	59
Ilustración 72. Métricas de desempeño modelo 10.....	60
Ilustración 73. Matriz de confusión modelo 10.....	60
Ilustración 74. Curvas de pérdida y precisión modelo 10.....	60
Ilustración 75. Métricas de desempeño modelo 11.....	61
Ilustración 76. Matriz de confusión modelo 11.....	61
Ilustración 77. Curvas de pérdida y precisión modelo 11.....	61
Ilustración 78. Métricas de desempeño modelo 12.....	62
Ilustración 79. Matriz de confusión modelo 12.....	62
Ilustración 80. Curvas de pérdida y precisión modelo 12.....	62
Ilustración 81. Métricas de desempeño modelo 13.....	63
Ilustración 82. Matriz de confusión modelo 13.....	63
Ilustración 83. Curvas de pérdida y precisión modelo 13.....	63
Ilustración 84. Métricas de desempeño modelo 14.....	64
Ilustración 85. Matriz de confusión modelo 14.....	64
Ilustración 86. Curvas de pérdida y precisión modelo 14.....	64

LISTA DE TABLAS

Tabla 1. Tipologías constructivas LADM-COL [13, p. 204].....	10
Tabla 2. Atributos relevantes de la base de datos.....	21
Tabla 3. Distribución de construcciones por Municipio.....	21
Tabla 4. Tipos de calificaciones según LADM-COL.....	22
Tabla 5. Rango de puntaje por tipología [2, p. 95].....	23
Tabla 6. Formato de fotografías.....	24
Tabla 7. Peso o Compresión de las fotografías.....	24
Tabla 8. Dimensiones de las fotografías.....	25
Tabla 9. Distribución de tipologías en el conjunto de datos.....	26
Tabla 10. Ejemplo de conformidad en nombre de fotografía.....	27
Tabla 11. Repositorio con etiquetas de fachadas original.....	30
Tabla 12. Repositorio con etiqueta de fachada depurada.....	31
Tabla 13. Especificaciones de modelos pre entrenados, tomado de [18].....	33
Tabla 14. Repositorio de fotografías de fachada con etiqueta.....	34
Tabla 15. Repositorio de fotografías de fachada con etiqueta depurada.....	35
Tabla 16. Puntuación de modelo convolucionales funcionales.....	40
Tabla 17. Puntuación de modelos SVM.....	45

INTRODUCCIÓN

En los procesos catastrales del país, y específicamente en la actividad de reconocimiento predial, es común que existan diferencias de criterio entre los profesionales e incluso entre los gestores catastrales o entidades que administran la información. Debe considerarse que el país se encuentra en un proceso de transición de políticas catastrales tradicionales a políticas con fines multipropósito, las cuales buscan estandarizar criterios. Las diferencias se dan en parte por la precaria interrelación entre las entidades que ejecutan un catastro tradicional, tal como se explica en el documento CONPES 3958 [1].

Dicha estandarización de criterios se implementa con la adopción de un modelo de datos único para el país, el modelo LADM-COL, el cual trata de un sistema de administración de tierras que se materializa como una base de datos relacional que busca la abstracción de la realidad territorial. La adopción de estos nuevos estándares no se da de manera rápida, sino que requiere un tiempo de transición para que los gestores adapten sus sistemas internos a los nuevos lineamientos.

En las actividades de reconocimiento predial, se identifican físicamente las construcciones de un predio. Dicha identificación comprende la calificación de las construcciones mediante una puntuación de ciertos aspectos reglamentados en el diccionario de datos del modelo LADM-COL. Esta puntuación lleva a la determinación de la tipología de construcción, la cual es un insumo fundamental para determinar el valor de los inmuebles.

En el presente documento se señala un problema de capacidad técnica de los profesionales que califican las construcciones, ocasionado por la cantidad limitada de profesionales disponibles con la experticia y criterio adecuado para generar productos de calidad. Esto se debe a que la demanda de dichos profesionales ha aumentado considerablemente debido a la puesta en marcha del catastro multipropósito, que tiene como meta lograr porcentajes altos de actualización catastral en el país. Basado en lo anterior, surge la necesidad de implementar estrategias que otorguen apoyo técnico al proceso catastral, específicamente en este proyecto, a la clasificación de tipologías dentro de actividades de reconocimiento predial de manera estandarizada y en línea con el catastro multipropósito.

De acuerdo con el problema y las necesidades identificadas, se desarrolló este proyecto con el objetivo de implementar un modelo de clasificación de imágenes utilizando técnicas de aprendizaje automático. Este modelo permitió llevar a cabo tareas de tipificación de construcciones siguiendo el estándar LADM-COL y podrá proporcionar un apoyo significativo al proceso de reconocimiento predial.

1. DEFINICIÓN DEL PROBLEMA

La calificación de construcciones es un producto del reconocimiento predial que se basa en dar puntajes a los elementos constructivos siguiendo un manual estandarizado. Por ejemplo, una construcción con uso residencial puede tener estructura en concreto, muros en bloque, cubierta en zinc y un estado de conservación regular, esa clasificación dará un valor numérico en forma de “puntos”, los cuales se suman a la calificación de acabados principales, baño y cocina, para obtener un puntaje total. Dicho puntaje define la tipología constructiva, el cual es un aspecto fundamental en los procesos catastrales, puesto que la determinación de valor de los inmuebles es una variable significativa [2].

En [2], el Instituto Geográfico Agustín Codazzi describe las etapas del reconocimiento predial, donde la calificación de construcciones es solo una parte de la identificación física, que, además de determinar la tipología constructiva, se identifica el uso, las dimensiones, el número de unidades de construcción, el estado de conservación y el número de plantas o niveles que componen la estructura.

Durante las visitas de campo es común que a los reconocedores prediales se les exija, como se menciona en [3], tomar registros fotográficos de todas las construcciones de un inmueble, incluyendo su fachada y las dependencias interiores, para posteriormente realizar la calificación en oficina apoyados en dichos soportes de campo. Esta actividad conlleva un tiempo de trabajo considerable y aunque se trata de un proceso estandarizado, el criterio y experiencia del reconocedor juega un papel importante, ya que en ocasiones la perspectiva de calificación puede variar entre profesionales generando así inconsistencias en la correcta calificación de las construcciones.

1.1 PLANTEAMIENTO DEL PROBLEMA

La situación actual de los procesos catastrales es un tema de interés nacional, lo anterior debido a la política pública de catastro multipropósito que sienta sus bases en el CONPES 3859 de 2016. Dicha política lleva consigo metas, planes de acción y estrategias, las cuales han tenido retrasos y dificultades en los tiempos de ejecución. Una de las dificultades proyectadas en el CONPES [4, p. 28] es la de encontrar personal capacitado para los procesos, como por ejemplo los reconocedores prediales, que como ya se mencionó, es el recurso indispensable que debe tener un criterio y experiencia adecuada para una correcta calificación de construcciones. Esto también ha sido identificado por las entidades colombianas, las cuales han propuesto iniciativas como cursos de capacitación de personal, o como plantea el director del IGAC en [5], crear registros reglamentados de profesionales avalados para dicha actividad.

Dado lo anterior, se evidencian dificultades con: el tiempo de ejecución de los procesos, la calidad de los productos generados y la disponibilidad de personal capacitado. En cuanto a la calidad, el Instituto Geográfico Agustín Codazzi y los gestores catastrales cuentan con procedimientos de evaluación específicos para el reconocimiento predial basados en muestreos de productos, en campo y oficina, que son ejecutados siguiendo estándares o manuales preestablecidos, con niveles de conformidad fundamentados en normas técnicas tal como se explica en [6].

Con base en los problemas mencionados, es necesario que se mejore la capacidad técnica de los profesionales en cuanto a la cantidad de personas disponibles y la calidad de sus productos. Una alternativa para superar dicha situación es involucrar nuevas tecnologías en los procesos catastrales que permitan reducir tiempos, implementar un trabajo más eficiente y de calidad por parte de los profesionales.

En resumen, la calificación de construcciones en el reconocimiento predial es un proceso basado en el aprendizaje y experiencia previa del profesional, lo cual conlleva tiempo y, además, si no se cuenta con la cantidad de profesionales y experticia necesaria de los mismos para evaluar grandes volúmenes de información, se dificulta la tarea, tal como sucede los barridos prediales masivos, donde se intervienen municipios enteros en un corto periodo de tiempo.

1.2 FORMULACIÓN DEL PROBLEMA

En diferentes momentos de la historia, la realidad física se ha intentado representar con el fin de solucionar problemas con los conocimientos y tecnologías disponibles para cada época, tal como menciona García et al. [7]; desde el nacimiento de la escritura, pasando por el desarrollo de las ciencias de la computación y las tecnologías, hasta la actualidad con disciplinas como el Big Data y la inteligencia artificial. La adopción de la tecnología disponible, como solución a los problemas en este proyecto de índole catastral, se aplicó como un proceso de adaptación a las necesidades.

En ese sentido el estudio y análisis del comportamiento humano ha llevado al desarrollo de simulaciones de actividades intelectuales del hombre en aspectos como la comprensión, percepción, resolución de problemas y la toma de decisiones, tal como señala Thomas en [8] y su definición de inteligencia artificial, con el objetivo de imitar o mejorar las capacidades humanas en ciertos aspectos.

En cuanto a la resolución de problemas, Wang *et al* [9] exponen que las imágenes llevan consigo mucha información y que, en los procesos masivos que involucran la captura y procesamiento de imágenes, es importante encontrar información útil. Dicha utilidad puede extraerse con algoritmos de inteligencia artificial, en especial con técnicas de aprendizaje automático o aprendizaje profundo.

De acuerdo con el problema relacionado con la capacidad técnica de los profesionales se planteó la pregunta de investigación:

- ¿Cómo construir un modelo de clasificación de imágenes a partir de técnicas de aprendizaje automático con el fin de llevar a cabo tareas de tipificación de construcciones de manera estandarizada y automatizada?

De manera detallada se formularon las preguntas de sistematización:

- ¿De qué manera se puede implementar una estrategia de captura y procesamiento de imágenes dentro de un conjunto de datos relacionado con construcciones?

- ¿Cómo se desarrolla un modelo de clasificación de imágenes a partir de técnicas de aprendizaje automático?

- ¿De qué manera se valida un modelo de clasificación de construcciones a partir de métricas de desempeño?

2. OBJETIVOS DEL PROYECTO

2.1 OBJETIVO GENERAL

Desarrollar un modelo de clasificación de imágenes a partir de técnicas de aprendizaje automático con el fin de llevar a cabo tareas de tipificación de construcciones siguiendo el estándar del Modelo LADM-COL.

2.2 OBJETIVOS ESPECÍFICOS

2.2.1. Implementar una estrategia de recolección y procesamiento de imágenes dentro un conjunto de datos relacionado con construcciones.

2.2.2. Desarrollar un modelo de clasificación de imágenes a partir de técnicas de aprendizaje automático.

2.2.3. Validar el modelo de clasificación construcciones a partir de métricas de desempeño.

3. MARCO DE REFERENCIA

3.1 MARCO TEÓRICO

A continuación, se presentan aspectos clave que se profundizaron para lograr el desarrollo de los objetivos propuestos en este proyecto, partiendo de conceptos catastrales, siguiendo con aspectos técnicos y las normas que los rigen, analizando también la importancia de los nuevos estándares en cuanto a la información catastral para posteriormente introducir conceptos de aprendizaje automatizado.

5.1.1. CONCEPTOS CATASTRALES

Catastro: Una de sus definiciones más actuales se da en el Decreto 148 de 2020 como: “El inventario o censo de los bienes inmuebles localizados en el territorio nacional, de dominio público o privado, independiente de su tipo de tenencia, el cual debe estar actualizado y clasificado con el fin de lograr su identificación física, jurídica y económica con base en criterios técnicos y objetivos” [10, p. 3].

Catastro Multipropósito: Una de sus primeras definiciones es dada por el Departamento Nacional de Planeación en [10, p. 3] como: “Es aquel en que la información que se genera a partir de su implementación, debe servir como un insumo fundamental en la formulación e implementación de diversas políticas públicas, contribuyendo a brindar una mayor seguridad jurídica, la eficiencia del mercado inmobiliario, el desarrollo y el ordenamiento territorial, integrada con el registro público de la propiedad inmueble, digital e interoperable con otros sistemas de información del territorio, y que provea instrumentos para una mejor asignación de los recursos públicos y el fortalecimiento fiscal de los territorios”.

Reconocimiento predial: Se define de diferentes maneras dependiendo del manual o normal que se consulte, sin embargo, en el Manual de Reconocimiento predial de Antioquia se tiene una definición que se puede considerar completa: “Es la verificación de los elementos físicos y jurídicos del predio, mediante la práctica de la inspección catastral para identificar su ubicación, linderos, extensión, mejoras por edificaciones y precisar el derecho de propiedad o de posesión” [3, p. 8].

Calificación de construcciones: Se define como una actividad propia de la verificación física del reconocimiento predial donde se identifica mediante visita en campo para cada construcción, el tipo (Residencial, Comercial o Industrial), los materiales, el estado de conservación, para cada elemento de la estructura, los acabados, el baño y la cocina se asigna un puntaje el cual define la tipología constructiva [11].

Tipologías constructivas: Se define de manera general por el Departamento Nacional de Planeación como “la clasificación o categorización de las características para las cuales fueron creadas las construcciones y/o edificaciones, que comprende la estructura, acabados, altura y los muros, entre otros.” [10, p. 4].

Modelo LADM-COL: Según la Superintendencia de Notariado y Registro y el Instituto Geográfico Agustín Codazzi [12], corresponde una adaptación para el caso colombiano del estándar de la norma ISO 19152:2012, la cual se podría definir como un modelo de administración territorial para Colombia, que de manera estandarizada, busca la interoperabilidad de los sistemas catastrales y que con modelos de aplicación específicos, como el modelo de levantamiento catastral, se materializan en bases de datos relacionales.

5.1.2. ASPECTOS TÉCNICOS Y NORMATIVIDAD

En cuanto a los aspectos técnicos y normatividad colombiana, es necesario a manera de contexto mencionar que desde el año 2016 se trazó un curso de acción diferente, en cuanto a las políticas catastrales, iniciando con el **Plan Nacional de Desarrollo (PND) 2014-2018**, que dio origen al **CONPES 3859** y sentó las bases de un Catastro Multipropósito, para luego sufrir algunas modificaciones con el **PND 2018-2022**, que dio origen al **CONPES 3958**; este último dio origen a la política pública de catastro multipropósito actual, a partir de allí han surgido a manera de Decretos, leyes y resoluciones cambios técnicos importantes. Dichos cambios se deben a profundos problemas estructurales en el catastro tradicional, tal como se menciona en el documento CONPES 3958 [1].

5.1.3. NUEVOS ESTÁNDARES

Basado en el nuevo curso de la política catastral surgen nuevos criterios, los cuales buscan interoperabilidad entre los sistemas y estandarización de la información catastral, como la resolución conjunta **SNR 04218 IGAC 499 de 2020** por la cual se adopta el modelo extendido de catastro registro del modelo LADM-COL [12].

En cuanto al proceso de reconocimiento predial en el catastro tradicional, llamando así al catastro antes de la política de catastro multipropósito, las entidades catastrales descentralizadas manejaban casi de manera independiente sus criterios técnicos. Aunque estos criterios eran muy similares en algunos aspectos, cada entidad tenía su propio manual, por ejemplo, el Instituto Geográfico Agustín Codazzi manejaba un manual de reconocimiento predial [2] y la gobernación de Antioquia manejaba otro diferente [3]. Esto obligaba a los reconocedores a estudiar y aprender distintos criterios según la zona del país, tal como menciona Zuluaga en [11].

Este tipo de diferencias en criterio se eliminan con la adopción de estándares como el modelo LADM-COL, por ejemplo, para el caso de las tipologías constructivas, la gobernación de Antioquia modificó sus tipologías y criterios técnicos mediante resolución para adaptarse al Modelo LADM-COL. Dado lo anterior, y para efectos de garantizar un criterio unificado, este proyecto se basa en la clasificación de tipologías constructivas del modelo LADM-COL.

De la mano del modelo LADM-COL y las nuevas políticas en materia de catastro se tiene la resolución 1040 de 2023 [13], la cual se establece como una resolución que unifica la gestión catastral multipropósito en un documento extenso. De dicha norma y para efectos de este proyecto se destaca el diccionario de datos del modelo LADM-COL vigente, donde es posible encontrar las definiciones más actuales de tipologías constructivas. A continuación, se presentan en la siguiente tabla, destacando únicamente aquellas de uso residencial.

Nombre	Alias	Descripción
Residencial	Residencial	
Tipo_1A	(Residencial) Tipo 1A	Armazón: Madera; Muro: Material de desecho, esterilla; Cubierta material desecho, telas asfálticas; Pisos tierra pisada.
Tipo_1B	(Residencial) Tipo 1B	Armazón: Madera; Muro: Material de bahareque, adobe, tapia; Cubierta: Zinc, teja de barro, Eternit rustico; Pisos: Cemento, madera burda.
Tipo_2A	(Residencial) Tipo 2A	Armazón: Madera; Muros en madera; Muro: Madera Cubierta: Zinc, teja de barro, Eternit rustico; Cubrimiento de Muros: Pañete, papel común, ladrillo presado; Pisos: Cemento, madera burda.
Tipo_2B	(Residencial) Tipo 2B	Armazón: Prefabricado; Muros: Concreto prefabricado; Cubierta: Zinc, teja de barro, Eternit rustico; Cubrimiento de Muros: Pañete, papel común, ladrillo presado; Pisos: Baldosa común de cemento, Tablón ladrillo.
Tipo_3A	(Residencial) Tipo 3A	Armazón: Ladrillo, bloque; Muro: ladrillo, Bloque; Cubierta: Eternit o teja de barro; Cubrimiento de muros: Pañete, papel común, ladrillo prensado; Piso: Baldosa común de cemento, tablón ladrillo.
Tipo_3B	(Residencial) Tipo 3B	Armazón: Ladrillo, bloque; Muro: ladrillo, Bloque; Cubierta: Azotea, aluminio, placa sencilla con Eternit o teja de barro; Cubrimiento muros: Estuco, cerámica, papel fino; Piso: Baldosa común de cemento, tablón ladrillo.
Tipo_4A	(Residencial) Tipo 4A	Armazón: Concreto hasta tres pisos, Muro: Bloque, ladrillo; Cubierta: Azotea, aluminio, placa sencilla con Eternit o teja de barro; Cubrimiento muro: estuco, cerámica, papel fino; Piso: Listón machi-embredo
Tipo_4B	(Residencial) Tipo 4B	Armazón: Concreto hasta tres pisos, Muro: Bloque, Ladrillo; Cubierta: Azotea, aluminio, placa sencilla con Eternit o teja de barro; Cubrimiento muro: estuco, cerámica, papel fino; Piso: Tableta, caucho, acrílico, granito, baldosa fina
Tipo_5A	(Residencial) Tipo 5A	Armazón: Concreto hasta tres pisos; Muro: Bloque, Ladrillo; Cubierta: Azote, Aluminio, placa sencilla con Eternit o teja de barro; Cubrimiento muro: Medra, Piedra ornamental; Piso: Tableta, caucho, acrílico, granito, baldosa fina.
Tipo_5B	(Residencial) Tipo 5B	Armazón: Concreto cuadro y más pisos; Muro: Bloque, Ladrillo; Cubierta: Placa impermeabilizada cubierta lujosa u ornamental; Cubrimiento muro: Mármol, lujoso, otros; Piso: Parquet, alfombra, retal de mármol (grano pequeño).
Tipo_6A	(Residencial) Tipo 6A	Armazón: Concreto hasta tres pisos; Muro: Bloque, ladrillo; Cubierta: Placa impermeabilizada cubierta lujosa u ornamental; Cubrimientos muros: Mármol, lujosos otros; Piso: Parquet, alfombra, Retal de mármol (grano pequeño)
Tipo_6B	(Residencial) Tipo 6B	Armazón: Concreto cuatro o más pisos; Muro: Bloque, ladrillo; Cubierta: Placa impermeabilizada cubierta lujosa u ornamental; Cubrimientos muros: Mármol, lujosos otros; Piso: Retal de mármol, mármol, otros.

Tabla 1. Tipologías constructivas LADM-COL [13, p. 204]

5.1.4. APRENDIZAJE AUTOMATIZADO

Inteligencia Artificial:

La inteligencia artificial (IA) tiene por objetivo el estudio y el análisis del comportamiento humano en los ámbitos de la comprensión, de la percepción, de la resolución de problemas y de la toma de decisiones, con el fin de poder reproducirlos con la ayuda de un computador. De esta manera, las aplicaciones de la IA se sitúan principalmente en la simulación de actividades intelectuales del hombre, es decir, imitar por medio de máquinas, normalmente electrónicas, tantas actividades mentales como sea posible, y quizás llegar a mejorar las capacidades humanas en estos aspectos tal como menciona Thomas en [8].

Aprendizaje Automático – Machine Learning

El aprendizaje automático es el núcleo del campo de la inteligencia artificial. Esta se describe como una disciplina multidisciplinaria e interdisciplinaria que involucra teoría de probabilidad, estadística, teoría de aproximación, análisis convexo, teoría de complejidad algorítmica y otras disciplinas; y se especializa en cómo las computadoras simulan o implementan el comportamiento de aprendizaje humano para adquirir nuevos conocimientos o habilidades y reorganizar las estructuras de conocimiento existentes para mejorar su propio rendimiento, tal como se menciona en [9].

Red Neuronal Artificial – Artificial neural network

En [9], Wang *et al.* señala que es un modelo matemático que abstrae y simula la forma de procesar información del sistema nervioso central biológico según el conocimiento de la topología de la red. La red neuronal tiene la capacidad de procesamiento de datos de forma paralela y distribuida, aprendizaje de características adaptativas, una buena tolerancia a fallos y una robustez considerable.

Modelo de Red Neuronal – Neural network model

En [9], Wang *et al.* mencionan que una red neuronal artificial está formada por una gran cantidad de unidades de procesamiento de información conectadas entre sí. Estas unidades de procesamiento de información se llaman neuronas artificiales. La construcción del modelo de neuronas artificiales se origina en la estructura de la neurona biológica. Los principales componentes de la neurona biológica incluyen el cuerpo celular, el axón y las dendritas. La neurona artificial establece la estructura artificial mediante la simulación de la neurona biológica. La neurona artificial se compone de tres partes: múltiples pesos de conexión, un término de suma y una función de activación no lineal.

Debido a que la topología de la red no es de bucle cerrado o bucle, el modelo de red es una red neuronal de alimentación directa. Dependiendo de la capacidad de estas numerosas neuronas para procesar y transmitir información, se pueden realizar tareas de extracción de características de información compleja y predicción utilizando el modo de procesamiento paralelo distribuido para mapear la señal no lineal [9].

Aprendizaje Profundo – Deep Learning

En [9], Wang *et al.* explican que es un nuevo campo en el aprendizaje automático. Su motivación es construir y simular la red neuronal del cerebro humano para análisis y aprendizaje. Imita el mecanismo del cerebro humano para interpretar datos, como texto, sonido e imagen. Es una forma de aprendizaje no supervisado. Su concepto proviene de la investigación de redes neuronales artificiales, por lo que también se le conoce como red neuronal profunda.

Red Neuronal Convolutacional - Convolutional Neural Network (CNN)

Wang *et al.* en [9], señalan que es un modelo de Deep Learning que tiene características especiales, diferentes a las redes neuronales tradicionales, puesto que debido a que se basa en el principio del procesamiento visual en la biología, tiene capacidades fuertes en extraer información de imágenes sin destruir la estructura de entrada y sin requerir tantos parámetros, lo que se traduce en menos tiempo de entrenamiento.

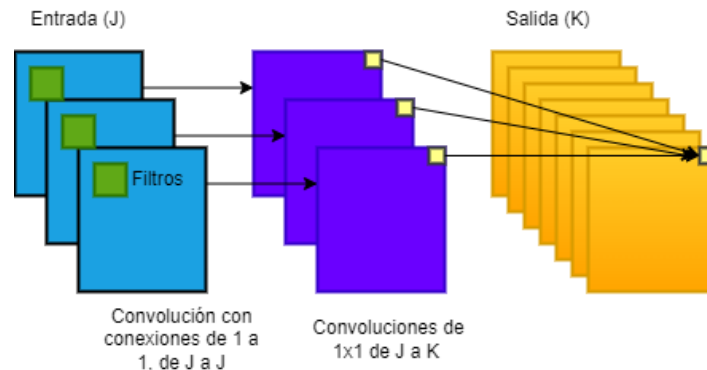


Ilustración 1. Representación Red Neuronal Convolutacional (CNN) Fuente: García et al [7]

Máquina de Vectores de Soporte – Support Vector Machine (SVM)

Es un poderoso modelo de Machine Learning tradicional, es común como método de discriminación. En comparación con otros modelos de clasificación, tiene una mejor capacidad de generalización. SVM es principalmente aplicable a dos situaciones; la primera categoría son datos linealmente separables, y la segunda categoría son datos linealmente inseparables, según lo establecido en [9].

3.2 ANTECEDENTES

A continuación, se exponen aquellas investigaciones que se consideraron relevantes para el desarrollo de este proyecto. El criterio de selección se fundamenta principalmente en su similitud con este proyecto, valorando los métodos utilizados, el aporte teórico, la claridad en el desarrollo de las etapas y los resultados obtenidos.

Utilización de las máquinas con vectores de soporte para regresión: m2 de construcción Bogotá (2009)

En esta investigación, Morales y Hernández [14], señalan que el procedimiento de calcular el valor por metro cuadrado de construcción es una actividad subjetiva, debido a que depende de factores como la experiencia de los profesionales; dicha aclaración va de la mano con lo planteado en este proyecto puesto que los reconocedores prediales y sus actividades de calificación-tipificación de construcciones también se basan en aspectos subjetivos que se definen con la experiencia de los individuos.

En [14], se evidencia que existe una relación entre la calificación de construcciones y la determinación del valor de m2 de construcciones; tal como señalan los autores, una de las variables que explica el comportamiento del valor del m2 está asociada al puntaje que se les asigna a las construcciones por medio de la calificación, que finalmente definen una tipología constructiva.

Los autores señalan que su investigación es la continuación de dos trabajos anteriores, en los cuales utilizaron dos técnicas diferentes, el primero utilizó un modelo econométrico con resultados poco precisos y el segundo redes neuronales con resultados adecuados. Basado en lo anterior, su objetivo fue utilizar una técnica diferente que obtenga resultados comparables al uso de redes neuronales. El modelo seleccionado, luego de varias exploraciones, fue un modelo basado en regresión Ridge con un Kernel Polinomial, el cual obtuvo un punto medio en cuanto a la comparación con los dos modelos [14].

En cuanto a la metodología, Morales y Hernández [14], utilizaron una función Kernel Polinomial de grado 5, obteniendo finalmente un error de 171.521 pesos colombianos frente a los 28.700 pesos que se obtuvieron con el uso de RNA. En el proyecto se utilizó un conjunto de datos con 2.626 registros, de los cuales el 80% se utilizó para el entrenamiento y el 20% para validaciones.

A manera de conclusión, se señala que los modelos de regresión no lineales presentan mejores resultados que los modelos lineales, como el econométrico, ya que los patrones encontrados en la investigación evidencian relaciones no lineales entre las variables explicativas.

Inteligencia artificial para la asignación automática de categorías constructivas (2018)

En esta investigación, los autores García et al. [7] tienen como objetivo construir una herramienta para clasificar las tipologías y categorías constructivas de viviendas urbanas de manera automática, lo anterior muestra una similitud considerable con el proyecto propuesto en el presente documento, por lo cual se toma como una referencia o punto de análisis relevante.

Los autores mencionan en [7] que es posible lograr un algoritmo con aplicaciones reales, pero para ello es necesario normalizar el formato de las imágenes, añadir metadatos a las mismas y tener acceso a cantidades considerables de ejemplos, lo anterior se podría resumir en que la etapa de implementación de una estrategia de captura y procesamiento de imágenes debe ser muy rigurosa.

El vincular datos estructurados, los cuales se entienden como las clases de tipologías constructivas y datos procedentes de imágenes, es una innovación para los procedimientos catastrales, entendiéndose que, lo que se pretende modelar son aspectos subjetivos. Dicha subjetividad puede tener un sesgo, según los autores en [7], el cual puede reproducirse o propagarse en el entrenamiento del algoritmo si no se controla qué tipo de datos son utilizados para la tarea.

En la investigación se utilizan dos variables, la tipología y la fotografía de fachada, en la identificación de la tipología se seleccionan aquellos parámetros que pueden interpretarse en la fotografía, ya que la calificación de la construcción determina la tipología constructiva al puntuar parámetros como la estructura, acabados principales, baño y cocina, pero para el desarrollo del presente proyecto, no es posible considerar como parámetros el baño y la cocina, si estos no se pueden identificar en las fotografías de fachada. A manera de conclusión, se pretende interpretar una variable discreta de carácter subjetivo y seleccionar los parámetros que pueden identificarse en una fotografía, para que, de esta manera aporten significativamente a una correcta tipificación [7].

En cuanto a los inconvenientes de los tratamientos de los datos, en [7] los autores señalan que, debe definirse un alcance respecto a las modalidades dentro del uso. Esto se debe a la disponibilidad de datos, ya que es necesario evaluar la cantidad de datos disponibles para entrenar clasificaciones de tipo comercial, industrial, institucional y residencial, puesto que estas no se presentan de manera heterogénea en el territorio.

En [7] se explica cómo en el proceso de clasificación de imágenes se identifican dos fases: el preprocesado y procesado, el preprocesado tiene que ver con captura de las fotografías, donde el ángulo, el formato, la compresión, el tamaño, el

color, los obstáculos o las alteraciones deben tenerse en cuenta para que, en el momento del procesamiento o entrenamiento del modelo, los datos tengan calidad.

En cuanto a la metodología, en [7] los autores utilizan una combinación de redes neuronales convolucionales para extraer las características de las imágenes y máquinas con vectores de soporte para las clasificaciones de las características. Esta decisión se tomó debido a que las redes neuronales han sido utilizadas para encontrar contornos o características de manera automática, y las máquinas con vectores de soporte son una técnica habitual, sencilla y de buen rendimiento los procesos de clasificación.

Para la construcción del modelo, los autores en [7] utilizan un conjunto de datos con 5802 imágenes de viviendas de uso únicamente residencial, obteniendo como resultado un índice de acierto superior al 85%. Las clasificaciones no consideradas conformes se contemplan como casos puntuales que deben ser evaluados por asistentes humanos, sirviendo la herramienta como un apoyo y no como una sustitución de las tareas de los profesionales, donde dicho respaldo puede ser en cuanto al control de calidad o en la producción de clasificaciones simples que solo deban aceptarse por operarios humanos.

Diseño y evaluación de un sistema experto difuso para la calificación de construcciones de tipología media y baja de uso residencial en el reconocimiento predial urbano (2020)

En [15], Cortés utiliza un sistema denominado experto difuso con el objeto de obtener calificaciones de construcciones de manera automatizada. Aunque este proyecto no involucra el análisis de fotografías para la determinación de tipologías, presenta un análisis interesante de los parámetros necesarios para la tipificación de las construcciones a través de las calificaciones por puntajes y la opinión de expertos, lo anterior es de suma importancia en el presente proyecto.

El autor señala en [15] como aspecto relevante el hecho de que la lógica difusa permite trabajar con información imprecisa o subjetiva, apoyada de una base de conocimientos que se modela a través de reglas lógicas, donde no necesariamente una clasificación debe basarse en 1 y 0, sino que puede también considerar puntos intermedios; por ejemplo, la tipología constructiva al ser un aspecto subjetivo, no debería evaluar sus características basándose en si está o no dentro de determinada clase, sino a la cercanía o grado de afinidad con alguna de ellas.

En [15] se explica cómo la base de conocimientos fue construida con encuestas a 10 profesionales experimentados en reconocimiento predial, con el objetivo de identificar falencias en los manuales o fichas prediales que estandarizan la clasificación por puntajes, los resultados de dichas encuestas aportan variables que no se tienen contempladas en los manuales. En la ilustración número 2, el autor plasmó la metodología del sistema difuso, mostrando cada una de las etapas del desarrollo del proyecto.

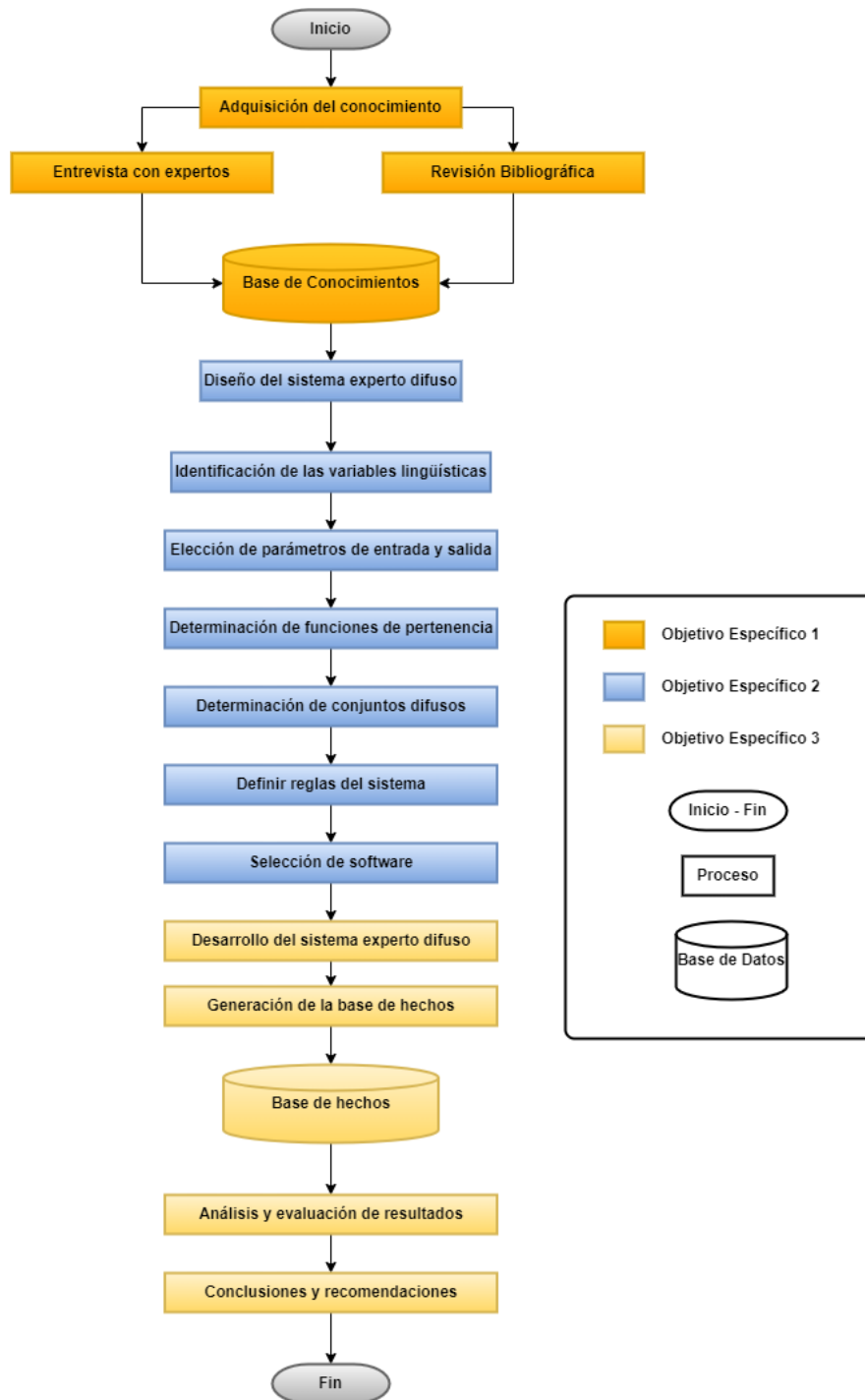


Ilustración 2. Sistema Difuso Metodología Fuente: Cortés [15, p. 62]

La metodología se fundamenta en Redes neuronales aplicadas en el software Matlab. Estas redes basadas en reglas “si, entonces”, siguen una estructura jerárquica de árbol en las cuales evalúan los parámetros seleccionados en la construcción de la base de conocimientos, dichos parámetros fueron seleccionados basados en las fichas prediales, manuales catastrales y la opinión de los expertos. En la Ilustración 3 se presentan los parámetros de entrada y los parámetros de salida:

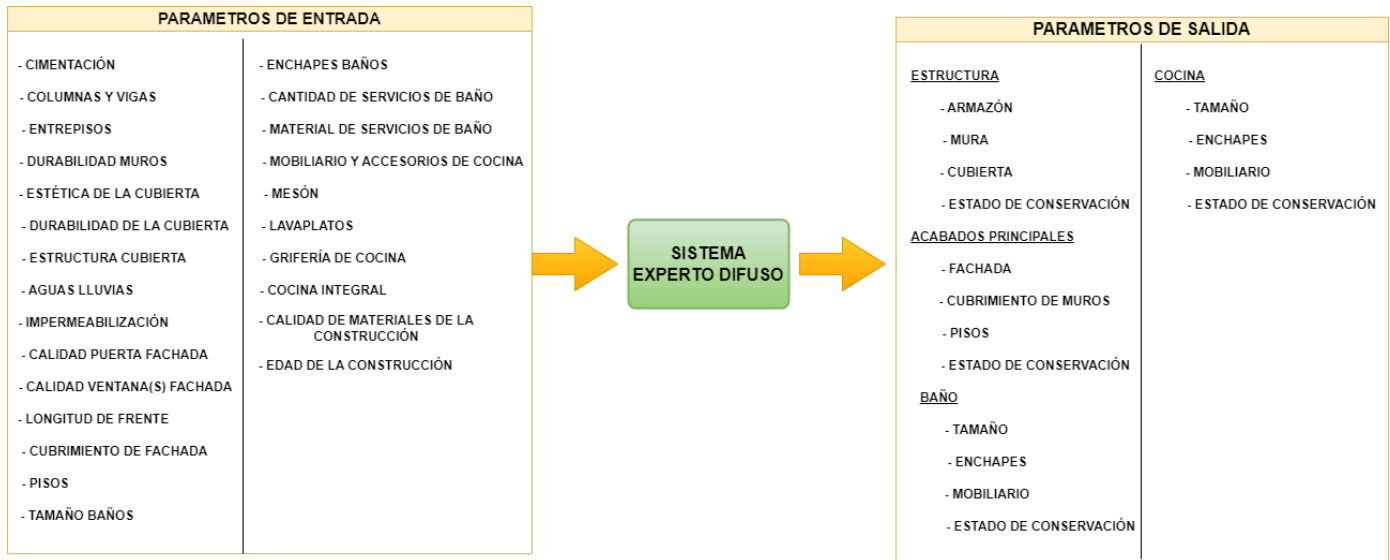


Ilustración 3. Parámetros del Sistema Difuso Fuente: Cortés [15]

Para la caracterización de cada parámetro, el sistema experto difuso desarrollado por Cortés [15] contempla un conjunto de fórmulas para cada clase, las cuales se denominan funciones de membresía, estas asignan un valor en el rango de los números reales alejándose así de una clasificación tipo booleana.

Como parte de la metodología empleada en [15], se utilizó una técnica de validación que consiste en la comparación de 50 calificaciones, dicha comparación se da entre las calificaciones dadas por el Instituto Geográfico Agustín Codazzi (IGAC), consideradas como verdaderas, la calificación de estudiantes de Ingeniería Catastral y los resultados del sistema difuso construido. Los resultados presentan una variabilidad considerable, que puede deberse a la poca experticia de la evaluación de los estudiantes, y a que, si bien la fuente del IGAC es oficial, esta puede presentar errores. Dado lo anterior, del proyecto se destaca no por sus resultados, sino por la metodología de identificación y clasificaciones de parámetros para la determinación de la tipología.

Determinación de características constructivas de un predio mediante el análisis de fachada por medio de una red neuronal artificial para una UPZ de Bogotá (2020)

En la investigación Venegas [16] tiene como objetivo “Diseñar una red neuronal artificial para la determinación masiva de las características constructivas de los predios de uso residencial basados en la fachada”, lo anterior tiene similitud con el proyecto estructurado en el presente documento, por lo cual se tomará también como un punto de referencia.

En [16], se explica que el alcance del proyecto dentro de la metodología planteada se limita a un área urbana de la ciudad de Bogotá, donde se utilizan redes neuronales artificiales y arboles de decisión únicamente para construcciones con uso residencial. En la siguiente ilustración el autor muestra el flujo metodológico planteado de la investigación:

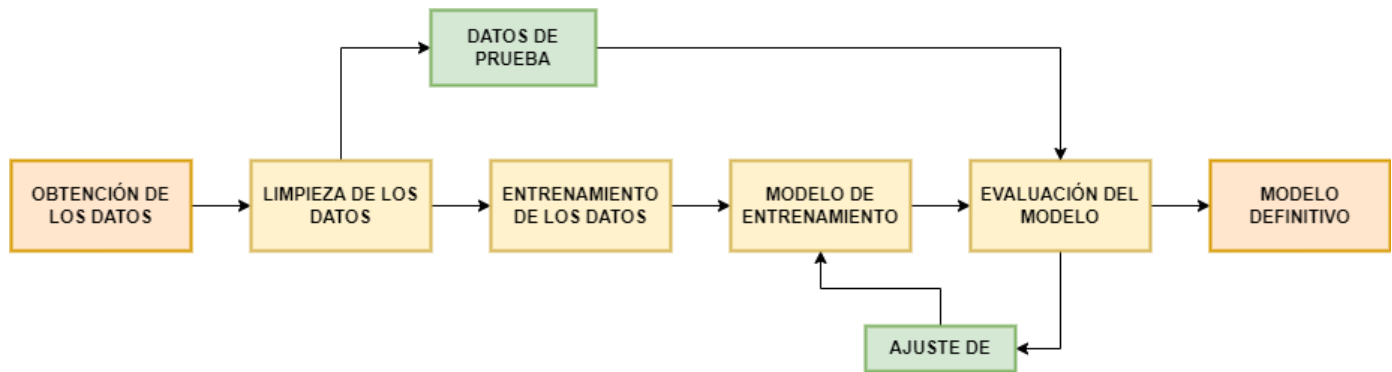


Ilustración 4. Diagrama de flujo metodológico Fuente: Venegas [15]

Respecto a la definición de los parámetros, Venegas en [16] considera únicamente los valores de la ficha predial del Instituto Geográfico Agustín Codazzi, seleccionando los 8 primeros campos correspondientes con la estructura y los acabados principales, ya que pueden ser distinguidos en una fotografía de fachada; adicional a estos, se considera otra variable que permite descartar los predios con régimen de propiedad horizontal, de los que no tienen esta condición; la variable anterior es importante en zonas urbanas, puesto que evaluar tipologías constructivas de un edificio o conjunto residencial puede ser una tarea con un nivel de dificultad superior.

En [16], Venegas aplica en el desarrollo de la metodología 20 combinaciones con diferentes neuronas y redes, obteniendo indicadores de pérdida, ajuste y tiempo de procesamiento, con los que discriminó el mejor resultado. Una vez seleccionado la mejor red con un valor de pérdida de 0.24 y un porcentaje de ajuste del 89.5 %, los datos de entrada fueron ingresados a dicha red de manera transformada por medio de un preprocesamiento, el cual indicó si se pertenecía o no a algunas de las 6 tipologías dispuestas para el uso residencial, obteniendo como resultado final un vector de probabilidades con 6 valores del cual se seleccionó como tipología final aquel con mayor probabilidad.

En cuanto a los resultados expuestos en [16], tanto para los obtenidos con la red neuronal como para aquellos provenientes de árboles de decisiones se evidencia sobre ajuste, el cual es mencionado, pero el autor no le da importancia por su magnitud. Es importante, y a manera de conclusión del autor, resaltar que los parámetros de la ficha predial se quedan cortos para realizar la tipificación de la fachada, por lo que se sugiere incorporar variables adicionales que permitan la obtención de mejores resultados.

Comparative analysis of image classification algorithms base on traditional machine learning and Deep learning (2020)

La investigación tiene como objetivo realizar una evaluación de desempeño en la clasificación de imágenes de los algoritmos de CNN y SVM, en cuanto a cuál de ellos presenta mejor precisión y tiempo de procesamiento; sumado a esto, Wang *et al.* [9] realiza una descripción teórica útil dando al lector una comprensión de las diferencias y similitudes entre los dos algoritmos, incluyendo las funciones y componentes de estos.

Respecto a la comparación realizada por Wang *et al.* [9] señalan que la red neuronal convolucional (CNN), por sus siglas en inglés, presenta características similares al método biológico de procesar imágenes por parte de las neuronas, resaltando que los algoritmos tradicionales, al contrario que CNN, rompen la estructura de los píxeles, lo que implica que estos sean reorganizados y presenten tiempos de procesamiento menos eficientes. Otro aspecto importante de CNN es que la imagen global puede ser convolucionada usando el operador de convolución para extraer las mismas características en diferentes posiciones de la imagen global. A través del análisis del comportamiento de parámetros, estos se reducen en la red considerablemente junto con el tiempo de entrenamiento. Los autores señalan que la

estructura comúnmente utilizada de CNN incluye la capa de entrada, la capa de convolución, la capa de activación, la capa de agrupación, la capa completamente conectada y la capa de salida.

Por otro lado, en [9] se explica que la Máquina de Vector de Soporte (SVM), por sus siglas en inglés, es un modelo de clasificación muy poderoso en el aprendizaje automático, siendo este un método discriminante común. En comparación con otros modelos de clasificación, tiene una mejor capacidad de generalización, y para datos no linealmente separables, cuenta con un conjunto de métodos teóricos avanzados para abordarlos. Para casos linealmente separables, el método específico consiste en encontrar la línea con la suma más grande de distancias entre los puntos vecinos para separarlos; para casos linealmente inseparables, se necesita una función de Kernel.

Respecto a la comparación realizada por Wang *et al.* en [9], se utilizó una metodología que consistía en cambiar el tamaño de las imágenes para validar el efecto de los algoritmos, en este caso, se obtuvo un mejor desempeño con la CNN en imágenes grandes, por otro lado, se compararon teniendo en cuenta diferentes tipos de imágenes y tamaños en el conjunto de datos de entrenamiento, donde las variaciones no fueron significativas pero tuvieron una mejor respuesta en tiempos de procesamiento con SVM. A manera de conclusiones, se obtiene entonces que en los conjuntos de datos de escala pequeña los modelos tradicionales como SVM tienen más ventajas en la clasificación, mientras que en conjuntos de datos a gran escala y en precisión de reconocimiento, los modelos de aprendizaje profundo como CNN tienen mejores capacidades.

4. DESARROLLO DEL PROYECTO

4.1 METODOLOGÍA

Las etapas mencionadas a continuación y sus correspondientes pasos, son la ruta a seguir para replicar la metodología sugerida en el presente proyecto:

Estrategia de recolección y procesamiento de imágenes

Paso 1: Planificación de la base de datos

En esta etapa se realizó la solicitud al Programa Nuestra Tierra Prospera de permitir el uso de los datos recopilados en los diferentes barridos prediales masivos que han impulsado en el territorio nacional. Tras una respuesta positiva y la firma de un acuerdo de confidencialidad, se compartió la siguiente información:

Fuente de datos 1: Archivos de transferencia XTF o Back ups de las bases de datos relacionales.

Contienen información geográfica y alfanumérica de construcciones estandarizadas bajo el modelo de datos LADM COL en su versión de levantamiento 1.0

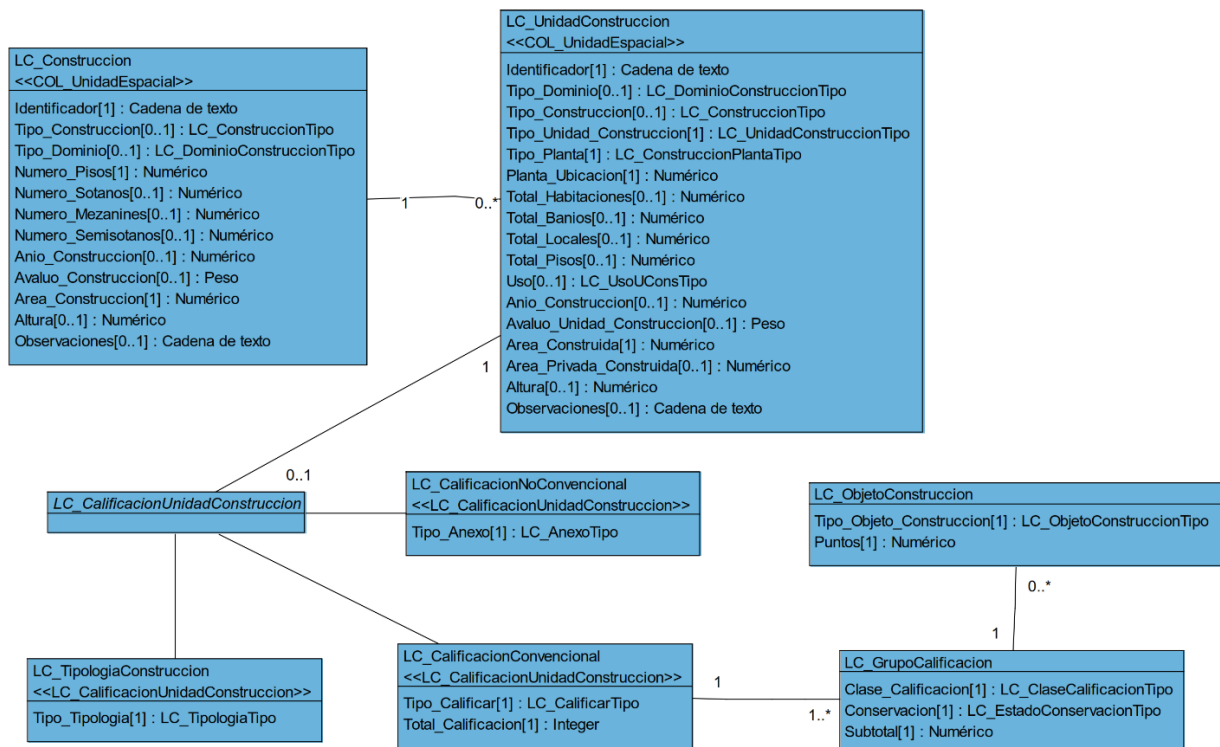


Ilustración 5. Modelo UML de Construcciones LADM-COL 1.0

nombre	fecha de modificacion	tipo
ataco_0606_2707.backup	22/08/2023 7:05 p. m.	Archivo BACKUP
Ataco_1107.xtf	2/10/2023 8:08 p. m.	Archivo XTF
FO_Final.xtf	29/09/2023 7:24 p. m.	Archivo XTF
nur_caceres_000_vf12_28062023_2.backup	22/08/2023 7:07 p. m.	Archivo BACKUP
V1_Urb_cabecera_001_vf.xtf	13/01/2023 1:49 p. m.	Archivo XTF
V1_Urb_guarumo_003_vf.xtf	13/01/2023 2:09 p. m.	Archivo XTF
V1_Urb_jardin_002_vf.xtf	13/01/2023 2:45 p. m.	Archivo XTF
V1_Urb_manizales_006_vf.xtf	13/01/2023 3:34 p. m.	Archivo XTF
V1_Urb_puertobelgica_005_vf.xtf	13/01/2023 3:56 p. m.	Archivo XTF
XTF_PLL_FINAL.xtf	29/09/2023 7:28 p. m.	Archivo XTF

Ilustración 6. Archivos XTF y Back up recopilados

Fuente de datos 2: Repositorios de fotografías de construcciones

Contienen fotografías de todo tipo relacionadas con construcciones, estas se almacenan en carpetas renombradas con un identificador único que hace referencia al predio, “Id operación”, donde recae cada construcción. Por otro lado, las fotografías están nombradas de tal manera que se logren relacionar con las unidades constructivas de la base de datos relacional, utilizando los identificadores “Id cons” y “Id uni”.

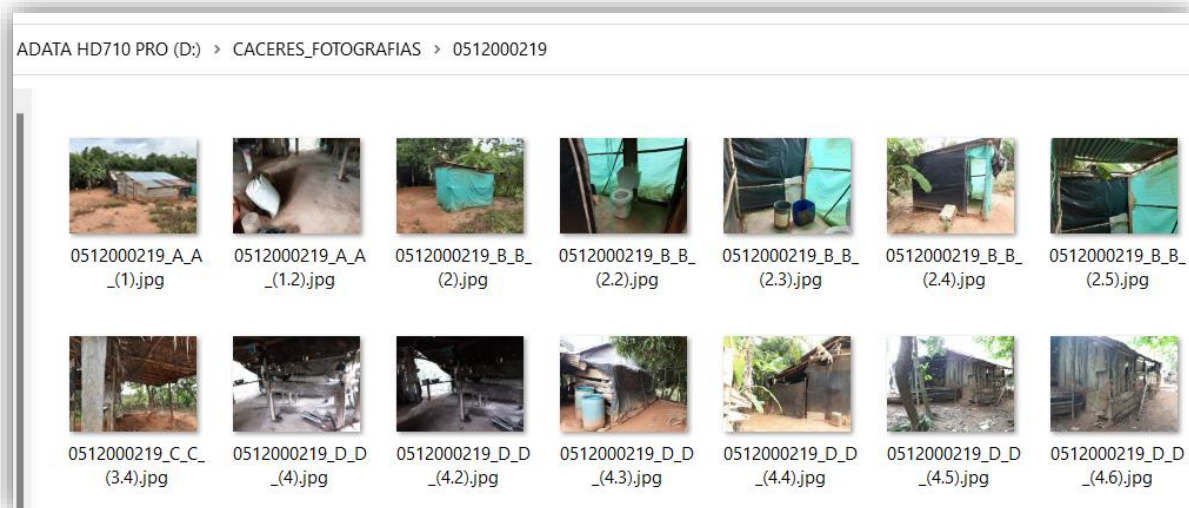


Ilustración 7. Repositorios de fotografías

Paso 2: Análisis exploratorio de datos y consolidación de datos

Paso 2.1: Análisis y consolidación de la fuente de datos 1

Una vez obtenidos los datos, se procedió a realizar un análisis exploratorio identificando las dos fuentes de información, las bases de datos relacionales y los repositorios de fotografías segmentadas para cuatro municipios: Ataco, Cáceres, Fuente de Oro y Puerto Lleras.

Para realizar el análisis exploratorio, fue necesario consolidar la información de construcciones de las diferentes bases de datos utilizando lenguaje SQL. Considerando la estructura relacional mostrada en la ilustración 5, se consolidaron los siguientes atributos:

t_id	Identificador único de Base de datos
XTF	Fuente de los datos
id_operacion	Identificador único de predio
id_cons	Identificador de construcción
id_uni	Identificador de Unidad de Construcción
tipo_unidad_construccion	Tipo de unidad constructiva
planta_ubicacion	Nivel de ubicación de la unidad
total_habitaciones	Cantidad de habitaciones en la unidad
total_banios	Cantidad de baños
total_locales	Cantidad de locales
total_pisos	Cantidad de pisos en la construcción
uso	Uso de la unidad
anio_construccion	Año de construcción
area_construida	Área construida
altura	Altura de la unidad
tipo_convencional	Tipo de unidad constructiva
total_calificacion	Puntaje de calificación detallada de la unidad
tipo_anexo	Clasificación de unidad no convencional o anexa
tipo_tipologia	Clasificación de unidad convencional no detallada

Tabla 2. Atributos relevantes de la base de datos

Obteniendo como un consolidado de 50988 construcciones, las cuales se distribuyen por municipio de la siguiente manera:

Municipio	Cantidad de Construcciones
ATACO	13142
CACERES RURAL	13564
CACERES URBANO	9848
FUENTE DE ORO Y PUERTO LLERAS	14434
<i>Total</i>	50988

Tabla 3. Distribución de construcciones por Municipio

Del consolidado destacan, por tipificar o clasificar las unidades según sus características constructivas, las siguientes variables:

- Tipo convencional y Total calificación

Estas dos variables son el resultado de la tipificación de las unidades constructivas convencionales de manera muy rigurosa considerando las tablas, LC objeto, LC grupo y LC calificación convencional del modelo LADM-COL, ver ilustración 5. Para poder determinarlas, es necesario tener acceso a información detallada de dependencias interiores de la construcción, sus materiales y su método de construcción.

- Tipo anexo

Esta variable es el resultado de la tipificación de las unidades constructivas no convencionales como, enramadas, galpones, establos, cocheras, silos, piscinas, tanques, entre otros. Su determinación es sencilla y se basa en la selección de dominios preestablecidos en una tabla y un atributo del modelo LADM-COL.

- Tipo tipología

Esta variable es el resultado de la tipificación de las unidades constructivas convencionales de manera sencilla o no detallada, considerando la tabla “LC Tipología Construcción” del modelo LADM-COL, ver ilustración 5. Se basa en la selección de dominios preestablecidos del modelo LADM-COL, por lo que no es necesario contar con información sobre dependencias interiores.

Dado lo anterior, es de interés saber por municipio cuantas calificaciones convencionales, no convencionales y por tipologías existen.

<i>Municipio</i>	Calificaciones Convencionales	Calificaciones No convencionales	Calificaciones por tipologías
ATACO	296	4834	8012
CÁCERES	5122	6772	1671
CÁCERES_URB	6685	3163	
FUENTE DE ORO Y PUERTO LLERAS	169	7391	6873
Total, general	12272	22160	16556

Tabla 4. Tipos de calificaciones según LADM-COL

La tabla 4 permitió tener un panorama claro frente al tipo de calificaciones empleadas para las unidades de construcción, siendo más representativas las calificaciones no convencionales con un 46.4 % del total. Estas calificaciones no convencionales se caracterizan por ser edificaciones que no cumplen con los métodos y materiales de construcción tradicionales, por lo que son difíciles de calificar de manera convencional.

La tipificación o clasificación de las unidades constructivas objeto de este trabajo, se centra en la calificación por tipologías, siendo esta la etiqueta objetivo representada por la variable “Tipo tipología”, la cual presenta un 32.4 % del consolidado de datos; cabe aclarar que es posible aumentar este porcentaje ya que la calificación convencional al ser detallada puede derivar en calificaciones por tipología, tal como se presenta a continuación:

Partiendo de las tipologías residenciales descritas en la Resolución 1040 de 2023 [13, p. 204] y la tabla de puntajes por tipología del manual de reconocimiento predial del IGAC [2, p. 95] se logró que 12272 construcciones con puntaje en “Total Calificación” migraran a tener registros de tipología en “Tipo tipología homologada”, este último atributo es un derivado de la combinación de todas aquellas calificaciones por tipología y las transformadas.

TIPO
1 (0 - 10)
2 (11 - 28)
3 (29 - 46)
4 (47 - 64)
5 (65 - 82)
6 (83 -100)

Tabla 5. Rango de puntaje por tipología [2, p. 95]

Se obtuvo como resultado un conjunto de datos consolidado para los cuatro municipios, los identificadores de cada unidad constructiva “Id_cons” y “Id_uni”, su identificador de predio relacionado “id_operacion”, atributos adicionales informativos y su respectiva variable objetivo “Tipología Homologada”.

En la ilustración 8 se señala en rojo y amarillo las variables de las cuales deriva la etiqueta final o variable objetivo, esta indica que de 50988 construcciones, se tienen 28827 con la etiqueta objetivo, lo que representa un 56.5% de los datos que son desarrollables en este proyecto.

```

rangeindex: 50988 entries, 0 to 50987
Data columns (total 22 columns):
#   Column                               Non-Null Count  Dtype
---  ---                               -
0   t_id                                  50988 non-null  int64
1   XTF                                   50988 non-null  object
2   id_operacion                         50988 non-null  int64
3   id_cons                              50988 non-null  object
4   id_uni                              50988 non-null  object
5   tipo_unidad_construccion            50988 non-null  object
6   planta_ubicacion                   50988 non-null  int64
7   total_habitaciones                 50987 non-null  float64
8   total_banios                       50988 non-null  int64
9   total_locales                      50988 non-null  int64
10  total_pisos                         50988 non-null  int64
11  uso                                 50988 non-null  object
12  anio_construccion                  50988 non-null  int64
13  area_construida                    50988 non-null  float64
14  altura                             50988 non-null  int64
15  tipo_convencional                   12272 non-null  object
16  total_calificacion                  12272 non-null  float64
17  tipo_anexo                          22161 non-null  object
18  tipo_tipologia                      16556 non-null  object
19  Tipologia_Homologada                28827 non-null  object
20  CARPETA                            35338 non-null  object
21  CANTIDAD_FTOS                      35338 non-null  float64
dtypes: float64(4), int64(8), object(10)

```

Ilustración 8. Análisis exploratorio del consolidado XTF

Paso 2.2: Análisis de fuente de datos 2

En el análisis exploratorio del repositorio de fotografías se pudo constatar que:

-No existe una estandarización frente a la extensión o formato de las fotografías siendo variable para cada municipio.

Municipio	Extensión	Total
ATACO	.heic	3
	.jpeg	2058
	.jpg	10122
	.pdf	115
	.png	56
CÁCERES	.2)jpg	2
	.jfif	20
	.jp	1
	.jpeg	1369
	.jpg	9571
	.pdf	1
	.png	231
CÁCERES URBANO	.jfif	3
	.jpeg	2
	.jpg	8015
	.png	3
FUENTE DE ORO Y PUERTO LLERAS	.jpeg	2778
	.jpg	5584
	.png	69

Tabla 6. Formato de fotografías

-Existe una variación considerable frente al tamaño en bytes de las fotografías por municipio, es decir que no existe un estándar claro sobre el peso o compresión de las fotografías.

Municipio	Promedio de Tamaño (Bytes)
ATACO_FOTOGRAFÍAS	2112897.307
CÁCERES_FOTOGRAFÍAS	2813721.889
CÁCERES_FOTOGRAFÍAS_URB	466547.163
FO_PLL_FOTOGRAFÍAS	2059732.798
Total general	1967629.053

Tabla 7. Peso o Compresión de las fotografías

-Existe una variación considerable frente a las dimensiones de las fotografías, esto hace referencia a la cantidad de pixeles y su distribución, donde la más común es la dimensión 3264 x 2448, sin embargo, existen 2598 dimensiones diferentes en el repositorio de fotografías.

<i>Dimensión</i>	<i>Total</i>
(3264, 2448)	10470
(1280, 960)	6223
(4000, 3000)	988
(1600, 1200)	825
(1280, 962)	790
(1600, 719)	782
(1200, 1600)	710
(1280, 720)	705
(4000, 1800)	641
(3264, 1836)	582
(1280, 800)	572
(3263, 2448)	499

Tabla 8. Dimensiones de las fotografías

Paso 3: Identificación de variables

El análisis exploratorio determinó algunas de las características de las fuentes de datos y atributos que pueden considerarse como relevantes para avanzar en el primer objetivo de este proyecto. Dado lo anterior, se procede a entrar en el detalle de la clasificación de las tipologías constructivas.

En el marco teórico, apartado 5.1.3 y la tabla 1, se describen las características que debe cumplir una unidad constructiva para clasificarse en la subcategorías o etiquetas de la tipología; de estas se destacan las siguientes variables:

- **Armazón:**
Estructura interna de un edificio o una construcción que proporciona soporte y estabilidad
- **Muros:**
Estructuras verticales que delimitan y protegen el espacio dentro de una construcción.
- **Cubierta:**
Elemento superior de una construcción que la protege de las inclemencias del tiempo, como la lluvia y el sol.
- **Cubrimiento de muros:**
Materiales y técnicas utilizados para revestir y proteger la superficie de los muros
- **Piso:**
Superficie horizontal sobre la que se camina dentro de una edificación.

Las anteriores variables y las características de cada una de ellas definen en conjunto la tipología, siendo estrictos respecto a lo normativo; es de mencionar que estas también son consideradas debido a que son variables de la ficha predial que podrían extraerse sin la necesidad de ingresar a las dependencias (Baño, Cocina, Habitaciones, entre otros) de la unidad constructiva o sin la identificación desde una fotografía de fachada.

A continuación, se muestra la distribución de las etiquetas de la variable “tipología homologada”, siendo evidente que las clases Residencial y sus subtipos 1A, 1B, 2A, 2B y 3A son las más comunes dentro del consolidado de base de datos obtenido para cuatro municipios.

Tipología Homologada	Cuenta de Tipología Homologada
Residencial.Tipo_5A	7
Residencial.Tipo_4B	4
Residencial.Tipo_4A	99
Residencial.Tipo_3B	305
Residencial.Tipo_3A	2712
Residencial.Tipo_2B	4876
Residencial.Tipo_2A	6961
Residencial.Tipo_1B	8271
Residencial.Tipo_1A	3876
Institucional.Tipo_4A	3
Institucional.Tipo_3B	66
Institucional.Tipo_3A	363
Institucional.Tipo_2B	183
Institucional.Tipo_2A	121
Institucional.Tipo_1B	17
Institucional.Tipo_1A	7
Industrial.Mediana_2	7
Industrial.Mediana_1	19
Industrial.Liviana_2	62
Industrial.Liviana_1	37
Comercial.Sectorial_2	25
Comercial.Sectorial_1	509
Comercial.Especializado_1	1
Comercial.Barrial_2	198
Comercial.Barrial_1	98
Total general	28827

Tabla 9. Distribución de tipologías en el conjunto de datos

Dado lo anterior, las tipologías residenciales 1,2 y 3 son aquellas que cuentan con 93.66 % de los datos por lo cual fueron las utilizadas para el desarrollo de este proyecto.

Paso 4: Validación de correspondencia

Para cumplir con los objetivos planteados, es necesario tener suficientes ejemplos de fotografías etiquetadas con su correspondiente tipología, tal como se analizó en los antecedentes y según lo propuesto por García et al. [7]. Por esta razón se realizó el proceso de identificar la tipología correspondiente de cada fotografía, teniendo en cuenta que las fotografías están nombradas con la estructura que muestra la ecuación 1.

$$ID_{OPERACIÓN} + ID_{CONSTRUCCIÓN} + ID_{UNIDAD} + CONSECUTIVO = NOMBRE FOTOGRAFÍA$$

Ecuación 1. Nombre estándar de fotografías

Considerando también que dentro del conjunto de datos consolidados se cuenta con los mismos atributos, se realizó la correspondencia para poder renombrar o clasificar las fotografías con su tipología, de la siguiente manera:

$$ID_{OPERACIÓN} + ID_{CONSTRUCCIÓN} + ID_{UNIDAD} + Tipología + CONSECUTIVO = NOMBRE FOTOGRAFÍA$$

Ecuación 2. Nombre estándar con tipología

Como resultado de lo anterior, no fue posible asociar 15763 unidades constructivas de las 28827 disponibles con la etiqueta "Tipología Homologada" a su correspondiente fotografía, debido a la ausencia de fotografías relacionadas con el "ID Operación". Por esta razón, se pasó de tener 26696 unidades constructivas con tipología residencial disponibles para el proyecto a descontar 7747, que no tenían fotografía, dando un consolidado parcial de 18949 unidades constructivas con fotografía, pertenecientes a las tipologías residenciales de interés.

De la validación de correspondencia con las fotografías versus la base de datos relacional, se evidenciaron las siguientes falencias en el consolidado de las fotografías:

- Ausencia de fotografías para 15763 unidades constructivas

Esta falencia generó que la información de la base de datos que no tiene fotografías no fuese útil para los fines de este proyecto.

- Fotografías que no cumplen con la estructura mostrada en la ecuación 1

Nombres de ejemplo	Conforme
0512002021_L_L_(1.4).jpg	SI
505770003000000050152000000000_E1.jpg	NO
8271_730670001000000130027000000000_II.jpg	NO
10126_VIII.jpeg	NO
0512007208_4.jpg	NO

Tabla 10. Ejemplo de conformidad en nombre de fotografía

La no conformidad con lo establecido en la ecuación 1 generó la necesidad de organizar, mediante filtros y consultas, los nombres de un gran número de fotografías. De las 55599 fotografías disponibles, en 21037 casos fue imposible relacionar los identificadores de operación, construcción o unidad que permitieran asociarlas con alguna tipología.

Lo anterior género que se tuviese un potencial de 32535 fotografías relacionadas con tipologías residenciales 1, 2 y 3.

- Múltiples fotografías para la misma unidad constructiva, ver ilustración 7

En los repositorios se evidenció que para cada unidad constructiva existía más de una fotografía, donde se incluyen fotografías de fachadas, panorámicas, baños, cocinas y dependencias internas. Lo anterior se presentó como un reto debido a que el proyecto necesitaba una fotografía de fachada por cada unidad constructiva y su correspondiente tipología.

Paso 5: Depuración, transformación y calidad de datos

En el proceso de depuración se buscó abordar primero la estandarización de las fotografías en cuanto a:

- Dimensiones:

Se probaron varias combinaciones de tamaño 150 x 150, 250 x 250 y 500 x 500, concluyendo que 500 x 500 tenía una relación de reducción adecuada, evitando la deformación y conservando los detalles de las fotografías.



Ilustración 9.1. Fotografía 150 x 150

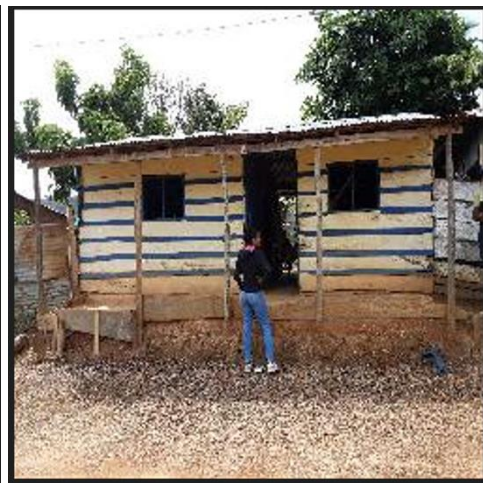


Ilustración 9.2. Fotografía 250 x 250



Ilustración 9.3. Fotografía 500 x 500

- Formatos:

Se seleccionó el formato JPG debido a la compatibilidad con diferentes motores y librerías de procesamiento de imágenes.

Los dos ítems anteriores se procesaron utilizando JUPYTER NOTEBOOK y sus librerías “OS” para el manejo de directorios, “PANDAS” para el manejo de conjuntos de datos y “PIL IMAGE” para el manejo y transformación de imágenes.

- Múltiples fotografías para una unidad constructiva

Para solucionar la situación se construyó un modelo de clasificación que discriminase entre fotografías de fachadas y no fachadas, apoyado previamente de la depuración manual del repositorio, eliminando fotografías con obstáculos, desenfocadas y alejadas del objetivo.

Es importante señalar que los procesos de captura de fotografías no se encontraban estandarizados en los cuatro municipios, lo que generó dificultades para vincular de manera efectiva la información con la base de datos relacional y su correspondiente tipología. En los siguientes apartados se explica la funcionalidad de un modelo de clasificación de fachadas, no fachadas y la utilidad que presenta en términos de calidad.

Paso 6: Aplicación de estrategia de recolección y procesamiento de datos

La selección de la muestra de entrenamiento y evaluación debe considerar una fotografía por unidad constructiva, para el caso de este proyecto, las fotografías de baños, cocinas, dependencias interiores y con mala calidad, asociadas a las 18949 unidades constructivas, deben ser descartadas, por lo cual fue necesario construir un clasificador de fachadas.

Para la construcción del clasificador se tomaron las fotografías del municipio de Cáceres área urbana, el cual dentro del nombre de cada fotografía adicional a la estructura de la ecuación 1, establecía si se trataba de una Fachada, Baño, Cocina, Cubierta, Sala u Otros, tal como se muestra en la siguiente ilustración:



Ilustración 10. Fotografías de Fachada Etiquetadas

Dado lo anterior se generó una clasificación binaria “Fachada: 0” y “No Fachada:1” de todo el repositorio urbano que contaba con etiquetas, obteniendo:

Etiqueta	Cantidad de Fotografías
Fachada: 0	9855
No Fachada: 1	17819
Total	27674

Tabla 11. Repositorio con etiquetas de fachadas original

Fue necesario realizar un proceso de depuración de las fotografías pre etiquetadas, debido a que se presentaban problemas de calidad como obstáculos, errores de etiquetados y objetivos muy alejados o desenfocados, tal como se muestra en la ilustración 11.

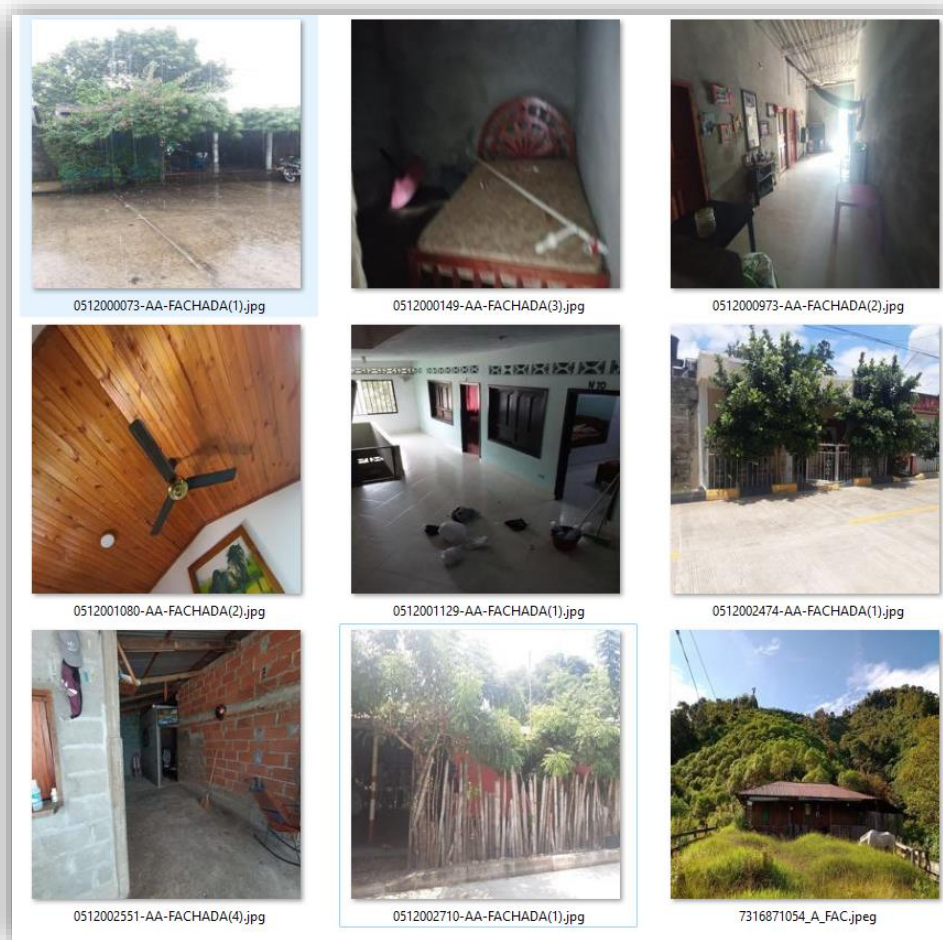


Ilustración 11. Fotografías con error en etiqueta

Dado lo anterior, se optó por realizar un proceso de depuración de las fotografías con el fin de entrenar el modelo con datos de mejor calidad, obteniendo el repositorio final de entrenamiento para el clasificador de fachadas, tal como se muestran en la siguiente tabla:

Etiqueta	Cantidad de Fotografías
Fachada: 0	2587
No Fachada: 1	16420
Total	19007

Tabla 12. Repositorio con etiqueta de fachada depurada

Considerando que se contaba con 19007 fotografías preclasificadas en fachada y no fachada, se procedió a realizar la construcción del modelo utilizando las siguientes herramientas y técnicas:

- **Lenguaje de Programación:** Python 3
- **Compilador de Lenguaje:** Google Colab
- **Librerías:**
 - “TensorFlow” para construir y entrenar modelos de aprendizaje profundo.
 - “Keras” para facilitar la construcción de modelos.
 - “TensorFlow Datasets” para acceder a conjuntos de datos predefinidos.
 - “ImageDataGenerator” para realizar aumentos de datos en imágenes
 - “Matplotlib” para visualización de datos.
 - “OS” y “SHUTIL” para manipulación de archivos y directorios.
 - “NumPy” para manipulación de datos en formato de matriz.
 - “Seaborn” para visualizaciones estadísticas.
 - “Scikit-learn” para evaluación del modelo y manejo de métricas de desempeño.
- **Técnicas de inteligencia artificial:** Redes Neuronales Convolucionales y “Transfer Learning”
- **Datos preclasificados:** Fotografías etiquetadas en fachadas y no fachadas.

Modelo base: Modelo VGG16 descargado usando KERAS y con parámetros modificados

```
base_model = tf.keras.applications.VGG16(weights='imagenet', include_top=False, input_shape=(500,500,3))
base_model.trainable = False
```

Ilustración 12. Modelo Base, clasificador de fachadas

Modelo Convolucional adicional: Al modelo base se adicionó

- 1 capa oculta plana
- 1 capa densa con 250 neuronas y función de activación “Relu”
- 1 parámetro de configuración “drop out” de 0.5
- 1 capa densa con 3 neuronas y función de activación “sigmoid”

```
model_DA2 = keras.Sequential([
    base_model,
    keras.layers.Flatten(),
    keras.layers.Dense(250, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(1, activation='sigmoid')
])
```

Ilustración 13. Modelo Convolucional clasificador de fachadas

Entrenamiento y validación:

1810 fotografías de entrenamiento y 777 de validación por cada etiqueta “Fachada” y “No Fachada”.

Posterior a la construcción de modelo de clasificación de fachadas, fue posible consolidar una estrategia completa de captura y recolección de imágenes dentro de un conjunto de datos relacionado con construcciones que deber seguir el siguiente flujo, con el fin de preparar la información para los siguientes pasos:

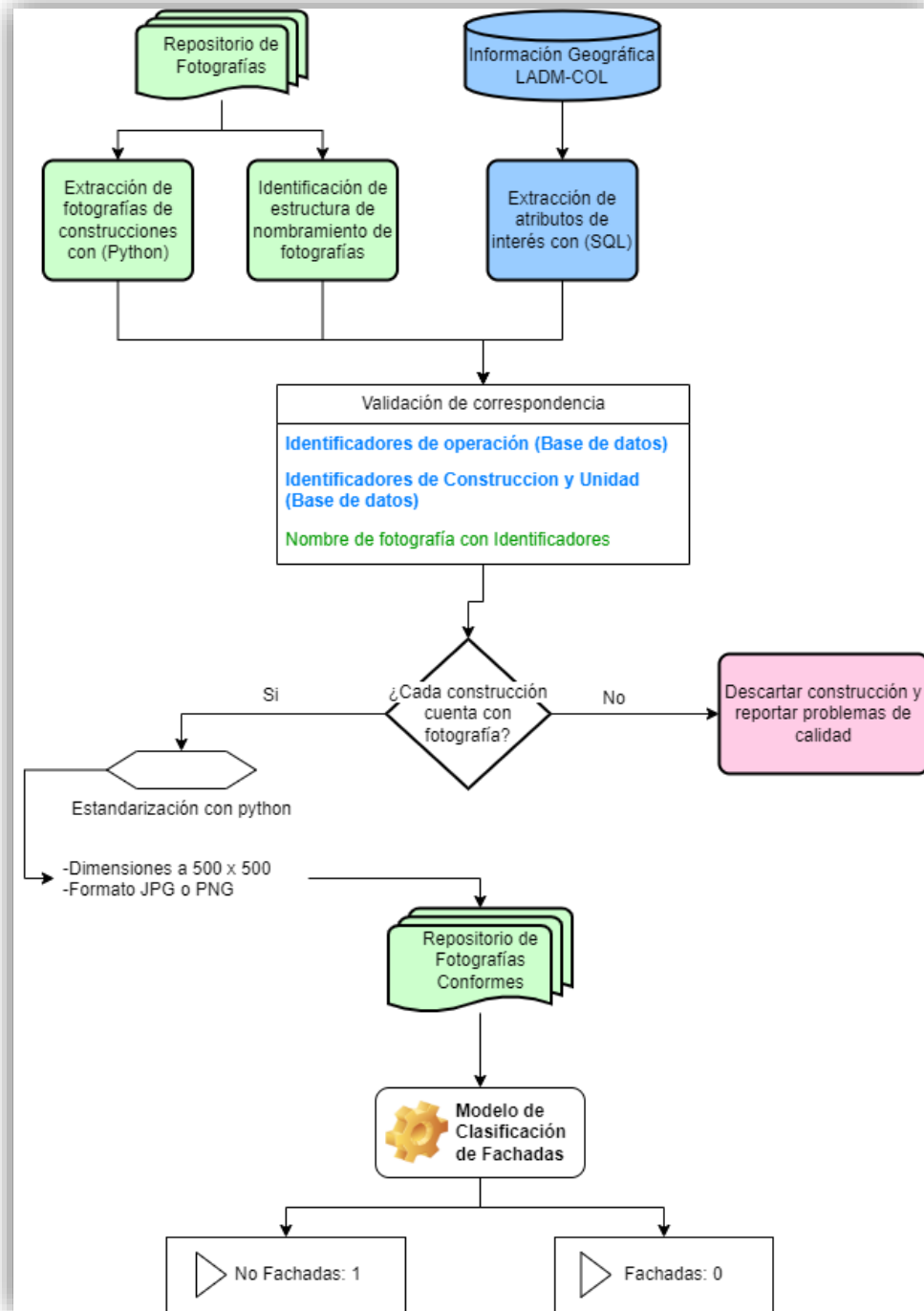


Ilustración 14. Estrategia de recolección y procesamiento

Modelo de clasificación de imágenes a partir de técnicas de aprendizaje automático.

Paso 7: Selección de técnicas adecuadas

Basado en el marco teórico del presente proyecto, las recomendaciones de otros proyectos estudiados en los antecedentes y la experiencia positiva adquirida en la construcción del modelo de clasificación de fachadas, se tomó como técnica principal el uso de Redes Neuronales Convolucionales - Convolutional Neural Network (CNN). Es importante señalar que la capacidad computacional y la disponibilidad de fotografías pueden ser limitantes para desarrollar modelos de calidad, en algunas investigaciones relacionadas con clasificación de imágenes, se aplica una técnica llamada “Transfer Learning”. En [17, p. 74] Pérez et al, explica que dicha técnica reutiliza la arquitectura y pesos de una red existente pre entrenada con grandes volúmenes de información y validadas o reconocidas en diferentes investigaciones, donde se demuestra que dichos modelos tienen niveles de exactitud altos como por ejemplo VGG16 y VGG19, lo anterior abre las posibilidades al modificar los parámetros del modelo pre entrenado para que pueda adaptarse a un caso en específico.

Debido a la posibilidad de usar modelos pre entrenados y la amplia gama de documentación sobre como modificar sus parámetros, se optó por utilizar el modelo VGG16, el cual cuenta con las siguientes especificaciones:

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
Xception	88	79.00%	94.50%	22.9M	81	109.4	8.1
VGG16	528	71.30%	90.10%	138.4M	16	69.5	4.2
VGG19	549	71.30%	90.00%	143.7M	19	84.8	4.4

Tabla 13. Especificaciones de modelos pre entrenados, tomado de [18].

Frente a los demás modelos evidenciados en el cuadro anterior, el modelo VGG16 presenta tiempos de procesamientos mucho más rápidos, por lo que este fue el seleccionado.

Paso 8: Construcción y entrenamiento del modelo

Previo a la construcción del modelo fue importante aplicar el flujo establecido en la ilustración 14, con el objetivo de obtener un repositorio de fotografías que cuenten con:

1. Una etiqueta o tipología constructiva definida por la base de datos LADM-COL.
2. Estandarización en dimensiones y formatos
3. Fotografías únicamente de fachadas

Respecto a los puntos 1 y 2, se contaron 18,949 unidades constructivas de la base de datos LADM-COL, que presentaban validación de correspondencia con 25,222 fotografías. Esto se debió a que, por cada unidad constructiva, existían múltiples fotografías, la mayoría de las dependencias interiores. Dado lo anterior, se aplicó el punto 3 utilizando el clasificador de fachadas, lo que resultó en un repositorio de 7,141 fotografías de fachadas que cuentan con una tipología definida en el modelo LADM-COL.

La distribución de fotografías por tipologías para el repositorio generado se presenta en la siguiente tabla:

Tipología General	Total Agrupado	Tipología LADM-COL	Cantidad de Fotografías de Fachada
(Residencial) Tipo 1	2032	(Residencial) Tipo 1A	412
		(Residencial) Tipo 1B	1620
(Residencial) Tipo 2	3785	(Residencial) Tipo 2A	2020
		(Residencial) Tipo 2B	1765
(Residencial) Tipo 3	1313	(Residencial) Tipo 3A	1184
		(Residencial) Tipo 3B	129
(Residencial) Tipo 4	11	(Residencial) Tipo 4A	8
		(Residencial) Tipo 4B	3
Total de Fotografías de Fachada			7141

Tabla 14. Repositorio de fotografías de fachada con etiqueta

Sobre le repositorio anterior, se realizó un muestreo aleatorio con el fin de identificar problemas de calidad en el etiquetado de fotografía de fachas. Del ejercicio se pudo evidenciar que existían problemas importantes con el etiquetado de fotografías ya que no seguían lo establecido en la Resolución 1040 del 2023, ver tabla 1. A continuación, en la ilustración 15 se muestran fotografías etiquetadas con tipología 1A que se caracterizan con cubiertas y muros en materiales de desecho, sin embargo, las fotografías evidencian una tipología diferente.



Ilustración 15. Ejemplos de fotografías con etiqueta incorrecta

Debido a los inconvenientes en el etiquetado, se realizó un proceso de depuración que consistía en etiquetar correctamente las fotografías, este fue un proceso puntual, manual y basado en el cumplimiento de las variables

descritas en la tabla 1; en el proceso se eliminaron también fotografía repetidas, dando como resultado la distribución de fotografías por tipologías depuradas de la tabla 15.

Tipología General	Total Agrupado	Tipología LADM-COL	Cantidad de Fotografías de Fachada
(Residencial) Tipo 1	901	(Residencial) Tipo 1A	170
		(Residencial) Tipo 1B	731
(Residencial) Tipo 2	922	(Residencial) Tipo 2A	587
		(Residencial) Tipo 2B	335
(Residencial) Tipo 3	4701	(Residencial) Tipo 3A	1142
		(Residencial) Tipo 3B	3559
(Residencial) Tipo 4	99	(Residencial) Tipo 4A	99
Total de Fotografías de Fachada			6623

Tabla 15. Repositorio de fotografías de fachada con etiqueta depurada

En la tabla anterior se evidencia que el 71 % de los datos corresponde a la tipología 3, lo que refleja la tipificación constructiva más común de las áreas rurales de donde se obtuvieron los datos.

Una vez seleccionadas las técnicas y depurado el repositorio, se selecciona el software y las librerías, ya que estas permiten ejecutar las técnicas mencionadas en el apartado anterior.

- **Lenguaje de Programación:** Python 3
- **Compilador de Lenguaje:** Google Colab
- **Librerías:**
 - “TensorFlow” para construir y entrenar modelos de aprendizaje profundo
 - “Keras” para facilitar la construcción de modelos
 - “TensorFlow Datasets” para acceder a conjuntos de datos predefinidos
 - “ImageDataGenerator” para realizar aumentos de datos en imágenes
 - “Matplotlib” para visualización de datos
 - “OS” y “SHUTIL” para manipulación de archivos y directorios
 - “NumPy” para manipulación de datos en formato de matriz
 - “Seaborn” para visualizaciones estadísticas
 - “Scikit-learn” para evaluación del modelo y manejo de métricas de desempeño
- **Técnicas de inteligencia artificial:** Redes Neuronales Convolucionales y “Transfer Learning”

Tomando como base el repositorio depurado, tabla 15, se evidenció que no existían suficientes datos para realizar un proceso de entrenamiento con las tipologías LADM-COL detalladas, por lo cual se optó por realizar un proceso de agrupación en tres categorías o “Tipologías Generales” y descartar la tipología número 4.

En el proceso de entrenamiento se probaron las siguientes estructuras:

Estructura 1:

Modelo base: Modelo VGG16 descargado usando KERAS y con parámetros modificados.

```
base_model = tf.keras.applications.VGG16(weights='imagenet', include_top=False, input_shape=(500,500,3))
base_model.trainable = False
```

Ilustración 16. Modelo Base, clasificador de tipologías

Modelo Convolucional adicional: Al modelo base se le adicionó

1 capa oculta plana

1 capa densa con 250 neuronas y función de activación “Relu”

1 parámetro de configuración “drop out” de 0.5

1 capa densa con 3 neuronas y función de activación “softmax”

```
[31]
model_DA2 = keras.Sequential([
    base_model,
    keras.layers.Flatten(),
    keras.layers.Dense(250, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(3, activation='softmax')
])
```

Ilustración 17. Modelo convolucional, clasificador de tipologías Estructura 1

Estructura 2:

Modelo base: Modelo VGG16 descargado usando KERAS y con parámetros modificados.

Modelo Convolucional adicional: Al modelo base se le adicionó

1 capa oculta plana

1 capa densa con 100 neuronas y función de activación “Relu”

1 parámetro de configuración “drop out” de 0.5

1 capa densa con 3 neuronas y función de activación “softmax”

```
[46]
model_DA2 = keras.Sequential([
    base_model,
    keras.layers.Flatten(),
    keras.layers.Dense(100, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(3, activation='softmax')
])
```

Ilustración 18. Modelo convolucional, clasificador de tipologías Estructura 2

Estructura 3:

Modelo base: Modelo VGG16 descargado usando KERAS y con parámetros modificados.

Modelo Convolutacional adicional: Al modelo base se le adicionó

1 capa oculta plana

1 capa densa con 100 neuronas y función de activación “Relu”

1 capa densa con 50 neuronas y función de activación “Relu”

1 parámetro de configuración “drop out” de 0.5

1 capa densa con 3 neuronas y función de activación “softmax”

```
[105]
model_DA2 = keras.Sequential([
    base_model,
    keras.layers.Flatten(),
    keras.layers.Dense(100, activation='relu'),
    keras.layers.Dense(50, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(3, activation='softmax')
])
```

Ilustración 19. Modelo convolutacional, clasificador de tipologías Estructura 3

Estructura 4:

Modelo base: Sin modelo base, no se aplicó “transfer learning”.

Modelo Convolutacional adicional:

1 capa oculta plana

1 capa densa con 100 neuronas y función de activación “Relu”

1 capa densa con 50 neuronas y función de activación “Relu”

1 parámetro de configuración “drop out” de 0.5

1 capa densa con 3 neuronas y función de activación “softmax”

```
model_DA2 = keras.Sequential([
    keras.layers.Flatten(input_shape=(150,150,3)),
    keras.layers.Dense(100, activation='relu'),
    keras.layers.Dense(50, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(3, activation='softmax')
])
```

Ilustración 20. Modelo convolutacional, clasificador de tipologías Estructura 4

Estructura 5:

Modelo base: Sin modelo base, no se aplicó “transfer learning”.

Modelo Convolutacional adicional:

1 capa oculta plana

1 capa densa con 250 neuronas y función de activación “Relu”

1 capa densa con 150 neuronas y función de activación “Relu”

1 capa densa con 100 neuronas y función de activación “Relu”

1 parámetro de configuración “drop out” de 0.5

1 capa densa con 3 neuronas y función de activación “softmax”

```
model_DA2 = keras.Sequential([
    keras.layers.Flatten(input_shape=(150,150,3)),
    keras.layers.Dense(250, activation='relu'),
    keras.layers.Dense(150, activation='relu'),
    keras.layers.Dense(100, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(3, activation='softmax')
])
```

Ilustración 21. Modelo convolutacional, clasificador de tipologías Estructura 5

Estructura 6:

Modelo base: Modelo VGG16 descargado usando KERAS y con parámetros modificados

Modelo Convolutacional adicional: Al modelo base se le adicionó

1 capa oculta plana

1 capa densa con 250 neuronas y función de activación “Relu”

1 capa densa con 150 neuronas y función de activación “Relu”

1 capa densa con 100 neuronas y función de activación “Relu”

1 parámetro de configuración “drop out” de 0.5

1 capa densa con 3 neuronas y función de activación “softmax”

```
model_DA2 = keras.Sequential([
    base_model,
    keras.layers.Flatten(),
    keras.layers.Dense(250, activation='relu'),
    keras.layers.Dense(150, activation='relu'),
    keras.layers.Dense(100, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(3, activation='softmax')
])
```

Ilustración 22. Modelo convolutacional, clasificador de tipologías Estructura 6

Entrenamiento y validación: En todas las estructuras se utilizó la misma distribución.

784 fotografías de entrenamiento y 116 de validación por cada tipología general 1,2 y 3.

Para el entrenamiento del modelo se trabajaron alrededor de 14 configuraciones diferentes, donde se variaron la estructura, cantidad de épocas, el tamaño de las imágenes y tamaños de lotes; en dichas configuraciones se obtuvieron limitaciones importantes en los tiempos de procesamiento y capacidad de máquina, donde algunos procesos no pudieron culminar con el entrenamiento ya que superaban la capacidad de 16 GB de RAM. Para los procesos que finalizaron correctamente se obtuvieron resultados funcionales, los cuales se analizarán más adelante.

4.2 RESULTADOS

Modelo Clasificador de Fachadas

Los primeros resultados del modelo de clasificación de fachada tuvieron indicadores por debajo del 50% en las métricas de evaluación “Precision”, “Recall”, “F1-score” y “Accuracy”, el problema se debía a que existían fotografías con problemas de etiquetado, fotografías tomadas en entornos con muchos obstáculos y muy alejadas como se muestra en la ilustración 11. Luego de la depuración del repositorio se logró un total de 19007 fotografías preclasificadas y con calidad.

Se tomó un 70 % de los datos para entrenamiento del modelo balanceando las dos etiquetas, para un total de 1810 fotografías de entrenamiento y 777 de evaluación interna, arrojando un “Accuracy” superior al 90%.

Para la evaluación del modelo entrenado se utilizaron las métricas de “Precision”, “Recall”, “F1-score”, “Accuracy” y se realizó la evaluación sobre la totalidad del repositorio depurado de la tabla 12, con 19007 fotografías, obteniendo las siguientes métricas de desempeño:

```
print(metrics.classification_report(y_test,y_pred, digits=4))
```

	precision	recall	f1-score	support
0	0.9573	0.8585	0.9052	2587
1	0.9781	0.9940	0.9860	16420
accuracy			0.9755	19007
macro avg	0.9677	0.9262	0.9456	19007
weighted avg	0.9752	0.9755	0.9750	19007

Ilustración 23. Métricas de desempeño del modelo

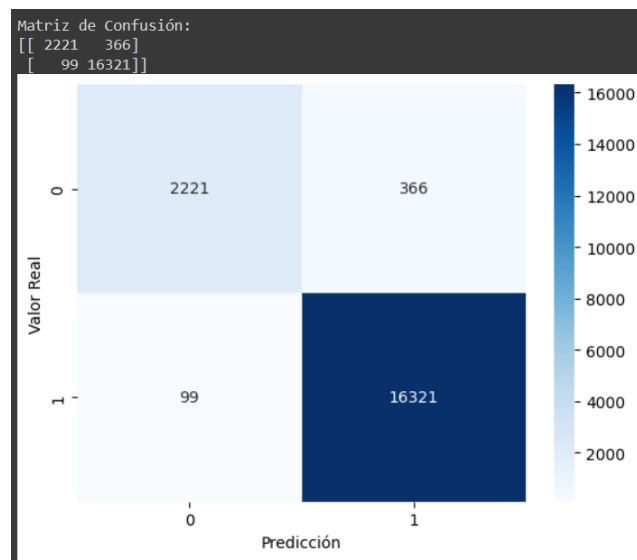


Ilustración 24. Matriz de confusión del modelo

Modelo Clasificador de Tipologías

El repositorio de fotografías depurado, siguiendo la estrategia de recolección y procesamiento y con sus correspondientes etiquetas de tipologías, fue de 6524 sin considerar la tipología 4, donde luego de balancear las etiquetas, 785 fotografías fueron de entrenamiento y 116 de validación interna por cada tipología general 1, 2 y 3.

Debido a que se establecieron modelos con 1 capa densa con 3 neuronas y función de activación “softmax” como salida, las predicciones del modelo fueron una matriz de 348 filas (116 fotografías de validación por cada tipología general) y 3 columnas que representan la probabilidad de pertenecer a la tipología 0, 1 o 2, que representan la tipología general 1, 2 y 3 respectivamente.

Como resultado del proceso de entrenamiento se obtuvieron 14 modelos funcionales los cuales se describen en la siguiente tabla:

Modelo funcional	Cantidad de épocas	Tamaño de imágenes	Tamaño de lote	Tiempo de procesamiento (Horas)	“Accuracy” en entrenamiento	Estructura utilizada
1	10	500 x 500	5	0.69	0.601	1
2	20	500 x 500	5	1.39	0.6156	1
3	40	500 x 500	5	2.78	0.6302	1
4	10	250 x 250	10	0.47	0.6293	1
5	40	250 x 250	10	1.89	0.6695	1
6	100	250 x 250	10	4.72	0.6839	1
7	200	250 x 250	10	9.44	0.681	1
8	400	250 x 250	10	19.5	0.729	1
9	100	150 x 150	10	0.603	0.704	2
10	400	150 x 150	10	8.65	0.637	2
11	100	150 x 150	10	0.662	0.660	3
12	10	150 x 150	10	0.060	0.333	4
13	10	150 x 150	10	0.062	0.333	5
14	100	150 x 150	10	0.665	0.660	6

Tabla 16. Puntuación de modelo convolucionales funcionales

En la tabla anterior se presentan las diferentes configuraciones de estructura, épocas, tamaño de imagen y lote, con su respectiva métrica “Accuracy”. Para cada estructura se probaron varias opciones y en la tabla se consolidaron aquellas que dieron mejor resultado. Se destacan por su mal desempeño los modelos 12 y 13, los cuales no usaron “Transfer Learning” como técnica base. Por el contrario, destacan de manera positiva los modelos 7, 8, 9 y 10 por su comportamiento y resultados durante el entrenamiento.

En la tabla anterior se evidencia que la métrica “Accuracy” mejoró conforme al aumento de la cantidad de épocas de procesamiento para la estructura 1, se redujo el tamaño de imágenes a 250x250 y se utilizó un tamaño de lote más grande, lo anterior se debe a que procesar tamaños de imágenes de 500x500 con más de 40 épocas y tamaños de lotes pequeños implica un tiempo de procesamiento más largo y un recurso de maquina no disponible.

Se opto por reducir el tamaño de las imágenes a 150 x 150 y probar otras estructuras, siendo la estructura 2 la que mejores resultados obtuvo con el modelo funcional 9 para dicho tamaño.

En el anexo de este proyecto, se muestran las métricas de desempeño, la matriz de confusión y las gráficas resultantes de los entrenamientos de los modelos funcionales del 1 al 14. Gráficamente se puede evidenciar que los modelos del 1 al 7, 9, 12 y 13 tienen una tendencia lineal en las líneas de pérdida y precisión en el entrenamiento, según Pérez [19], esto puede ser un indicio de sub ajuste de los modelos, lo cual evidencia que no se tiene la flexibilidad suficiente y es muy general respecto a los datos, por otro los modelos 8 y 10 tiene un comportamiento más flexible y natural en el entrenamiento, aumentando la precisión y disminuyendo la pérdida conforme aumentan las épocas.

De los modelos funcionales 8 y 9 no se evidencia sobre ajuste, ya que este se caracteriza por valores de precisión muy cercanos a 1 o 100%, curvas de validación muy distantes y valores de pérdida que no disminuyen conforme avanza el entrenamiento. Es importante mencionar que existen picos y variaciones importantes en las curvas de validación, en [19], la autora analiza esta situación y busca suavizar las curvas para el caso estudiado, sin embargo, tras probar varios métodos no se consigue una mejora significativa, por lo cual opto por cambiar la base de datos y entrenar con los mismos parámetros, obteniendo mejores resultados gráficos pero peores resultados en métricas de desempeño y matriz de confusión, llegando a la conclusión que los picos pueden darse debido a la complejidad del modelo y la variabilidad de características de los datos, siendo el caso de este proyecto, que utiliza un modelo VGG16 con millones de parámetros en su entrenamiento y fotografías de fachadas muy diversas.

A partir de la tabla 16 se pudo evidenciar que los mejores modelos en cuanto la métrica “Accuracy” en el entrenamiento fueron el 8 y el 9, dicha métrica evaluó a 116 fotografías por cada clase 1, 2 y 3, en total 348, debido a que se trataba de la muestra de evaluación; sin embargo, para validar con mayor certeza cuál de ellos podría tener mejores métricas de desempeño y evaluar datos que no haya entrenado el modelo, se realizó la predicción sobre el universo de fotografías etiquetadas con 6524, ver tabla 15, descontando la tipología general 4, obteniendo mejores resultados para el modelo funcional 8, con una precisión de 83.57 %, frente al 76.99 % del modelo 9.

En las siguientes ilustraciones se pueden ver a detalle las métricas de desempeño evaluadas, donde se resalta que la métrica f1-score para la tipología 3, alcanzan valores de 88.42% y 84.69%, para la tipología 1 valores de 84.87% y 75.81%, y para la tipología 2 valores de 66.69% y 53.61%; en esta última es evidente que existe una dificultad para diferenciarse de la tipología 3.

	precision	recall	f1-score	support
0	0.7844	0.9245	0.8487	901
1	0.5173	0.9382	0.6669	922
2	0.9995	0.7986	0.8842	4701
accuracy			0.8357	6524
macro avg	0.7641	0.8871	0.8000	6524
weighted avg	0.8952	0.8357	0.8486	6524

	precision	recall	f1-score	support
0	0.6805	0.8557	0.7581	901
1	0.4111	0.7701	0.5361	922
2	0.9667	0.7535	0.8469	4701
accuracy			0.7699	6524
macro avg	0.6861	0.7931	0.7137	6524
weighted avg	0.8487	0.7699	0.7907	6524

Ilustración 25. Métricas de desempeño globales modelo 8 y modelo 9

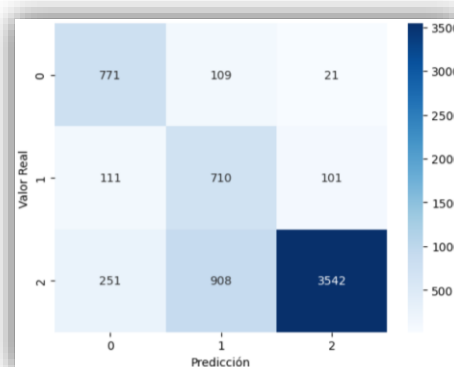
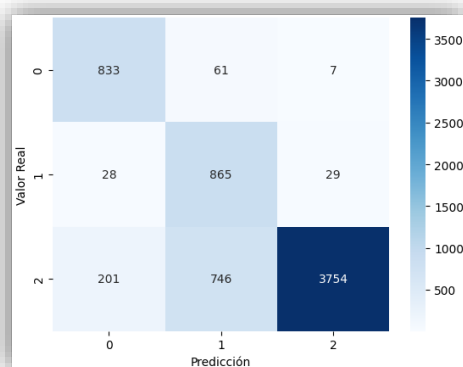


Ilustración 26. Matriz de confusión global modelo 8 y modelo 9

En todos los modelos funcionales probados se evidenció que la tipología general 2, presentó más dificultades en todas las ocasiones para generar resultados satisfactorios, por lo cual se procedió a realizar tres modelos con la estructura 1, usada en el modelo funcional 8, con una neurona de salida, omitiendo primero la tipología 1, luego la tipología 2 y finalmente la tipología 3. Para ver el comportamiento de las métricas de evaluación en cada caso, en la siguiente ilustración se muestra la estructura del modelo de prueba:

```

model_DA2 = keras.Sequential([
    base_model,
    keras.layers.Flatten(),
    keras.layers.Dense(250, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(1, activation='sigmoid')
])
  
```

Ilustración 27. Modelo convolucional, estructura de pruebas

Modelo de prueba 1, se elimina tipología 1: Con los mismos datos de entrenamiento, pero omitiendo la tipología 1, se realizó el modelo de clasificación de tipologías entre la tipología 2 y 3.

Siendo la tipología 2 (0) y la tipología 3 (1), se evidencia que la métrica “F1-score” es baja para la tipología 2, y al analizar la matriz de confusión, 1181 fotografías fueron clasificadas con la tipología 2, cuando realmente eran tipología 3.

	precision	recall	f1-score	support
0	0.4330	0.9783	0.6003	922
1	0.9944	0.7488	0.8543	4701
accuracy			0.7864	5623
macro avg	0.7137	0.8635	0.7273	5623
weighted avg	0.9023	0.7864	0.8126	5623

Ilustración 28. Métricas de desempeño, modelo de prueba 1

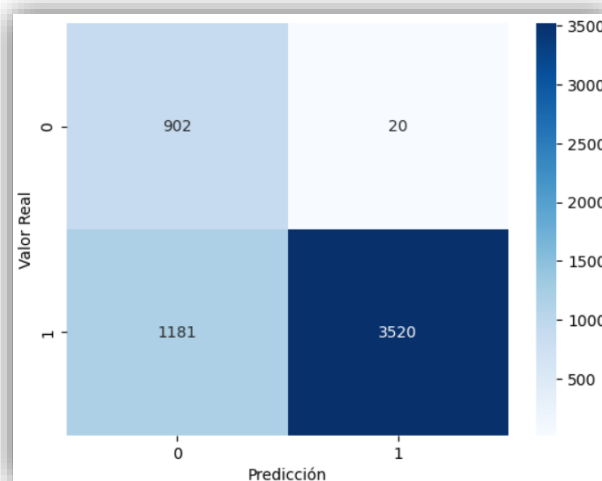


Ilustración 29. Matriz de confusión, modelo de prueba 1

Modelo prueba 2, se elimina tipología 2: Con los mismos datos de entrenamiento, pero omitiendo la tipología 2, se realizó el modelo de clasificación de tipologías entre la tipología 1 y 3.

Siendo la tipología 1 (0) y la tipología 3 (1), se evidencia que la métrica “F1-score” es superior a 80 % para ambas tipologías, y al analizar la matriz de confusión, 287 fotografías fueron clasificadas con la tipología 1, cuando realmente eran tipología 3.

	precision	recall	f1-score	support
0	0.7081	0.9612	0.8154	901
1	0.9920	0.9241	0.9568	4701
accuracy			0.9300	5602
macro avg	0.8501	0.9426	0.8861	5602
weighted avg	0.9463	0.9300	0.9341	5602

Ilustración 30. Métricas de desempeño, modelo de prueba 2

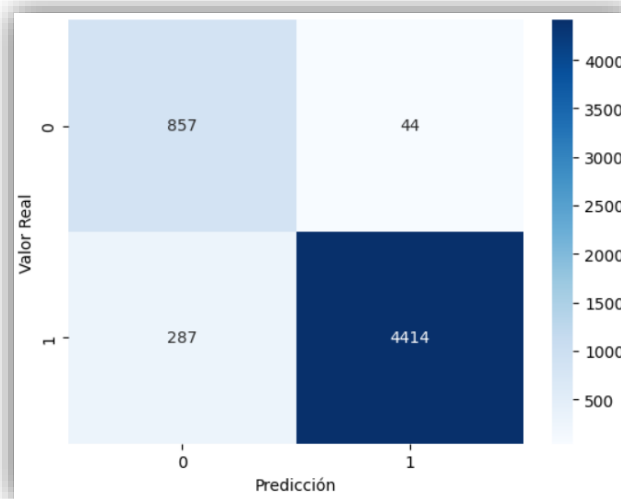


Ilustración 31. Matriz de confusión, modelo de prueba 2

Modelo prueba 3, se elimina tipología 3: Con los mismos datos de entrenamiento, pero omitiendo la tipología 3, se realizó el modelo de clasificación de tipologías entre la tipología 1 y 2.

Siendo la tipología 1 (0) y la tipología 2 (1), se evidencia que la métrica “F1-score” es superior a 90 % para ambas tipologías, y al analizar la matriz de confusión 105 fotografías fueron clasificadas con la tipología 2, cuando realmente eran tipología 1.

	precision	recall	f1-score	support
0	0.9454	0.8835	0.9134	901
1	0.8930	0.9501	0.9207	922
accuracy			0.9172	1823
macro avg	0.9192	0.9168	0.9170	1823
weighted avg	0.9189	0.9172	0.9171	1823

Ilustración 32. Métricas de desempeño, modelo de prueba 3

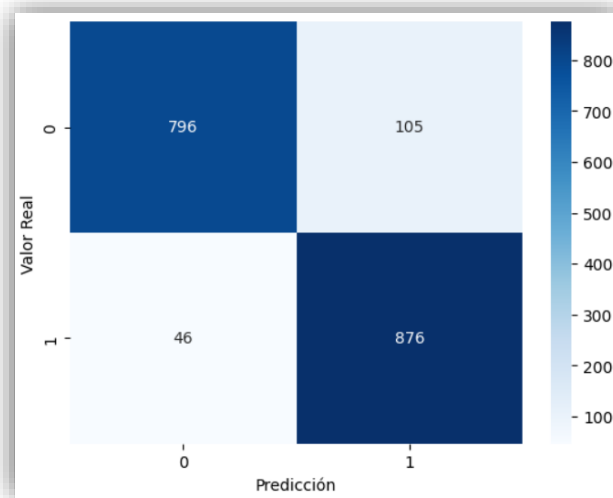


Ilustración 33. Matriz de confusión, modelo de prueba 3

Dados los resultados de los tres modelos de prueba anteriores, cuando se tiene en un modelo la tipología 2 y 3, la calidad de la clasificación baja considerablemente, tal como ocurre en el modelo de prueba 1; dicha disminución de calidad en la clasificación se debe a la particularidad de la tipología 2B, la cual describe en la tabla 1, como estructuras y muros prefabricados, los cuales son difícil de identificar o discriminar de tipologías como la 3A y 3B cuando los muros prefabricados se encuentran pañetados o pintados, tal como se muestra en la siguiente ilustración:



Ilustración 34. Casos con difícil discriminación de tipología

Comparación de resultados con otras técnicas

Según lo establecido en los antecedentes, en especial por Wang *et al.* [9] y García *et al.* [7], otra técnica común en la clasificación de imágenes son las Máquinas de Vectores de Soporte – Support Vector Machine (SVM), por lo cual se tomó como la técnica indicada para realizar la comparación con la técnica aplicada en este proyecto (Transfer Learning y Redes Convolucionales), en la siguiente tabla se muestra la comparación de 4 modelos de SVM aplicados al mismo conjunto de datos de entrenamiento y validación.

Modelo SVM	Kernel	Tamaño de imágenes	Tamaño de lote	Tiempo de procesamiento (Horas)	“Accuracy” en entrenamiento
1	Polinomial	250 x 250	5	0.50	0.68
2	Lineal	250 x 250	5	0.20	0.66
3	Radial	250 x 250	5	0.45	0.74
4	Sigmoidal	250 x 250	10	0.30	0.73

Tabla 17. Puntuación de modelos SVM

En el entrenamiento de los modelos anteriores, algo a resaltar es el tiempo de procesamiento; SVM es una técnica mucho más rápida y que consume menos recurso de máquina. Respecto a las métricas de desempeño dada en la tabla 17, esta parece tener mejor comportamiento que las métricas de desempeño de la tabla 16, sin embargo, es importante considerar todas las métricas en conjunto, para evitar que, por la naturaleza en el cálculo de “Accuracy” se puedan cometer errores en la interpretación, por lo cual, en las siguientes imágenes se muestra cada modelo a detalle.

Modelo SVM 1 y SVM 2

	precision	recall	f1-score	support
0	0.2373	0.1138	0.1538	123
1	0.2373	0.1273	0.1657	110
2	0.7524	0.8929	0.8166	616
accuracy			0.6808	849
macro avg	0.4090	0.3780	0.3787	849
weighted avg	0.6110	0.6808	0.6363	849

	precision	recall	f1-score	support
0	0.2045	0.1463	0.1706	123
1	0.2500	0.1455	0.1839	110
2	0.7518	0.8506	0.7982	616
accuracy			0.6572	849
macro avg	0.4021	0.3808	0.3842	849
weighted avg	0.6075	0.6572	0.6277	849

Ilustración 35. Métricas de desempeño modelo SVM1 y modelo SVM2

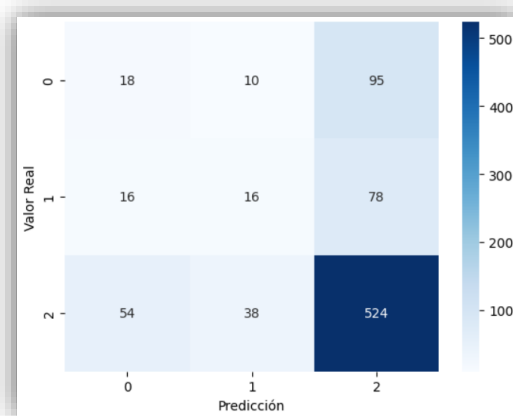
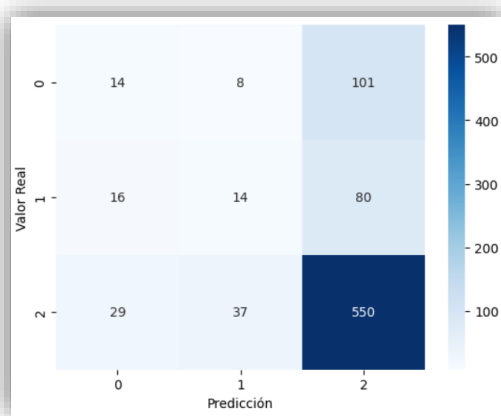


Ilustración 36. Matriz de confusión modelo SVM1 y modelo SVM2

Modelo SVM 3 y SVM 4

	precision	recall	f1-score	support
0	0.7500	0.0244	0.0472	123
1	0.7778	0.0636	0.1176	110
2	0.7356	0.9984	0.8471	616
accuracy			0.7362	849
macro avg	0.7545	0.3621	0.3373	849
weighted avg	0.7432	0.7362	0.6367	849

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	123
1	0.0000	0.0000	0.0000	110
2	0.7256	1.0000	0.8410	616
accuracy			0.7256	849
macro avg	0.2419	0.3333	0.2803	849
weighted avg	0.5264	0.7256	0.6102	849

Ilustración 37. Métricas de desempeño modelo SVM3 y modelo SVM4

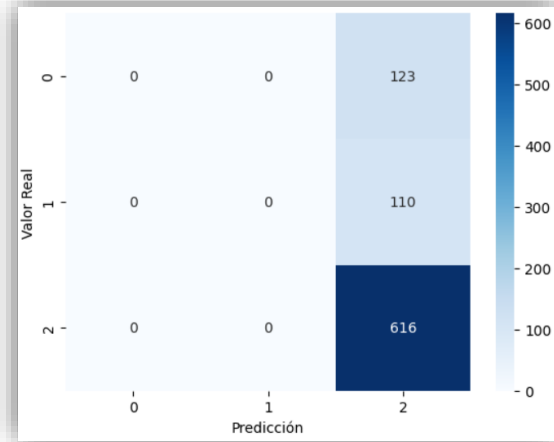
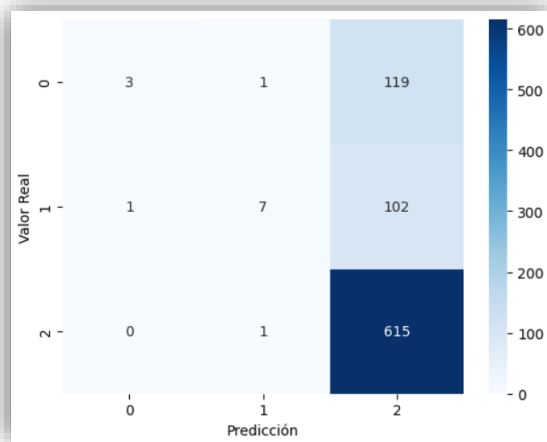


Ilustración 38. Matriz de confusión modelo SVM3 y modelo SVM4

En las métricas de desempeño evaluadas en las ilustraciones anteriores se evidencia que de manera general los resultados no son buenos, en especial para las tipologías generales 1 y 2, ya que en la métrica F1-Score no superan el 20%; en cuanto a la tipología general 3, se tienen de un 79% hasta un 84%, lo anterior indica que los modelos de la técnica SVM no generan una buena clasificación frente al modelo con las técnicas transfer learning y redes convolucionales.

Presentación del modelo y resultados de simulación

Debido a que el modelo funcional 8, presento mejor desempeño se presentan los ejemplos de aquellas imágenes clasificadas correctamente:

Tipología 1



Ilustración 39. Tipología 1 Modelo funcional 8

Tipología 2





Ilustración 40. Tipología 2 Modelo funcional 8

Tipología 3



Ilustración 41. Tipología 3 Modelo funcional 8

Se presentan de igual manera los ejemplos de aquellas imágenes clasificadas incorrectamente:

Tipología 1 clasificada como tipología 2



Ilustración 42. Errores de clasificación en tipología 1

Tipología 2 clasificada como tipología 3



Ilustración 43. Errores de clasificación en tipología 2

Tipología 3 clasificada como tipología 1

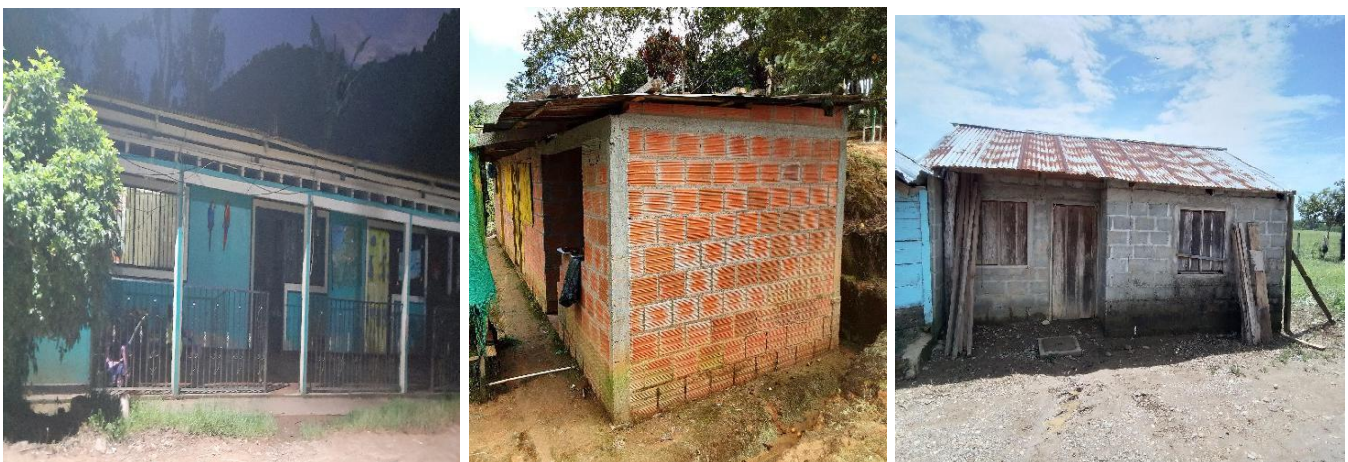


Ilustración 44. Errores de clasificación en tipología 3

5. CONCLUSIONES Y TRABAJOS FUTUROS

5.1 CONCLUSIONES

- El flujo de trabajo implementado en el presente proyecto refleja la necesidad de seguir una estrategia de recolección y procesamiento de imágenes rigurosa y estandarizada que permita identificar problemas de calidad. Frente a los problemas de calidad en las fotografías antes de ingresar la información a un repositorio final, es importante evaluar el tamaño y el formato, siendo 500 x 500 un tamaño adecuado que puede generar ahorro en almacenamiento, y JPG o PNG formatos aptos al presentar compatibilidad con muchos sistemas
- Al aplicar modelos como el clasificador de fachadas es posible identificar fotografías mal capturadas, con obstáculos, alejadas, borrosas o dañadas. Respecto a los problemas de calidad en las bases de datos, se debe controlar la correspondencia entre fotografías (su nombre) y los identificadores de base de datos a los que se les relaciona la etiqueta de tipología.
- En el desarrollo de modelos con técnicas de aprendizaje automático, enfocados en la tipificación de construcciones, se requieren tiempos extensos y recursos de cómputo elevados, fotografías de calidad y bien etiquetadas, por lo cual es fundamental aplicar de manera rigurosa la estrategia de recolección y procesamiento de imágenes antes de cualquier proceso de entrenamiento. Adicional a lo anterior, la cantidad de ejemplos disponibles limitan el entrenamiento por las tipologías detalladas del modelo LADM-COL, aún al utilizar la tipología residencial que es la más común y generalizar las tipologías del modelo LADM, se pueden tener problemas, tal como ocurrió con la tipología 4, la cual no fue posible implementar.
- Frente a la validación de los diferentes modelos estudiados, en este proyecto se obtuvieron buenos resultados gracias al uso de modelos pre entrenados ajustados a nuevos parámetros en combinación con redes convolucionales; lo anterior permitió con los datos y recursos de cómputo disponibles obtener un modelo funcional, ya que aumentando la cantidad de épocas de procesamiento se lograban mejorías considerables, pese a la existencia de similitud de fachadas de tipología 2 con las tipologías 1 y 3.

5.2 TRABAJOS FUTUROS

Los resultados obtenidos en este proyecto pueden ser la base de un proyecto futuro más robusto, donde se pueda implementar el flujo de recolección y procesamiento de imágenes como un control de calidad dentro de un proyecto de barrido predial o actualización catastral en implementación. Este podría recrear con la aplicación de modelos similares al de clasificación de fachada, considerando además de las fotografías de fachada, fotografías de materiales y variables categóricas adicionales, un modelo de clasificación que logre una discriminación más clara entre tipologías intermedias.

6. ANEXO

Modelo funcional 1

	precision	recall	f1-score	support
0	0.5217	0.9635	0.6769	137
1	0.5000	0.1898	0.2751	137
2	0.8396	0.6496	0.7325	137
accuracy			0.6010	411
macro avg	0.6205	0.6010	0.5615	411
weighted avg	0.6205	0.6010	0.5615	411

Ilustración 45. Métricas de desempeño modelo 1

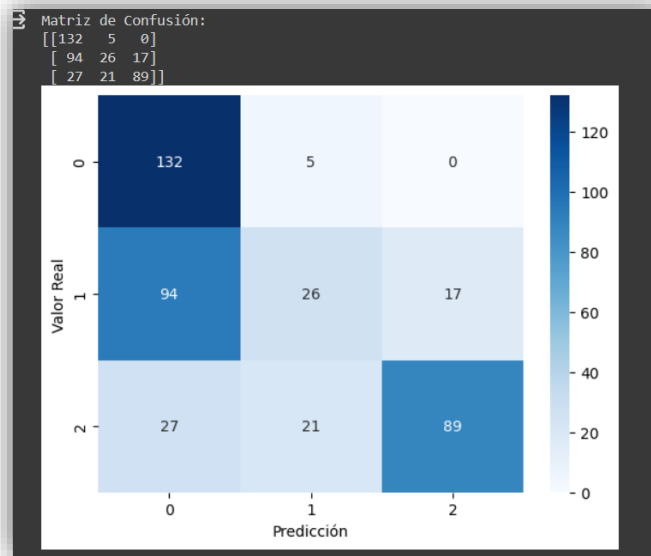


Ilustración 46. Matriz de confusión modelo 1

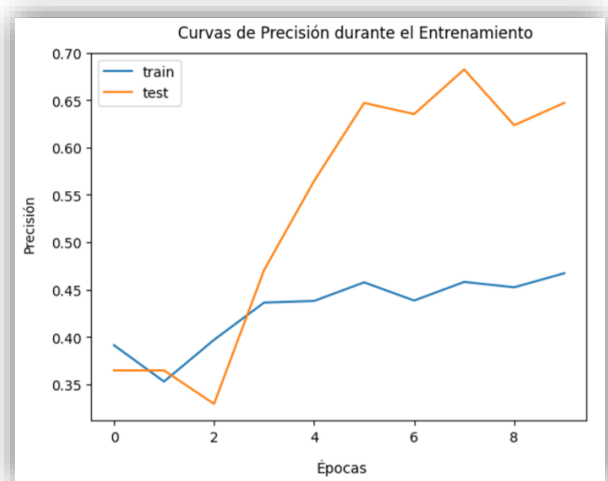
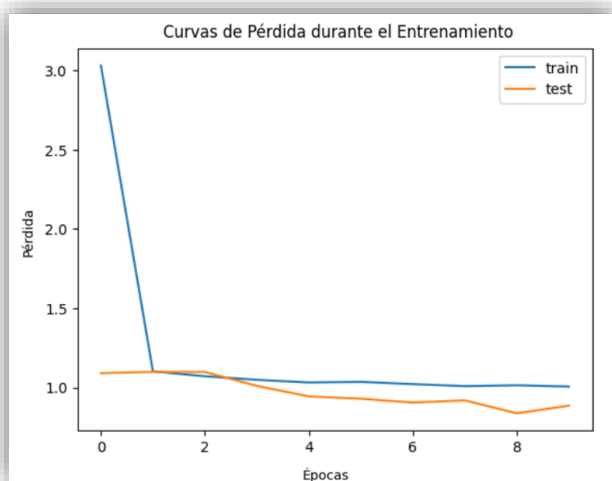


Ilustración 47. Curvas de pérdida y precisión modelo 1

Modelo funcional 2

	precision	recall	f1-score	support
0	0.7015	0.6861	0.6937	137
1	0.4556	0.2993	0.3612	137
2	0.6310	0.8613	0.7284	137
accuracy			0.6156	411
macro avg	0.5960	0.6156	0.5945	411
weighted avg	0.5960	0.6156	0.5945	411

Ilustración 48. Métricas de desempeño modelo 2

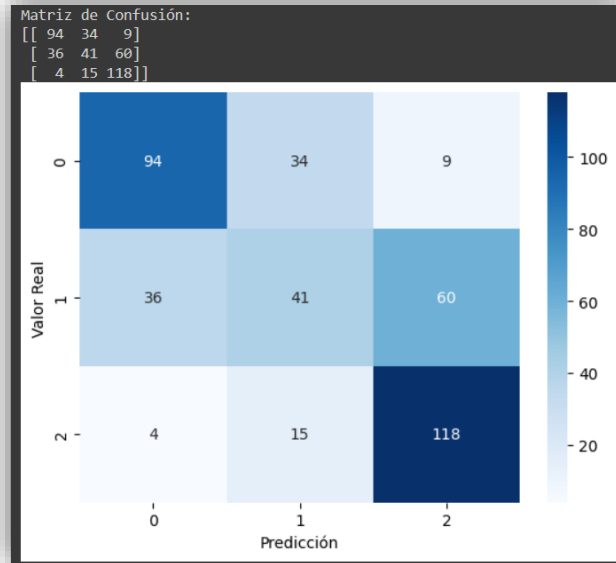


Ilustración 49. Matriz de confusión modelo 2

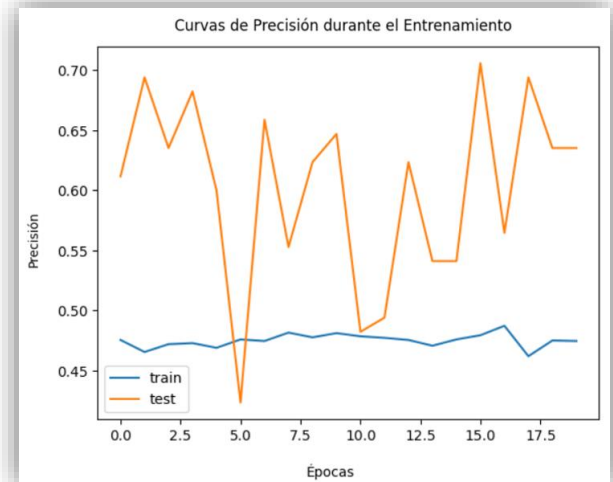
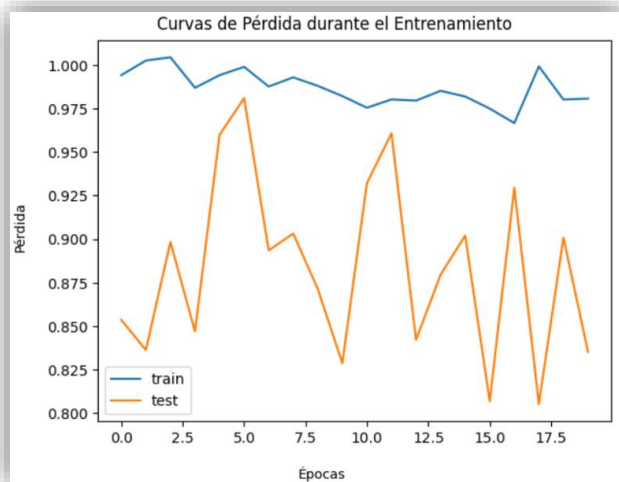


Ilustración 50. Curvas de pérdida y precisión modelo 2

Modelo funcional 3

	precision	recall	f1-score	support
0	0.5470	0.9343	0.6900	137
1	0.6579	0.1825	0.2857	137
2	0.7626	0.7737	0.7681	137
accuracy			0.6302	411
macro avg	0.6558	0.6302	0.5813	411
weighted avg	0.6558	0.6302	0.5813	411

Ilustración 51. Métricas de desempeño modelo 3

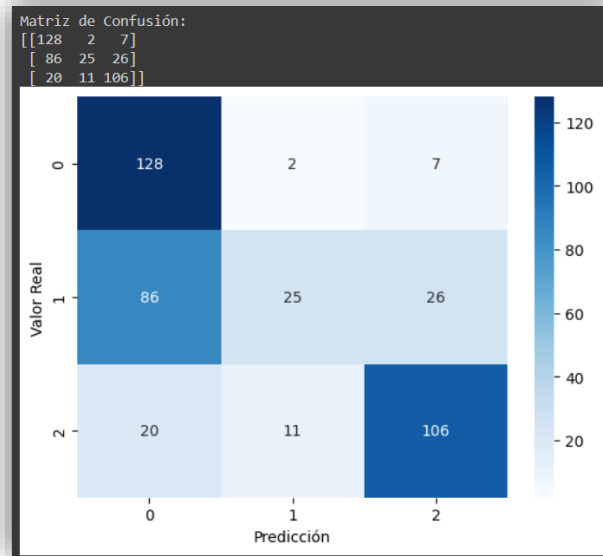


Ilustración 52. Matriz de confusión modelo 3

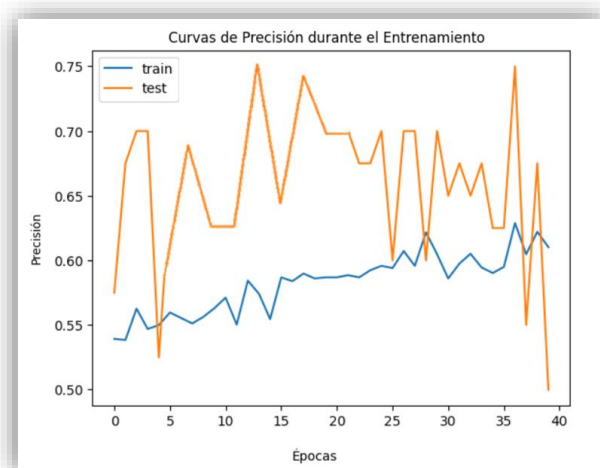
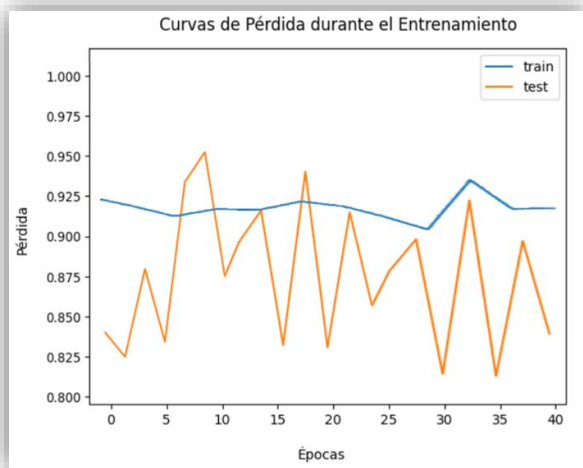


Ilustración 53. Curvas de pérdida y precisión modelo 3

Modelo funcional 4

	precision	recall	f1-score	support
0	0.7429	0.6724	0.7059	116
1	0.4894	0.3966	0.4381	116
2	0.6376	0.8190	0.7170	116
accuracy			0.6293	348
macro avg	0.6233	0.6293	0.6203	348
weighted avg	0.6233	0.6293	0.6203	348

Ilustración 54. Métricas de desempeño modelo 4

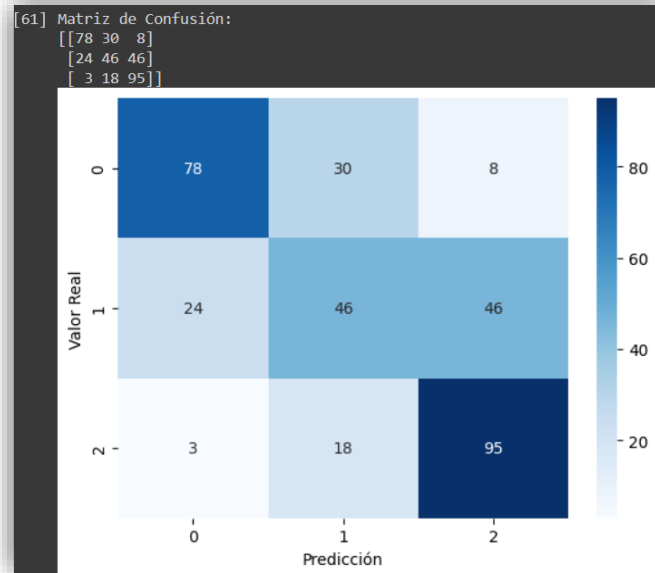


Ilustración 55. Matriz de confusión modelo 4

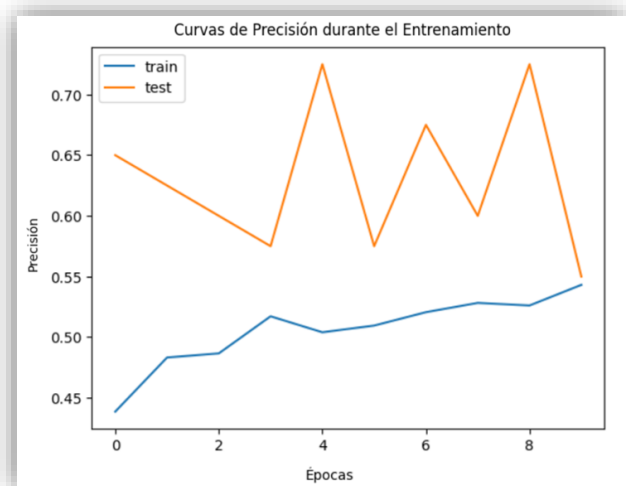
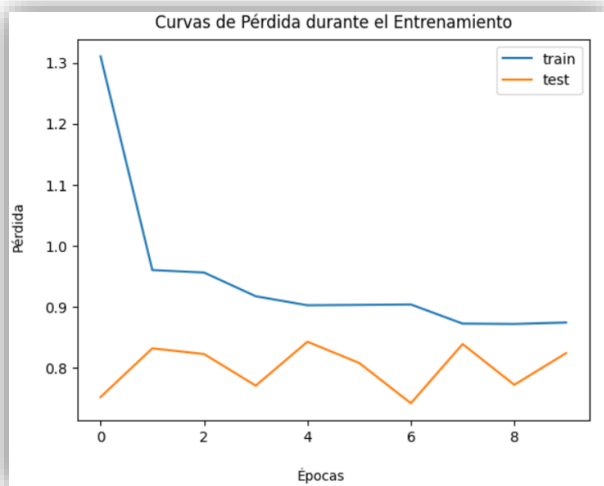


Ilustración 56. Curvas de pérdida y precisión modelo 4

Modelo funcional 5

	precision	recall	f1-score	support
0	0.7742	0.6207	0.6890	116
1	0.5263	0.6897	0.5970	116
2	0.7864	0.6983	0.7397	116
accuracy			0.6695	348
macro avg	0.6956	0.6695	0.6752	348
weighted avg	0.6956	0.6695	0.6752	348

Ilustración 57. Métricas de desempeño modelo 5

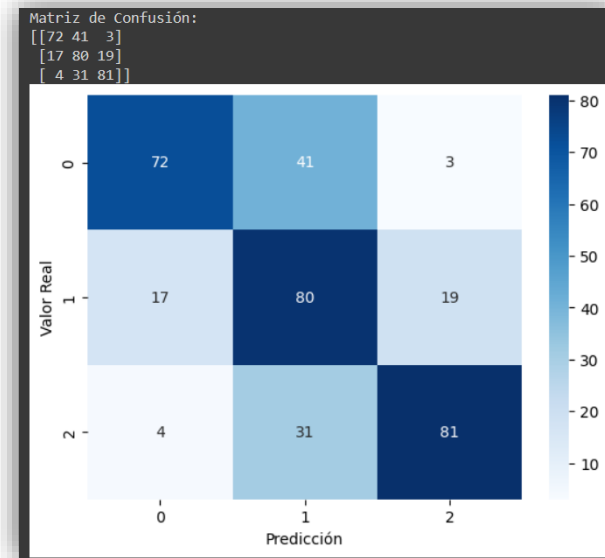


Ilustración 58. Matriz de confusión modelo 5

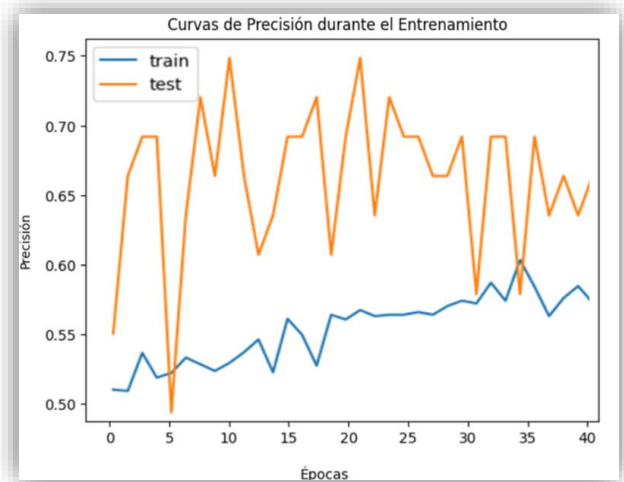
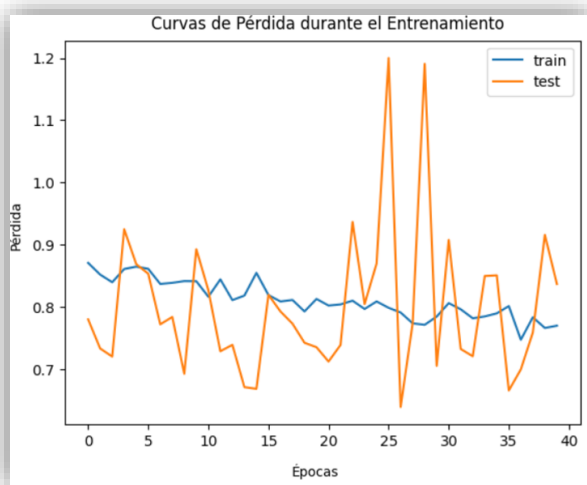


Ilustración 59. Curvas de pérdida y precisión modelo 5

Modelo funcional 6

	precision	recall	f1-score	support
0	0.8090	0.6207	0.7024	116
1	0.5346	0.7328	0.6182	116
2	0.8100	0.6983	0.7500	116
accuracy			0.6839	348
macro avg	0.7179	0.6839	0.6902	348
weighted avg	0.7179	0.6839	0.6902	348

Ilustración 60. Métricas de desempeño modelo 6

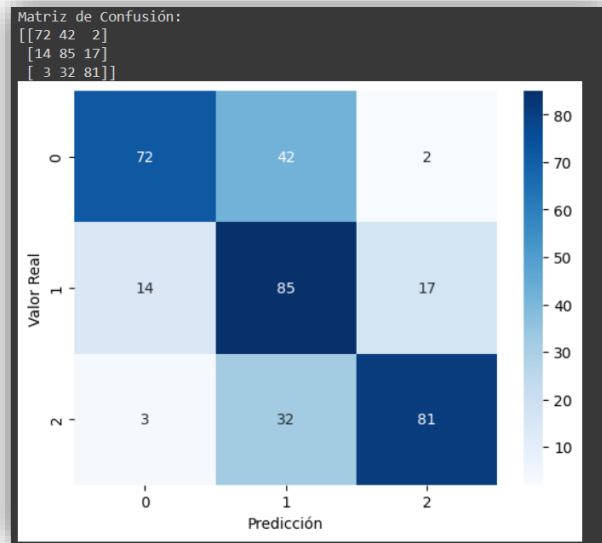


Ilustración 61. Matriz de confusión modelo 6

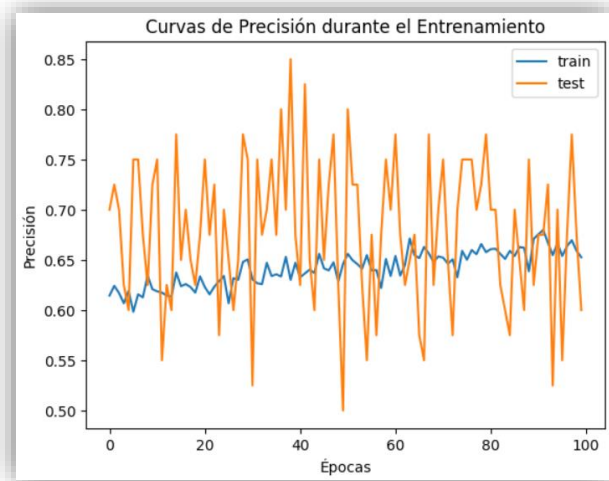
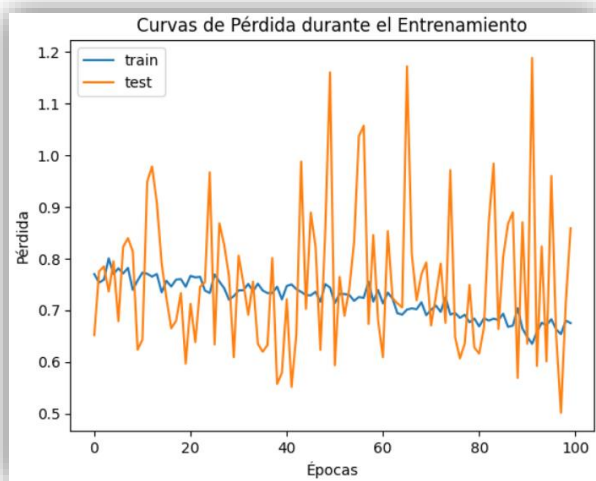


Ilustración 62. Curvas de pérdida y precisión modelo 6

Modelo funcional 7

	precision	recall	f1-score	support
0	0.8000	0.6207	0.6990	116
1	0.5556	0.6466	0.5976	116
2	0.7317	0.7759	0.7531	116
accuracy			0.6810	348
macro avg	0.6958	0.6810	0.6833	348
weighted avg	0.6958	0.6810	0.6833	348

Ilustración 63. Métricas de desempeño modelo 7

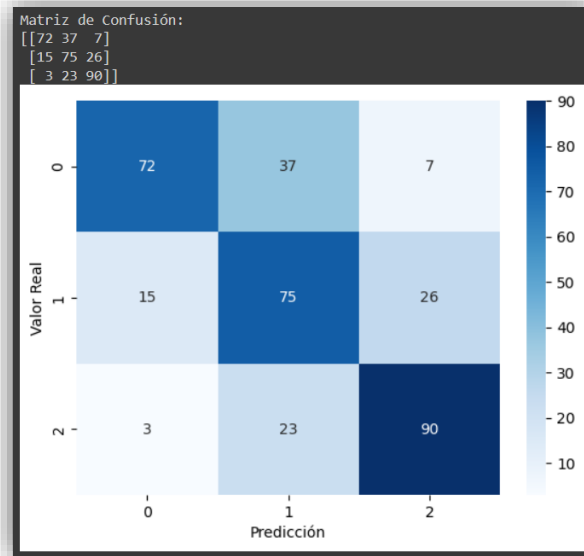


Ilustración 64. Matriz de confusión modelo 7

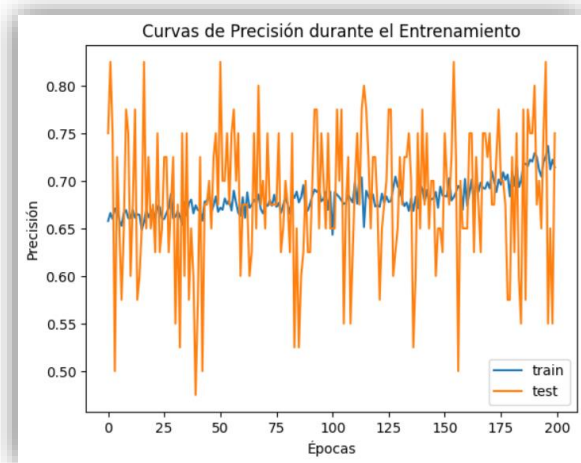
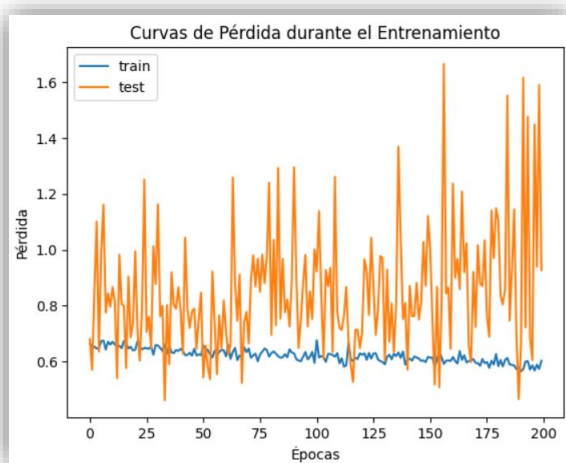


Ilustración 65. Curvas de pérdida y precisión modelo 7

Modelo funcional 8

	precision	recall	f1-score	support
0	0.8660	0.7241	0.7887	116
1	0.6183	0.6983	0.6559	116
2	0.7417	0.7672	0.7542	116
accuracy			0.7299	348
macro avg	0.7420	0.7299	0.7329	348
weighted avg	0.7420	0.7299	0.7329	348

Ilustración 66. Métricas de desempeño modelo 8

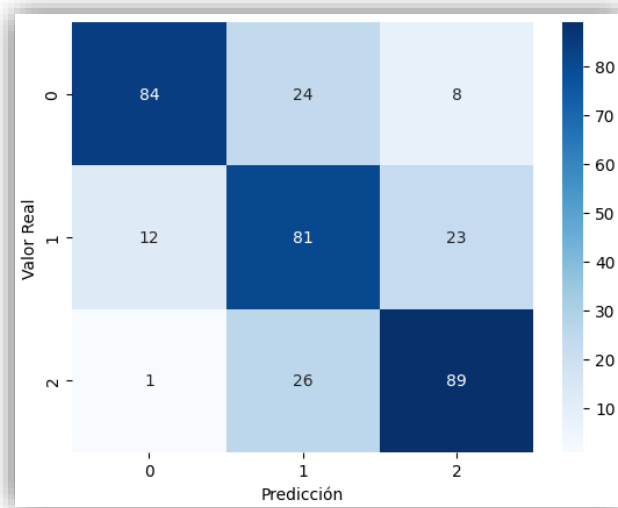


Ilustración 67. Matriz de confusión modelo 8

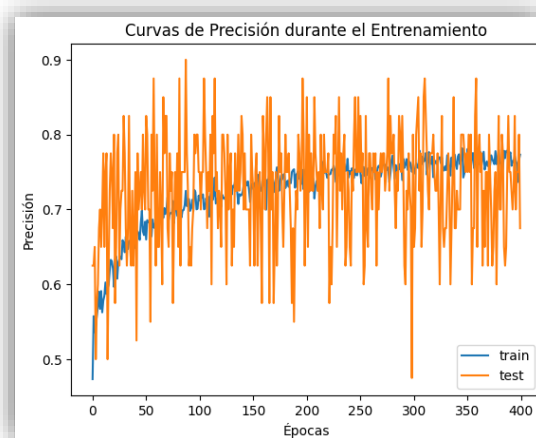
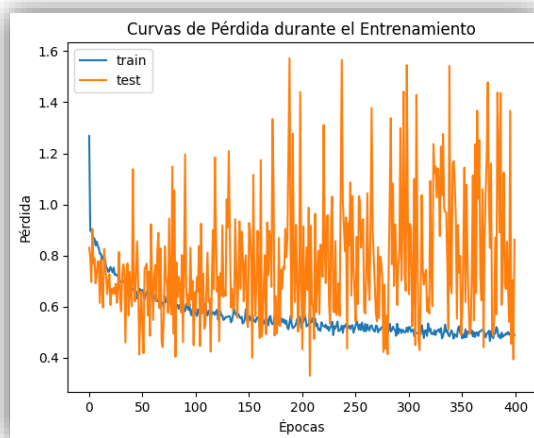


Ilustración 68. Curvas de pérdida y precisión modelo 8

Modelo funcional 9

	precision	recall	f1-score	support
0	0.7670	0.6810	0.7215	116
1	0.5969	0.6638	0.6286	116
2	0.7672	0.7672	0.7672	116
accuracy			0.7040	348
macro avg	0.7104	0.7040	0.7058	348
weighted avg	0.7104	0.7040	0.7058	348

Ilustración 69. Métricas de desempeño modelo 9

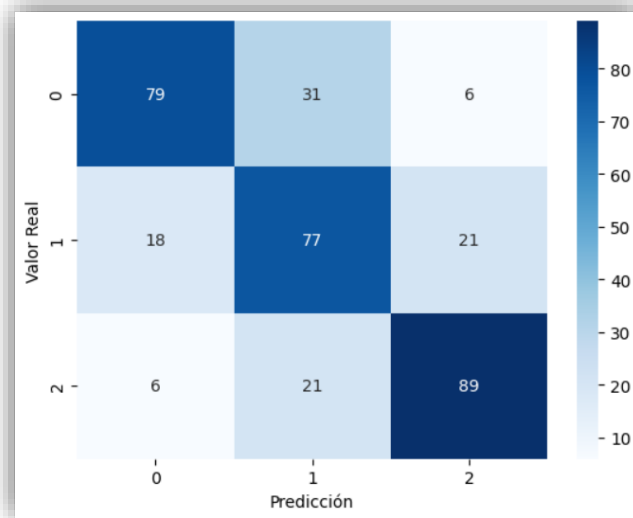


Ilustración 70. Matriz de confusión modelo 9

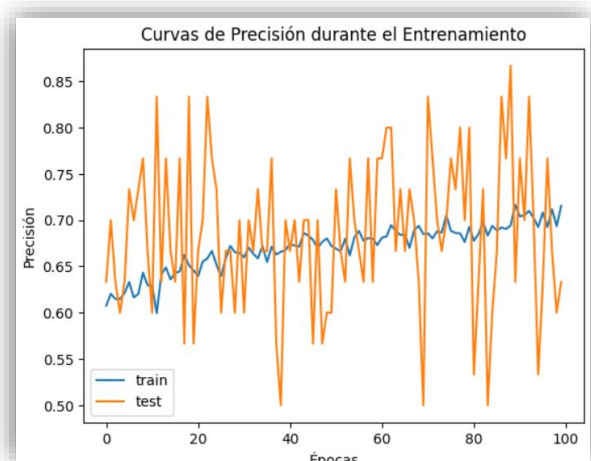
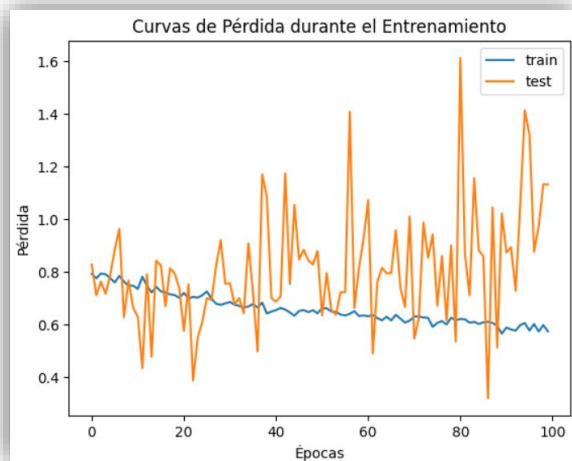


Ilustración 71. Curvas de pérdida y precisión modelo 9

Modelo funcional 10

	precision	recall	f1-score	support
0	0.7654	0.5345	0.6294	116
1	0.5455	0.5690	0.5570	116
2	0.6438	0.8103	0.7176	116
accuracy			0.6379	348
macro avg	0.6516	0.6379	0.6347	348
weighted avg	0.6516	0.6379	0.6347	348

Ilustración 72. Métricas de desempeño modelo 10

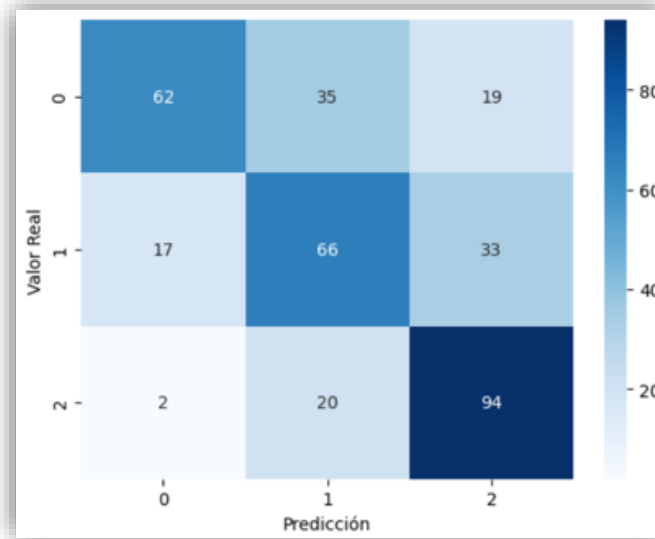


Ilustración 73. Matriz de confusión modelo 10

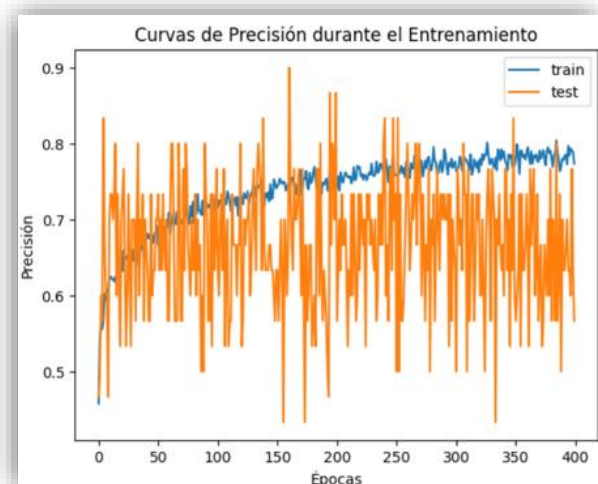
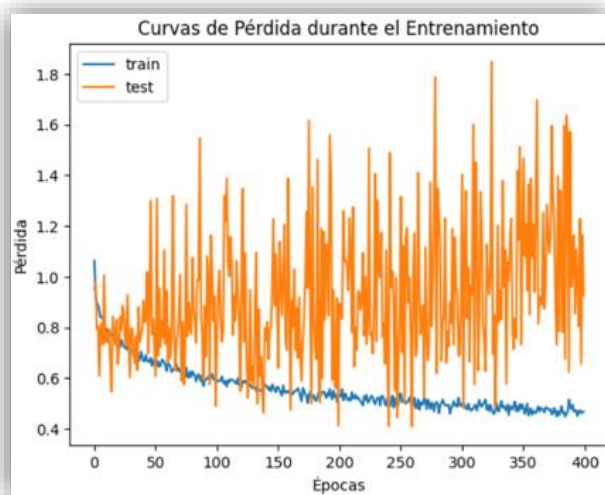


Ilustración 74. Curvas de pérdida y precisión modelo 10

Modelo funcional 11

	precision	recall	f1-score	support
0	0.7755	0.6552	0.7103	116
1	0.5441	0.6379	0.5873	116
2	0.7018	0.6897	0.6957	116
accuracy			0.6609	348
macro avg	0.6738	0.6609	0.6644	348
weighted avg	0.6738	0.6609	0.6644	348

Ilustración 75. Métricas de desempeño modelo 11

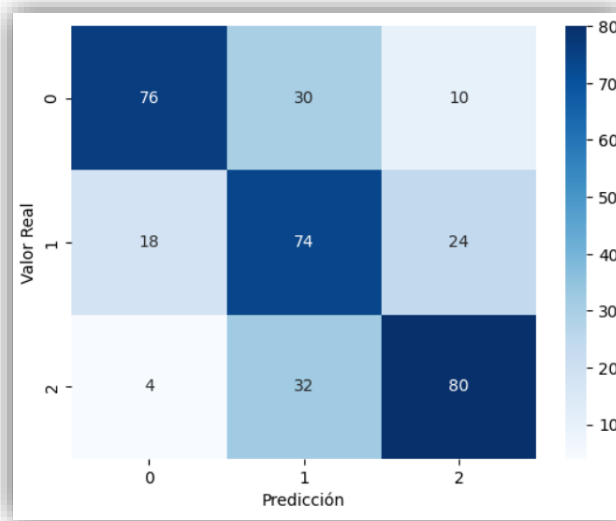


Ilustración 76. Matriz de confusión modelo 11

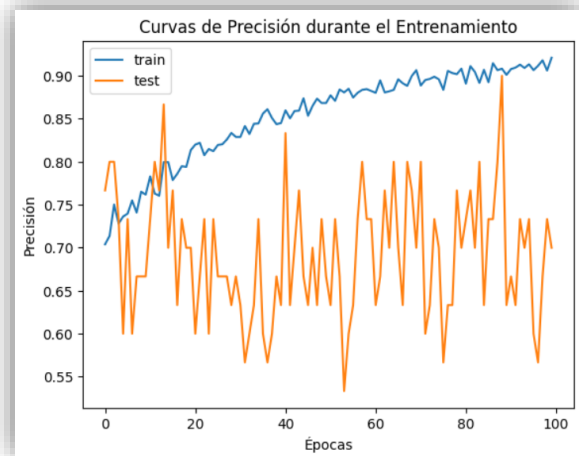
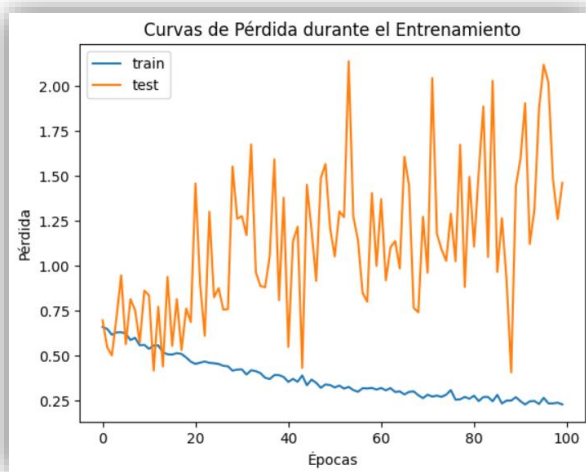


Ilustración 77. Curvas de pérdida y precisión modelo 11

Modelo funcional 12

	precision	recall	f1-score	support
0	0.3333	1.0000	0.5000	116
1	0.0000	0.0000	0.0000	116
2	0.0000	0.0000	0.0000	116
accuracy			0.3333	348
macro avg	0.1111	0.3333	0.1667	348
weighted avg	0.1111	0.3333	0.1667	348

Ilustración 78. Métricas de desempeño modelo 12

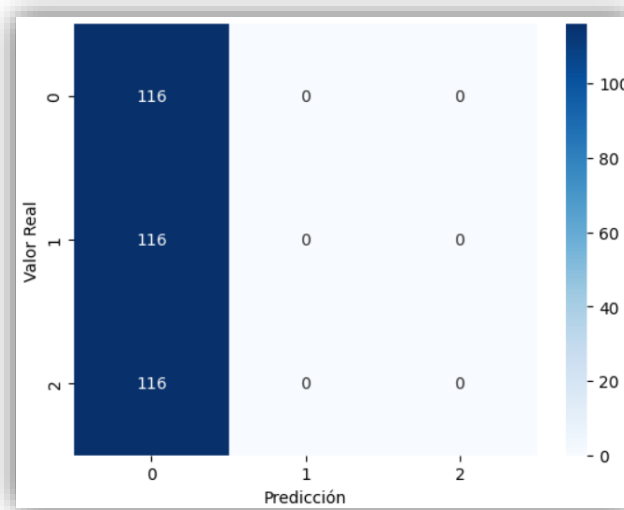


Ilustración 79. Matriz de confusión modelo 12

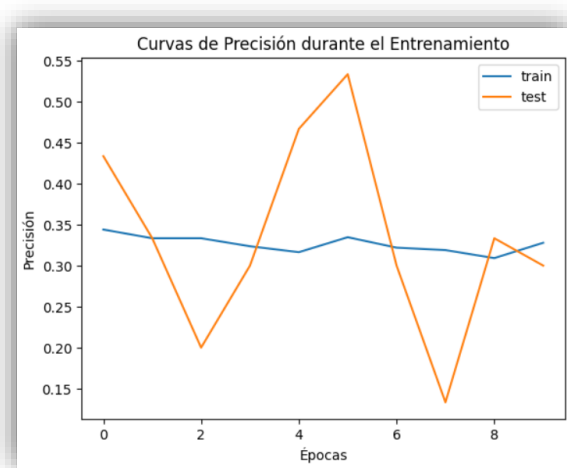
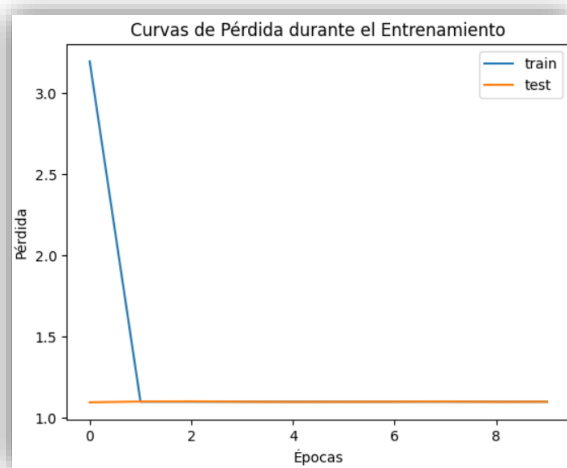


Ilustración 80. Curvas de pérdida y precisión modelo 12

Modelo funcional 13

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	116
1	0.0000	0.0000	0.0000	116
2	0.3333	1.0000	0.5000	116
accuracy			0.3333	348
macro avg	0.1111	0.3333	0.1667	348
weighted avg	0.1111	0.3333	0.1667	348

Ilustración 81. Métricas de desempeño modelo 13

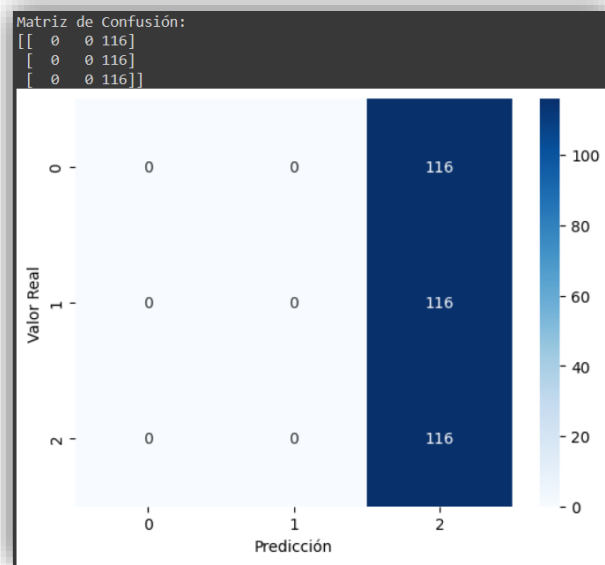


Ilustración 82. Matriz de confusión modelo 13

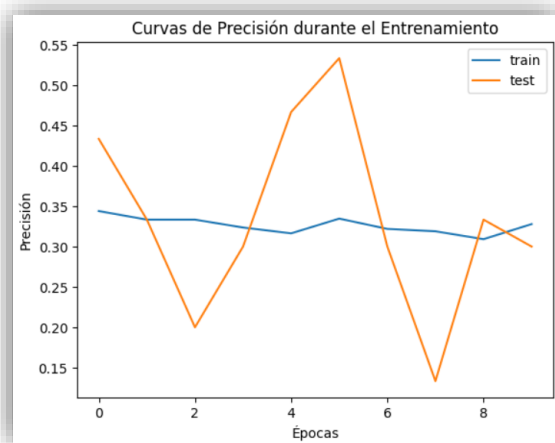
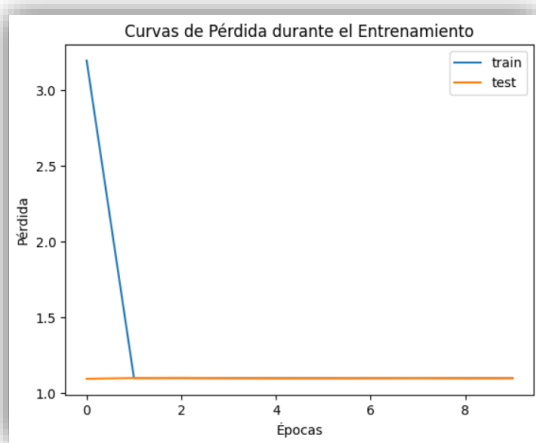


Ilustración 83. Curvas de pérdida y precisión modelo 13

Modelo funcional 14

	precision	recall	f1-score	support
0	0.7379	0.6552	0.6941	116
1	0.5476	0.5948	0.5702	116
2	0.7143	0.7328	0.7234	116
accuracy			0.6609	348
macro avg	0.6666	0.6609	0.6626	348
weighted avg	0.6666	0.6609	0.6626	348

Ilustración 84. Métricas de desempeño modelo 14

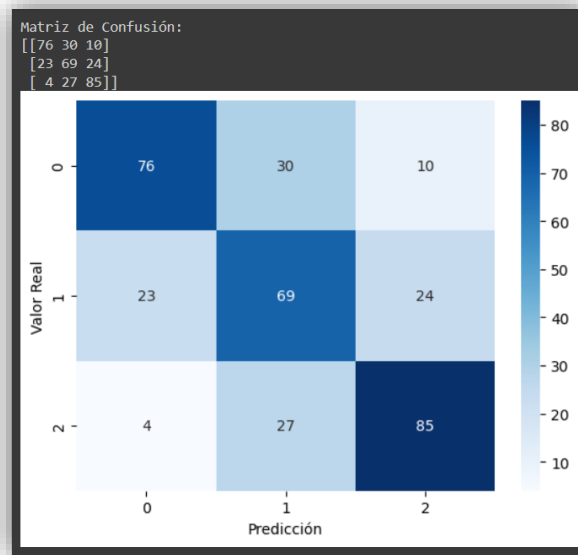


Ilustración 85. Matriz de confusión modelo 14

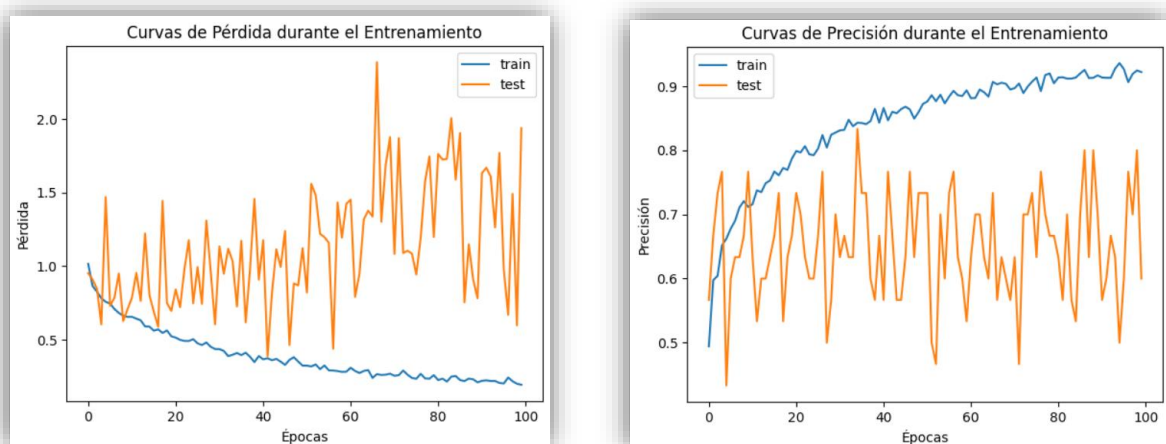


Ilustración 86. Curvas de pérdida y precisión modelo 14

7. REFERENCIAS BIBLIOGRÁFICAS

- [1] Departamento Nacional de Planeación, «DPN.gov.com,» 26 03 2019. [En línea]. Available: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3958.pdf>. [Último acceso: 09 06 2023].
- [2] Instituto Geográfico Agustín Codazzi, «Observatorio del espacio público de Bogotá,» [En línea]. Available: https://observatorio.dadep.gov.co/sites/default/files/documentos/mg48_manual_de_reconocimiento_predial.pdf. [Último acceso: 19 05 2023].
- [3] Dirección de Sistemas de Información y Catastro, «Alcaldía de Bello,» 2010. [En línea]. Available: <http://intranet.bello.gov.co:8081/intranet/start/sig/procesos/estrategicos/direccionamiento%20estrategico/documentos%20externos/DA%20catastro/48%20Manual%20de%20Reconocimiento%20Predial.pdf>. [Último acceso: 19 05 2023].
- [4] Departamento Nacional de Política Económica y Social, CONPES 3859, Bogotá, D.C.: Consejo Nacional de Política Económica y Social, 2016.
- [5] Instituto Geográfico Agustín Codazzi, «Reconocedores prediales serán cruciales para la transformación de Colombia en los tiempos de la paz: Director del IGAC,» IGAC, Bucaramanga, 2023.
- [6] Instituto Geográfico Agustín Codazzi, «IGACNET,» 07 2017. [En línea]. Available: <http://igacnet2.igac.gov.co/intranet/UserFiles/File/procedimientos/Methodologias%202008/2017/M51500-01%20Control%20de%20calidad%20en%20la%20etapa%20de%20reconocimiento%20predial.pdf>. [Último acceso: 19 05 2023].
- [7] R. A. B. J. P. T. J. C. David García-Retuerta, «Dialnet,» 12 2018. [En línea]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=7289190>. [Último acceso: 09 06 2023].
- [8] H. Thomas, «IA: Inteligencia Artificial,» *POLIS, Revista de la Universidad Bolivariana*, vol. 1, n° 2, p. 0, 2001.
- [9] E. F. P. W. Pin Wang, «ScienceDirect,» 1 08 2020. [En línea]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865520302981>. [Último acceso: 09 06 2023].
- [10] Departamento Administrativo Nacional de Estadística, «Función Pública,» 4 02 2020. [En línea]. Available: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=105952>. [Último acceso: 09 06 2023].
- [11] J. C. Z. Galla, «Biblioteca Digital Universidad de Antioquia,» 2021. [En línea]. Available: <https://bibliotecadigital.udea.edu.co/handle/10495/24988>. [Último acceso: 09 06 2023].
- [12] Superintendencia de Notariado y Registro, Instituto Geográfico Agustín Codazzi, «IGAC.com,» 28 05 2020. [En línea]. Available: <https://igac.gov.co/es/contenido/resolucion-conjunta-snr-04218-igac-499-de-2020>. [Último acceso: 09 06 2023].
- [13] Instituto Geográfico Agustín Codazzi, «Resolución 1040 de 2023,» 08 08 2023. [En línea]. Available: <https://antiguo.igac.gov.co/es/contenido/resolucion-1040-de-2023>. [Último acceso: 28 01 2024].
- [14] G. H. Abraham Gómez Morales, «Redalyc,» 9 2009. [En línea]. Available: <https://www.redalyc.org/pdf/1331/133113598004.pdf>. [Último acceso: 09 06 2023].
- [15] L. V. C. Luis, «Repositorio Universidad Distrital Francisco Jose de Caldas,» 2020. [En línea]. Available: <https://repository.udistrital.edu.co/handle/11349/22980>. [Último acceso: 09 06 2023].
- [16] D. F. V. Avila, «Repositorio Universidad Distrital Francisco Jose de Caldas,» 2020. [En línea]. Available: <https://repository.udistrital.edu.co/handle/11349/29164>. [Último acceso: 09 06 2023].
- [17] P. A. Daniel, R. R. Redy y C. P. Luis, «Transfer learning en la clasificación binaria de imágenes térmicas,» 01 07 2021. [En línea]. Available: <https://www.redalyc.org/journal/5055/505567902007/>. [Último acceso: 06 04 2024].
- [18] Keras, «Keras Applocations,» [En línea]. Available: <https://keras.io/api/applications/#usage-examples-for-image-classification-models>. [Último acceso: 06 04 2024].
- [19] M. P. O. d. Landaluce, «Biblioteca,» Universidad de Sevilla, 2021. [En línea]. Available: <https://biblus.us.es/bibing/proyectos/abreproy/93772/fichero/TFG-3772+P%C3%89REZ+ORTIZ+DE+LANDALUCE%2C+MARTA.pdf>. [Último acceso: 06 06 2024].