



UNIVERSIDAD
DE GRANADA

TRABAJO FIN DE GRADO
INGENIERÍA INFORMÁTICA

Algoritmos meméticos para reducir datos de entrenamiento en modelos de aprendizaje profundo convolucionales

Autor

José Ruiz López (alumno)

Directores

Daniel Molina Cabrera (tutor)



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, Junio de 2025

Algoritmos meméticos para reducir datos de entrenamiento en modelos de aprendizaje profundo convolucionales

José Ruiz López (alumno)

Palabras clave: Algoritmos meméticos, Imágenes, Modelos de Aprendizaje profundo convolucionales

Resumen

Los **modelos de Aprendizaje Profundo** (Deep Learning) han supuesto un verdadero hito en la **Inteligencia Artificial**, ya que son capaces de procesar grandes volúmenes de datos, además de reconocer patrones sumamente complejos. Dentro de estos, los **modelos convolucionales** se han destacado como particularmente efectivos a la hora de identificar objetos y características en imágenes —una capacidad esencial para muchas aplicaciones modernas—. Sin embargo, a diferencia de los seres humanos, estos modelos requieren una gran cantidad de datos de entrenamiento para cada categoría que deben aprender. Esto implica un proceso de entrenamiento más largo y, muchas veces, la recolección de los datos necesarios puede ser problemática, según el tipo de información que se necesite.

Además de la dificultad en la obtención de datos, las nuevas normativas europeas en torno a la inteligencia artificial establecen la necesidad de auditar no solo los modelos, sino también los datos utilizados para entrenarlos, especialmente cuando se trata de aplicaciones de IA que manejan datos sensibles. Estas auditorías, por su propia naturaleza, se volverán más complejas conforme aumente el tamaño del conjunto de entrenamiento. Por lo tanto, se vuelve completamente necesario desarrollar estrategias que permitan **reducir el tamaño de los conjuntos de datos de entrenamiento** intentado comprometer la calidad del modelo lo mínimo posible.

En este trabajo, proponemos el uso de **algoritmos meméticos** para establecer un proceso de reducción del conjunto de **entrenamiento**, lo que se conoce como **selección de instancias**. La idea es seleccionar un conjunto reducido de imágenes representativas que, junto con las técnicas de aumento de datos, sean suficientes para entrenar modelos convolucionales de manera óptima. De este modo, se podría reducir significativamente el tamaño del conjunto de entrenamiento, manteniendo la calidad del aprendizaje y, a su vez, facilitando tanto el proceso de auditoría como la eficiencia computacional del sistema.

Memetic Algorithms for Reducing Training Data in Convolutional Deep Learning Models

José, Ruiz López (student)

Keywords: Memetic Algorithms, Images, Convolutional Deep Learning Models

Abstract

Deep Learning models have marked a true milestone in the field of **Artificial Intelligence**, as they are capable of processing large volumes of data and recognizing highly complex patterns. Among these, **convolutional models** have stood out as particularly effective in identifying objects and features in images—an essential capability for many modern applications. However, unlike humans, these models require a large amount of training data for each category they need to learn. This implies a longer training process, and in many cases, collecting the necessary data can be problematic depending on the type of information required.

In addition to the difficulty of obtaining data, new European regulations on artificial intelligence establish the need to audit not only the models but also the data used to train them, especially in AI applications that handle sensitive data. These audits, by their very nature, will become increasingly complex as the size of the training set grows. Therefore, it becomes essential to develop strategies that allow for **reducing the size of training datasets**, while minimizing any compromise in model quality.

In this work, we propose the use of **memetic algorithms** to implement a **training data reduction process**, also known as **instance selection**. The idea is to select a small set of representative images that, along with data augmentation techniques, are sufficient to optimally train convolutional models. In this way, it would be possible to significantly reduce the size of the training set while maintaining learning quality and, at the same time, facilitating both the auditing process and the system's computational efficiency.

Yo, **José Ruiz López**, alumno de la titulación INGENIERÍA INFORMÁTICA de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI **77964364E**, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: José Ruiz López

Granada a 8 de junio de 2025.

D. **Daniel Molina Cabrera** (tutor, Profesor del Departamento Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada).

Informan:

Que el presente trabajo, titulado ***Algoritmos meméticos para reducir datos de entrenamiento en modelos de aprendizaje profundo convolucionales***, ha sido realizado bajo su supervisión por **José Ruiz López (alumno)**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 8 de junio de 2025.

Los directores:

Daniel Molina Cabrera (tutor)

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que han hecho posible la realización de este Trabajo de Fin de Grado.

En primer lugar, me gustaría agradecer al profesor Daniel Molina Cabrera, mi tutor, por su valiosa guía, por su apoyo continuo durante todo el proceso y por brindarme acceso a los recursos necesarios para llevar a cabo esta investigación. Su experiencia y disponibilidad han sido fundamentales para que este proyecto pudiera desarrollarse de forma rigurosa y enriquecedora.

También quiero destacar lo desafiante que ha sido compaginar este trabajo académico con mis responsabilidades laborales. Ha requerido un esfuerzo constante y una gran capacidad de organización, pero también me ha permitido valorar aún más el proceso y el aprendizaje adquirido.

Agradezco de corazón a mis padres y amigos por su comprensión, paciencia y apoyo emocional durante los momentos más exigentes del camino. Y, sobre todo, a mi pareja, sin ella no habría sido posible superar el reto que supone realizar un trabajo como este. Su apoyo incondicional, su confianza en mí y la calma que me ha transmitido en los momentos más difíciles han sido clave para llegar hasta el final.

Asimismo, extendiendo mi gratitud a todas las personas y compañeros que, directa o indirectamente, han contribuido con sus comentarios, sugerencias y tiempo. Gracias a ellos, este trabajo ha podido alcanzar una mayor profundidad y claridad.

Por último, quiero agradecer a todas aquellas comunidades de código abierto y herramientas libres que han permitido que este proyecto se desarrollara sin barreras tecnológicas, facilitando el aprendizaje y la innovación.

Índice general

1. Introducción	1
1.1. Contexto	1
1.2. Motivación	1
1.3. Objetivos	2
2. Metodología y Presupuesto	5
2.1. Metodología	5
2.1.1. Enfoque Ágil Basado en Scrum	5
2.1.2. Organización de Tareas	5
2.1.3. Ciclos de Trabajo	6
2.1.4. Seguimiento del Progreso	7
2.1.5. Ajuste de los Tiempos	7
2.2. Presupuesto	8
2.2.1. Mano de obra	8
2.2.2. Recursos computacionales	9
2.2.3. Software y licencias	10
2.2.4. Gastos indirectos	10
2.2.5. Presupuesto total	10
3. Fundamentos de Aprendizaje Profundo	13
3.1. Definición de Aprendizaje Profundo	13
3.2. Redes Neuronales Artificiales	14
3.2.1. Componentes de una Red Neuronal	14

3.3. Redes Neuronales Convolucionales	16
3.3.1. Componentes principales de una CNN	16
3.3.2. Funcionamiento General de una CNN	17
3.3.3. Aplicaciones de las CNN	18
3.4. Modelos	18
3.4.1. ResNet50	18
3.4.2. MobileNetV2	19
4. Repaso Bibliográfico	21
4.1. Reducción de Conjuntos de Datos en Aprendizaje Profundo .	21
4.2. Selección de Instancias mediante Algoritmos Evolutivos y Meméticos	22
4.3. Algoritmos Meméticos en Modelos Convolucionales	22
4.4. Aplicaciones y Desafíos de los Algoritmos Meméticos	23
4.5. Perspectivas Futuras en la Optimización de CNN con Algoritmos Meméticos	23
5. Descripción de los Algoritmos	25
5.1. Algoritmo Aleatorio	25
5.1.1. Descripción	25
5.1.2. Pseudocódigo	26
5.1.3. Aplicación en la reducción de datos	26
5.1.4. Resultados esperados	26
5.2. Algoritmo Búsqueda Local	27
5.2.1. Descripción	27
5.2.2. Pseudocódigo	27
5.2.3. Aplicación en la reducción de datos	28
5.2.4. Ventajas y limitaciones	28
5.3. Algoritmo Genético	28
5.3.1. Descripción	28
5.3.2. Componentes del Algoritmo	29
5.3.3. Pseudocódigo	30

5.3.4.	Aplicación en la Reducción de Datos	30
5.3.5.	Ventajas y Limitaciones	31
5.4.	Algoritmo Genético con Cruce Ponderado	31
5.4.1.	Descripción	31
5.4.2.	Componentes del Algoritmo	32
5.4.3.	Pseudocódigo	32
5.4.4.	Aplicación en la Reducción de Datos	32
5.4.5.	Ventajas y Limitaciones	32
5.5.	Algoritmo Genético con Mutación Adaptativa	33
5.5.1.	Descripción	33
5.5.2.	Componentes del Algoritmo	34
5.5.3.	Pseudocódigo	34
5.5.4.	Aplicación en la Reducción de Datos	34
5.5.5.	Ventajas y Limitaciones	35
5.6.	Algoritmo Genético con Reinicio Poblacional	35
5.6.1.	Descripción	35
5.6.2.	Componentes del Algoritmo	36
5.6.3.	Pseudocódigo	36
5.6.4.	Aplicación en la Reducción de Datos	37
5.6.5.	Ventajas y Limitaciones	37
5.7.	Algoritmo Memético	38
5.7.1.	Descripción	38
5.7.2.	Pseudocódigo	38
5.7.3.	Aplicación en la reducción de datos	38
5.7.4.	Ventajas y limitaciones	38
5.8.	Versiones Libres de los Algoritmos Evolutivos	40
5.8.1.	Algoritmo Genético con Cruce Ponderado (Versión Libre)	40
5.8.2.	Algoritmo Genético con Reinicio Poblacional (Versión Libre)	40
5.8.3.	Algoritmo Memético (Versión Libre)	41

6. Implementación	43
6.1. Descripción del Sistema	43
6.2. Herramientas y Lenguajes de Programación	44
6.3. Gestión de Dependencias	45
6.4. Arquitectura de la Implementación	46
6.4.1. Esquema General de Funcionamiento de los Algoritmos	46
6.4.2. Configuración de Modelos Preentrenados	47
6.4.3. Módulo de Algoritmos	48
6.4.4. Núcleo de Ejecución	48
6.4.5. Módulo de Utilidades	49
6.4.6. Scripts de Ejecución en GPU	49
6.5. Consideraciones de Optimización	50
6.6. Métricas de Evaluación	51
7. Entorno Experimental	53
7.1. Datasets utilizados	53
7.1.1. Rock, Paper, Scissors (Piedra, Papel, Tijera)	53
7.1.2. PAINTING (Art Images: Drawing/Painting/Sculptures/Engravings)	56
7.1.3. Comparación entre datasets	58
7.2. Diseño de los experimentos	59
7.3. Procedimiento de Ejecución y Evaluación	59
7.3.1. Métricas de Evaluación	59
7.3.2. Evaluaciones por Ejecución	60
7.3.3. Repeticiones y Semillas	60
7.3.4. Tiempos de Ejecución	60
7.4. Visualización de resultados	61
7.5. Nomenclatura simplificada de los algoritmos	62
7.6. Estructura de las Tablas de Resultados	63
7.7. Parámetros de los Algoritmos y del Entrenamiento	64
7.7.1. Parámetros Generales de los Algoritmos	64

7.7.2. Parámetros del Entrenamiento de Modelos	65
7.7.3. Parámetros Específicos por Algoritmo	65
8. Resultados y Análisis	67
8.1. Resultados con el conjunto completo	67
8.2. Comparativa inicial de modelos	68
8.3. Resultados de la búsqueda local	70
8.4. Resultados del Algoritmo Genético	71
8.5. Mejorando el operador de cruce	72
8.6. Mejorando el operador de mutación	73
8.7. Resultados del reinicio poblacional	75
8.8. Resultados con versiones libres	77
8.9. Resultados del algoritmo memético	81
8.10. Resultados finales entre enfoques	83
8.10.1. Análisis comparativo de accuracy	83
8.10.2. Porcentaje final de datos seleccionados	85
8.10.3. Estabilidad en la distribución de clases	86
8.10.4. Síntesis final	87
8.11. Validación con el dataset PAINTING	89
8.11.1. Comparación de <i>accuracy</i> entre algoritmos	89
8.11.2. Impacto del porcentaje inicial	90
8.11.3. Equilibrio en la distribución de clases	93
8.11.4. Evolución del tamaño del subconjunto	93
9. Conclusiones	97
10. Bibliografía	101

Capítulo 1

Introducción

1.1. Contexto

En la actualidad, vivimos en una era marcada por una constante y acelerada generación de datos. Este fenómeno ha incrementado la necesidad de desarrollar métodos eficaces para el procesamiento y análisis de grandes volúmenes de información. En este contexto, los **modelos de aprendizaje profundo**, y en particular las **redes neuronales convolucionales (CNN)**, han demostrado un notable rendimiento en tareas como la **clasificación de imágenes**, el **reconocimiento de patrones** y diversas aplicaciones de alta complejidad. No obstante, el entrenamiento de estos modelos suele requerir volúmenes elevados de datos, lo que plantea desafíos significativos tanto en términos de **tiempo** como de **costes** asociados a su obtención.

Conforme los sistemas de inteligencia artificial evolucionan hacia estructuras más sofisticadas y precisas, la disponibilidad de conjuntos de datos amplios y adecuados se vuelve un requisito cada vez más crucial. Sin embargo, la recopilación, almacenamiento y tratamiento de estos datos suponen obstáculos importantes, especialmente para aquellas instituciones u organizaciones que cuentan con recursos limitados. Esta situación pone de relieve la necesidad de investigar estrategias innovadoras que permitan **reducir y optimizar los conjuntos de datos** sin comprometer el rendimiento de los modelos entrenados.

1.2. Motivación

La necesidad de **reducir los conjuntos de datos de entrenamiento** responde al objetivo de mejorar la **eficiencia** en el desarrollo de modelos de aprendizaje profundo. Aunque las redes neuronales convolucionales han

demostrado un rendimiento notable en diversas tareas, su implementación conlleva **altos costes computacionales** y una gran demanda de datos, lo que supone un obstáculo importante en muchos escenarios, especialmente aquellos con recursos limitados. Una estrategia prometedora consiste en entrenar estos modelos utilizando únicamente una fracción de los datos disponibles, seleccionados de manera óptima. Esta aproximación permitiría disminuir significativamente el consumo de recursos sin comprometer la precisión del modelo, lo que supondría un avance importante para la **inteligencia artificial**, en particular en aplicaciones donde existen **restricciones de recursos**.

En este contexto, la selección inteligente de subconjuntos representativos constituye una estrategia eficaz para reducir tanto los tiempos de entrenamiento como el uso de recursos, sin afectar negativamente el rendimiento. Es precisamente en este punto donde las **metaheurísticas** adquieren un papel relevante. Estas técnicas de optimización están diseñadas para abordar problemas complejos en los que los métodos tradicionales resultan ineficaces, gracias a sus **estrategias de búsqueda y exploración del espacio de soluciones**. Al combinar diversas heurísticas, permiten encontrar soluciones aproximadas en tiempos razonables, lo que las convierte en una herramienta especialmente útil cuando la obtención de una solución exacta resulta inviable desde el punto de vista computacional.

Por tanto, las metaheurísticas se posicionan como una alternativa sólida para mejorar la eficiencia y accesibilidad del aprendizaje profundo, incluso en contextos con recursos limitados. Esto es especialmente relevante para avanzar hacia una inteligencia artificial más abierta, **democrática** y aplicable en una mayor variedad de escenarios.

Este Trabajo de Fin de Grado tiene como objetivo aplicar técnicas metaheurísticas para realizar una selección inteligente de ejemplos, con el fin de reducir el tamaño de los conjuntos de entrenamiento sin que ello afecte significativamente a los resultados obtenidos. La investigación aspira a contribuir al desarrollo de modelos más **eficientes**, accesibles y económicamente sostenibles, fomentando así un futuro en el que la inteligencia artificial sea más **inclusiva y sostenible**.

1.3. Objetivos

El objetivo principal de este TFG es investigar la aplicación de **metaheurísticas** para la **reducción de conjuntos de datos de entrenamiento** en modelos de **aprendizaje profundo convolucionales**. Este estudio permitirá evaluar el impacto de dichos algoritmos en la **eficiencia computacional** y en el **rendimiento de los modelos**. Para lograr estos objetivos,

se desarrollarán y compararán distintas técnicas de selección de instancias aplicadas sobre modelos convolucionales, incluyendo enfoques aleatorios, de búsqueda local, genéticos y meméticos.

Para cumplir con este objetivo general, se plantean los siguientes **objetivos específicos**:

- **Estudiar** la conveniencia del uso de metaheurísticas para la reducción de conjuntos de imágenes en modelos de aprendizaje profundo.
- **Desarrollar y mejorar** algoritmos metaheurísticos orientados a la selección de ejemplos representativos, con el objetivo de optimizar el entrenamiento de modelos convolucionales sin comprometer su rendimiento.
- **Evaluar** el impacto de la reducción de datos en el rendimiento de los modelos, comparando aspectos clave como la precisión, eficacia y el tiempo de entrenamiento, en modelos entrenados con conjuntos de datos completos frente a conjuntos reducidos.
- **Comparar** el rendimiento de metaheurísticas con porcentajes de selección fijos frente a aquellos que permiten una selección flexible del número de ejemplos, analizando sus ventajas y limitaciones.

A través de este estudio, se busca no solo mejorar el rendimiento y la eficiencia de los modelos convolucionales, sino también fomentar el desarrollo de soluciones más **sostenibles y accesibles** en el ámbito de la inteligencia artificial, mediante la incorporación de estrategias metaheurísticas como **propuestas innovadoras** para la reducción de datos de entrenamiento.

Capítulo 2

Metodología y Presupuesto

2.1. Metodología

Para organizar el desarrollo del proyecto se optó por una metodología inspirada en el marco ágil **Scrum** [1], ampliamente utilizado en entornos donde se requiere adaptación constante y entregas progresivas. Dado que el desarrollo del trabajo se ajustaba progresivamente según los resultados obtenidos en cada fase, este enfoque demostró ser adecuado para mantener una estructura flexible y favorecer un avance iterativo

2.1.1. Enfoque Ágil Basado en Scrum

El proyecto se dividió en ciclos de trabajo breves (**sprints**), generalmente de dos semanas. Cada sprint comenzaba con una planificación en la que se establecían objetivos concretos y finalizaba con una revisión para evaluar el progreso y hacer ajustes si era necesario. Esta dinámica permitió mantener un ritmo constante, al tiempo que se dejaba margen para adaptarse a posibles imprevistos o nuevas ideas surgidas durante el desarrollo.

2.1.2. Organización de Tareas

Para la organización y seguimiento de tareas se empleó una herramienta digital de apoyo (Notion [2]), que facilitó la planificación personal y visualización de actividades pendientes.

Este tipo de plataformas contribuyen a mejorar la organización y la colaboración en proyectos complejos, tal como se ha demostrado en entornos profesionales mediante el uso de tecnologías de la información orientadas a la gestión colaborativa [3].

Esta plataforma facilitó la visualización de las actividades pendientes y cumplidas, lo que ayudó a no perder de vista los plazos y prioridades de cada fase.

En paralelo, se empleó **Git** [4] para el control de versiones del código, lo que resultó fundamental para mantener un seguimiento detallado de los cambios realizados en cada etapa del proyecto. También se utilizó **GitHub** como repositorio central, facilitando así la colaboración con el tutor y permitiendo una trazabilidad clara de todas las modificaciones. El código fuente completo y los recursos del proyecto están disponibles en el repositorio: <https://github.com/JoseRuizLopez/TFG>.

2.1.3. Ciclos de Trabajo

Los sprints se estructuraron en varias fases repetidas en cada iteración:

- **Planificación:** En esta etapa se definían los objetivos del sprint, basándose en lo ya realizado y en las tareas pendientes más prioritarias. El análisis del **backlog** permitía seleccionar actividades realistas que se pudieran completar en el plazo previsto.
- **Desarrollo e implementación:** Se ejecutaban las tareas previstas, como la programación de nuevos módulos, mejoras en algoritmos o la integración de componentes. El enfoque fue incremental, es decir, se añadían funcionalidades poco a poco, asegurando que cada avance se pudiera probar por separado.
- **Pruebas y ajustes:** Una vez desarrolladas las funcionalidades, se realizaban pruebas (unitarias, de integración o empíricas) para validar el funcionamiento del sistema y ajustar los parámetros si fuese necesario. A veces, los resultados de esta fase daban lugar a nuevas tareas que se añadían al backlog.
- **Revisión y análisis de resultados:** Al finalizar el sprint, se revisaban los objetivos cumplidos, se analizaban los resultados obtenidos y se valoraba la necesidad de replantear enfoques. Esta revisión también ayudaba a identificar puntos de mejora y reforzar los que ya funcionaban bien.
- **Documentación:** Durante todo el proceso se fue registrando la evolución del proyecto, desde los cambios en el código hasta los resultados de las pruebas. Este esfuerzo permitió tener una memoria bien estructurada, facilitar la reproducción de experimentos y justificar cada decisión tomada.

2.1.4. Seguimiento del Progreso



Figura 2.1: Diagrama de Gantt que muestra la planificación de dos sprints del proyecto, con sus respectivas fases: planificación, desarrollo, pruebas, revisión y documentación. Este esquema representa la estructura iterativa adoptada a lo largo del desarrollo

Para tener una visión clara del avance del proyecto, se elaboró un **diagrama de Gantt** [2.1] que recogía dos sprints como ejemplo del proceso general. En él se reflejan las distintas fases mencionadas, así como la duración aproximada de cada una. Aunque se realizaron más sprints a lo largo del trabajo, este diagrama sirve como muestra del esquema de planificación utilizado.

2.1.5. Ajuste de los Tiempos

Durante el desarrollo surgieron algunos retrasos que obligaron a reajustar los tiempos previstos para los sprints. Gracias a la flexibilidad que permite **Scrum** y al uso de herramientas como **Git**, permitió reorganizar prioridades y redistribuir tareas sin comprometer los objetivos del proyecto sin comprometer la calidad del trabajo.

2.2. Presupuesto

El presupuesto estimado incluye tanto el tiempo dedicado al proyecto (mano de obra) como los recursos computacionales y otros gastos relacionados. Aunque muchos recursos utilizados son gratuitos o de bajo coste, se ha realizado una estimación considerando el valor del tiempo invertido y el uso de recursos computacionales.

2.2.1. Mano de obra

El proyecto ha requerido un esfuerzo considerable distribuido entre tareas de análisis, desarrollo e implementación técnica, así como documentación y validación experimental. En total, se estima que se han dedicado aproximadamente **400 horas** al trabajo, repartidas de forma flexible a lo largo del calendario del TFG.

La estimación del coste de la mano de obra se ha calculado a partir de una tarifa de **20 euros** por hora, en línea con valores orientativos para trabajos técnicos similares [5].

A continuación se detallan las tareas principales:

- **Planificación inicial y diseño del proyecto:** Establecimiento de objetivos, enfoque metodológico y cronograma.
- **Repaso bibliográfico:** Revisión de trabajos previos relacionados con algoritmos meméticos y reducción de datos en aprendizaje profundo.
- **Comprensión del problema y formulación:** Definición precisa del problema y formalización de los objetivos computacionales.
- **Selección de datasets y preprocesado:** Búsqueda, limpieza y organización de los conjuntos de datos utilizados.
- **Implementación inicial de algoritmos:** Desarrollo de versiones base de los algoritmos genético, memético y de búsqueda local.
- **Desarrollo de versiones mejoradas:** Ajustes iterativos, experimentación con nuevas variantes y mecanismos de exploración.
- **Evaluación experimental y análisis de resultados:** Ejecución masiva de pruebas, análisis comparativo y visualización de métricas.
- **Generación de gráficos:** Creación de boxplots, barplots y figuras de evolución del fitness.

- **Escritura de la memoria del TFG:** Redacción estructurada del documento final, incluyendo capítulos técnicos, resultados y conclusiones.
- **Revisión y corrección:** Mejora del contenido tras la retroalimentación del tutor, ajuste de redacción, formato y bibliografía.

Tarea realizada	Horas dedicadas	Coste por hora (€)	Coste total (€)
Planificación inicial y diseño del proyecto	20	20	400
Repaso bibliográfico	30	20	600
Comprensión del problema y formulación	25	20	500
Selección de datasets y preprocesado inicial	20	20	400
Implementación inicial de algoritmos	40	20	800
Desarrollo de propuestas y versiones mejoradas	50	20	1.000
Evaluación experimental y análisis de resultados	60	20	1.200
Generación de gráficos	30	20	600
Escritura de la memoria del TFG	80	20	1.600
Revisión y corrección final del documento	45	20	900
Total	400 horas		8.000 €

Tabla 2.1: Estimación detallada del coste de la mano de obra distribuido por tareas del proyecto.

A modo de resumen, en la Tabla 2.1 se presenta una distribución detallada de las horas dedicadas a cada una de estas tareas, junto con el coste estimado asociado. Esta descomposición permite visualizar de forma clara cómo se ha estructurado el esfuerzo invertido, diferenciando tareas de investigación, desarrollo técnico y documentación. El resultado total asciende a 8.000 €, lo que proporciona una visión aproximada del valor económico del trabajo llevado a cabo.

2.2.2. Recursos computacionales

El proyecto ha requerido el uso de recursos computacionales para entrenar los modelos de aprendizaje profundo, especialmente para evaluar la eficacia de los algoritmos en la reducción de datos.

El tutor **Daniel Molina Cabrera** me puso en disposición el acceso a un servidor de investigación con disponibilidad de GPU, de manera que no ha supuesto ningún coste.

Aunque el acceso al servidor de investigación con GPU no implicó coste real, se ha realizado una estimación hipotética utilizando precios de **Google Cloud** [6] usando un **Compute Engine** [7] en Bélgica para valorar el uso de recursos.

Los componentes equivalentes del servidor en Compute Engine serían:

- Un **Intel Xeon E-2226G** equivaldría a un **c2-standard-8**, cuyo coste es de 0.43 EUR/h.

- Un **NVIDIA TITAN Xp** equivaldría a una **NVIDIA K80 1 GPU**, GDDR5 de 12 GB cuyo coste es de 0.42 EUR/h.

De modo que los gastos estimados serían:

Recurso	Horas utilizadas	Coste por hora (€)	Coste total (€)
CPU (c2-standard-8)	600	0.43	258
GPU (NVIDIA K80 1 GPU)	600	0.42	252
Total	600	0.85	510

Tabla 2.2: Simulación del coste de recursos computacionales equivalentes en Google Cloud.

2.2.3. Software y licencias

Para este proyecto, todas las herramientas de desarrollo utilizadas han sido de **código abierto**, por lo que no se han incurrido en costes de licencias.

2.2.4. Gastos indirectos

Aunque muchos de los recursos utilizados durante el desarrollo del proyecto han sido gratuitos o provistos por el entorno universitario, se ha considerado el coste indirecto asociado al uso continuado del equipo personal.

En concreto, se estima el coste proporcional derivado del uso de un ordenador portátil **Lenovo Ideapad 530S-14KB**, empleado intensivamente durante todo el curso académico para la programación, análisis y documentación del TFG. El cálculo se ha realizado prorrateando su valor de adquisición estimado (aproximadamente 900 €) a lo largo de una vida útil de 5 años, lo que da lugar a un coste anual aproximado de 180 €.

Concepto	Coste estimado (€)
Uso del portátil (Lenovo Ideapad 530S)	180

Tabla 2.3: Estimación de gastos indirectos asociados al desarrollo del proyecto.

2.2.5. Presupuesto total

Sumando los costes de mano de obra, los recursos computacionales y otros gastos, el presupuesto total estimado es el siguiente:

Concepto	Coste (€)
Mano de obra	8.000
Recursos computacionales	510
Gastos indirectos	180
Total	8.690 €

Tabla 2.4: Resumen del presupuesto total estimado, incluyendo mano de obra, recursos y gastos indirectos.

En conjunto, la planificación metodológica y la estimación de recursos permitieron garantizar el avance ordenado del proyecto y anticipar su viabilidad técnica y económica.

Capítulo 3

Fundamentos de Aprendizaje Profundo

El presente capítulo tiene como objetivo proporcionar una base teórica sólida sobre los conceptos clave del **aprendizaje profundo**, que constituyen el núcleo metodológico del presente Trabajo de Fin de Grado. Dado que el proyecto se centra en optimizar el entrenamiento de modelos de redes neuronales convolucionales mediante técnicas de reducción de datos, resulta fundamental comprender el funcionamiento interno de este tipo de modelos y sus componentes principales.

Esta revisión técnica permitirá comprender mejor los retos que supone entrenar dichos modelos y la motivación detrás de aplicar algoritmos meméticos para optimizar el proceso.

3.1. Definición de Aprendizaje Profundo

El **aprendizaje profundo** (Deep Learning) [8] es una subcategoría del aprendizaje automático que se basa en el uso de **redes neuronales artificiales** con muchas capas (de ahí el término “profundo”). Estas redes están diseñadas para imitar el funcionamiento del cerebro humano, lo que les permite aprender representaciones complejas de los datos de manera jerárquica.

La principal diferencia entre el **aprendizaje automático tradicional** y el aprendizaje profundo es la manera en que se manejan las características de los datos:

- En los enfoques tradicionales, el ingeniero o científico de datos debe extraer manualmente las características más importantes para entrenar al modelo (por ejemplo, bordes, formas, texturas en imágenes). En el

aprendizaje profundo, las redes neuronales son capaces de **aprender automáticamente las representaciones de los datos** a partir de los datos crudos (por ejemplo, imágenes, texto, sonido).

- Este proceso es denominado **aprendizaje de características** (feature learning), lo que reduce la necesidad de intervención humana.

El aprendizaje profundo ha mostrado un rendimiento sobresaliente en diversas tareas, como el reconocimiento de imágenes, el procesamiento del lenguaje natural, la conducción autónoma y el diagnóstico médico, gracias a su capacidad para **capturar patrones complejos** en grandes volúmenes de datos.

3.2. Redes Neuronales Artificiales

Las **redes neuronales artificiales** (ANN) [9] son el corazón del aprendizaje profundo. Estas redes están compuestas por neuronas artificiales, que son unidades matemáticas inspiradas en las neuronas biológicas. Cada neurona toma varias entradas, las procesa mediante una **función de activación**, y produce una salida. Cuando se combinan muchas de estas neuronas en capas, forman una red neuronal.

3.2.1. Componentes de una Red Neuronal

1. Neuronas o Unidades:

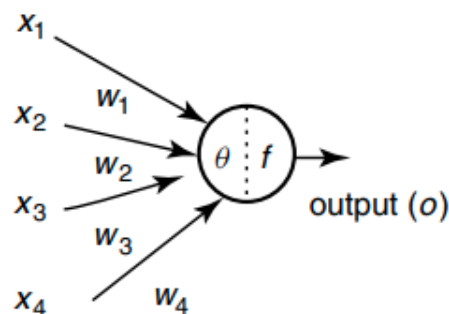


Figura 3.1

Cada **neurona** realiza una operación simple: recibe varias entradas, las pondera por medio de **pesos** w_i , suma estos valores junto con un

sesgo b , y aplica una función de activación. La salida de la neurona se expresa como:

$$z = w_1x_1 + w_2x_2 + \dots + w_nx_n + b_z = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$

Luego, el valor z pasa por una función de activación, que introduce la no linealidad en el sistema, permitiendo que las redes neuronales modelen relaciones complejas.

2. Capas de la Red:

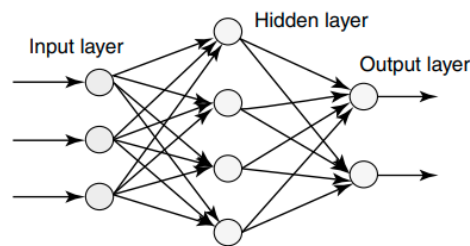


Figura 3.2

- **Capa de entrada:** Es la primera capa de la red neuronal, que recibe los datos crudos (por ejemplo, píxeles de una imagen).
 - **Capas ocultas:** Estas capas intermedias entre la entrada y la salida aprenden representaciones abstractas de los datos. En una red profunda, hay múltiples capas ocultas, lo que permite la **transformación jerárquica** de los datos.
 - **Capa de salida:** Produce la predicción final, que puede ser una clase (en problemas de clasificación) o un valor numérico (en problemas de regresión).
3. **Pesos y Bias:** Los **pesos** son parámetros ajustables que determinan la importancia de cada entrada en la neurona. El **bias** es otro parámetro que se suma al valor ponderado para desplazar la activación de la neurona y permitir que el modelo ajuste mejor los datos.
4. **Funciones de Activación:** Las funciones de activación son fundamentales para que las redes neuronales puedan aprender relaciones no lineales. Entre las más comunes se encuentran:
- **ReLU (Rectified Linear Unit):** $ReLU(x) = \max(0, x)$, que activa solo valores positivos.
 - **Sigmoide:** Que transforma los valores en un rango entre 0 y 1.
 - **Tanh (Tangente hiperbólica):** Transforma los valores en un rango entre -1 y 1.

El uso de **backpropagation** o retropropagación permite ajustar los pesos y sesgos durante el entrenamiento mediante un algoritmo de optimización, como el descenso de gradiente. De esta manera, la red aprende minimizando la diferencia entre sus predicciones y las respuestas correctas.

3.3. Redes Neuronales Convolucionales

Las **Redes Neuronales Convolucionales** (Convolutional Neural Networks, CNN) son una clase de redes neuronales profundas especialmente efectivas para el procesamiento de datos que tienen una estructura de tipo rejilla, como las imágenes. Fueron inspiradas por el sistema visual de los mamíferos, donde diferentes capas de neuronas responden a estímulos visuales de manera jerárquica.

Las CNN son ampliamente utilizadas en tareas de **visión por computador**, como el reconocimiento de imágenes, la segmentación de objetos y la clasificación de imágenes. Lo que diferencia a las CNN de las redes neuronales tradicionales es su capacidad para detectar **patrones espaciales** como bordes, texturas, y formas, sin necesidad de un procesamiento manual de las características.

3.3.1. Componentes principales de una CNN

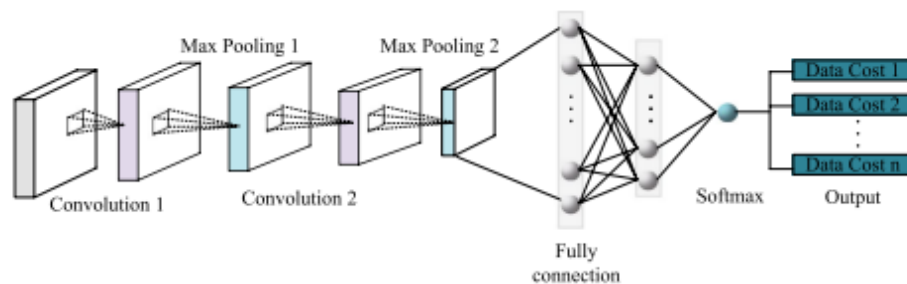


Figura 3.3: En esta imagen extraída de [10] puede observarse de la estructura de una red neuronal convolucional. Junto a sus capas convolucionales.

1. **Capas Convolucionales:** Estas capas aplican **filtros o kernels** sobre las imágenes de entrada para detectar características locales, como bordes, esquinas o texturas. Un filtro convolucional es una pequeña matriz que se mueve a lo largo de la imagen, calculando productos escalares en cada posición para producir un mapa de características.

Las convoluciones son útiles porque explotan la **localidad de las características**, es decir, las relaciones espaciales entre píxeles cercanos. Además, la cantidad de parámetros se reduce drásticamente en comparación con las capas densas, ya que el filtro se comparte a lo largo de la imagen.

2. **Pooling (Submuestreo o Agrupamiento)**: Las capas de pooling reducen la dimensionalidad de las características extraídas por las capas convolucionales, lo que hace que las representaciones sean más manejables y robustas frente a pequeños cambios o desplazamientos en la imagen.

El max-pooling es la técnica de pooling más común, donde se toma el valor máximo dentro de una ventana de píxeles, reduciendo el tamaño de la imagen, pero reteniendo las características más importantes.

3. **Capas Densas**: Después de varias capas convolucionales y de pooling, se agregan una o más **capas densas** (fully connected) para realizar la clasificación o predicción final. Estas capas toman todas las características aprendidas en las capas convolucionales y las combinan para generar una decisión final.
4. **Batch Normalization**: Esta técnica se utiliza para **normalizar** las salidas de las capas intermedias de una red neuronal. Batch Normalization ayuda a **acelerar el entrenamiento** y a hacer que la red sea más estable, al reducir el **desplazamiento covariante** (cambios en las distribuciones de las entradas de las capas intermedias a lo largo del entrenamiento). Esto se logra al normalizar las entradas de cada capa convolucional o densa antes de aplicar la activación, ajustando su media y varianza.
5. **Dropout**: El Dropout es una técnica de **regularización** que se utiliza para prevenir el **sobreajuste** (overfitting) durante el entrenamiento de una red neuronal. Durante cada iteración del entrenamiento, Dropout **desactiva aleatoriamente** un porcentaje de las neuronas, lo que obliga a la red a no depender excesivamente de ciertas neuronas y a ser más robusta. Esta técnica mejora la generalización de la red, lo que la hace funcionar mejor en datos no vistos.

3.3.2. Funcionamiento General de una CNN

Al pasar una imagen a través de varias capas convolucionales, la red aprende a identificar características simples como líneas y bordes. Conforme avanza a capas más profundas, las características se vuelven más abstractas, capturando patrones más complejos como formas, texturas y, finalmente, estructuras completas como objetos.

Por ejemplo, en una red entrenada para reconocer caras, las primeras capas pueden detectar bordes o contornos, las capas intermedias pueden aprender a reconocer ojos, nariz o boca, y las últimas capas pueden identificar una cara completa.

3.3.3. Aplicaciones de las CNN

- **Clasificación de imágenes:** Etiquetar imágenes en distintas categorías, como identificar animales o vehículos.
- **Detección de objetos:** Identificar y localizar objetos en imágenes.
- **Reconocimiento facial:** Utilizado en sistemas de seguridad, como el desbloqueo de teléfonos móviles.

Las CNN son fundamentales en muchas aplicaciones modernas debido a su capacidad para procesar y entender datos visuales de manera eficiente y automática.

3.4. Modelos

En el ámbito del aprendizaje profundo, existen diversas arquitecturas de redes neuronales convolucionales que han demostrado un rendimiento excepcional en diversas tareas de visión por computadora. Estas arquitecturas están diseñadas para abordar problemas complejos y variados, desde la clasificación de imágenes hasta la detección de objetos y el segmentado de imágenes.

A continuación, se explorarán algunas de estas arquitecturas que representan avances significativos en la eficiencia y efectividad del aprendizaje profundo.

3.4.1. ResNet50

ResNet50 [11] es una arquitectura de red neuronal convolucional introducida por Kaiming He et [12]. La principal innovación de ResNet es la introducción de **bloques de residualidad**, que permiten la construcción de redes extremadamente profundas sin el problema de la degradación del rendimiento.

La idea básica de los bloques residuales es permitir que la red aprenda funciones de identidad, facilitando así la propagación de la información y el gradiente a través de la red. En términos prácticos, esto se traduce

en un rendimiento mejorado en tareas de clasificación de imágenes, donde ResNet50 ha logrado resultados sobresalientes en competencias como ImageNet [13]. Esta arquitectura se compone de 50 capas, de las cuales 49 son convolucionales y una es totalmente conectada, incluyendo capas de normalización y activación (ReLU), así como conexiones residuales que facilitan el entrenamiento de redes profundas.

3.4.2. MobileNetV2



Figura 3.4: Diagrama de la Arquitectura de ResNet50

MobileNet [14] es una arquitectura de red neuronal diseñada específicamente para aplicaciones móviles y de visión por computadora en dispositivos con recursos limitados. Introducida por Andrew G. Howard et al. en 2017, MobileNet se basa en el principio de **convoluciones separables en profundidad** (depthwise separable convolutions), que dividen el proceso de convolución en dos pasos: primero, se aplica una convolución a cada canal de la entrada (depthwise), y luego, se combinan los resultados con una convolución 1x1 (pointwise).

Esta técnica reduce significativamente el número de parámetros y el costo computacional, lo que permite ejecutar modelos de visión por computadora en dispositivos móviles sin sacrificar drásticamente la precisión.

MobileNetV2

MobileNetV2 [15], introducido por Sandler et al. en 2018, mejora la arquitectura de MobileNet original al incorporar varias innovaciones. La principal contribución de MobileNetV2 es la introducción de los bloques de **residualidad invertida** (inverted residuals), que permiten que la red



Figura 3.5: Diagrama de la Arquitectura de MobileNetV2

mantenga una mayor capacidad de representación y flujo de información a través de las capas.

Además, MobileNetV2 utiliza una función de activación llamada **linear bottleneck**, que ayuda a preservar la información durante la propagación a través de las capas, lo que mejora aún más el rendimiento del modelo en tareas de clasificación y detección. Esta arquitectura se optimiza para ser altamente eficiente, permitiendo que sea utilizada en aplicaciones de tiempo real en dispositivos con limitaciones de hardware.

MobileNetV2 ha demostrado ser una opción popular para aplicaciones en dispositivos móviles y sistemas embebidos, ofreciendo un buen equilibrio entre precisión y eficiencia computacional.

Capítulo 4

Repaso Bibliográfico

En este capítulo se revisa la literatura relevante en torno a los modelos de **aprendizaje profundo**, la necesidad de reducir los conjuntos de **datos de entrenamiento** y el rol de los **algoritmos meméticos** en esta tarea. A lo largo de este capítulo, se analizan estudios previos que aplican técnicas de **optimización** para reducir los datos necesarios en el entrenamiento de **redes neuronales convolucionales** (CNN) y se exploran las soluciones existentes en la selección de instancias mediante algoritmos **evolutivos** y **meméticos**.

4.1. Reducción de Conjuntos de Datos en Aprendizaje Profundo

El **aprendizaje profundo**, especialmente en el campo de la **visión por computadora**, se ha basado en grandes volúmenes de **datos** para alcanzar su éxito. En particular, las redes neuronales convolucionales (CNN) han demostrado un rendimiento notable en tareas de **clasificación**, **detección** y **segmentación** de imágenes [16]. Sin embargo, el volumen de datos necesario para entrenar estos modelos supone retos en cuanto a la disponibilidad de **almacenamiento**, el **tiempo de entrenamiento** y el **costo computacional**, especialmente en sistemas con **recursos limitados** [17].

La necesidad de reducir estos volúmenes de datos sin comprometer la **precisión** del modelo ha impulsado investigaciones en torno a técnicas de **selección de instancias**, donde el objetivo es identificar y mantener solo las instancias de datos que aportan valor al modelo. A este respecto, las técnicas de selección de instancias combinadas con estrategias de **aumento de datos** [18] permiten reducir el tamaño de los conjuntos de datos manteniendo una **diversidad** adecuada, lo cual es crucial para evitar el **so-**

breajuste y la pérdida de **generalización** en las redes neuronales. Además, otros enfoques, como el **submuestreo de datos** y la generación de **conjuntos sintéticos**, se han propuesto como soluciones complementarias en el contexto de aprendizaje profundo, siendo útiles en escenarios con conjuntos de datos desequilibrados o insuficientes.

4.2. Selección de Instancias mediante Algoritmos Evolutivos y Meméticos

La **selección de instancias** es una técnica de reducción de datos que se centra en eliminar aquellas instancias **redundantes** o **irrelevantes**, mejorando la **eficiencia** y manteniendo la **precisión** en el modelo entrenado. Los **algoritmos evolutivos**, como los **algoritmos genéticos**, han demostrado eficacia en la selección de instancias al simular procesos de **selección natural**, con la ventaja de que pueden explorar un espacio de **soluciones** de manera eficiente a través de **operadores genéticos** como la **selección**, el **cruce** y la **mutación** [19].

Dentro de este campo, los **algoritmos meméticos** ofrecen un avance significativo al combinar los principios de **optimización evolutiva** con técnicas de **búsqueda local**, lo que permite una adaptación más precisa de las instancias seleccionadas [20]. Estos algoritmos representan un enfoque **híbrido**, ya que aprovechan la capacidad **exploratoria** de los algoritmos evolutivos con la capacidad **explotatoria** de las búsquedas locales, lo cual es crucial para alcanzar una **convergencia óptima**. Al introducir esta **dualidad**, los algoritmos meméticos pueden enfocarse en subconjuntos de datos más representativos y potencialmente relevantes para el modelo, maximizando la reducción sin comprometer la calidad del aprendizaje.

4.3. Algoritmos Meméticos en Modelos Convolucionales

La capacidad de los **algoritmos meméticos** para optimizar la selección de datos se ha aplicado a **modelos convolucionales**, en particular en contextos de redes profundas que requieren grandes cantidades de datos para alcanzar un rendimiento óptimo [21]. La selección de instancias mediante algoritmos meméticos permite mantener la precisión del modelo, pero con un volumen de datos significativamente reducido, lo cual es especialmente útil en aplicaciones donde los recursos computacionales y el tiempo son limitados [22].

Algunos estudios han demostrado que los algoritmos meméticos no solo

reducen los tiempos de **entrenamiento** y el tamaño de los conjuntos de **datos**, sino que también ayudan a minimizar el riesgo de **sobreajuste**. Al reducir los datos redundantes, el modelo **convolucional** puede centrarse en **patrones** más específicos, potenciando su capacidad de **generalización** y **robustez**. Estas ventajas son especialmente valiosas en escenarios como el **reconocimiento facial** o la **medicina**, donde la precisión y la eficiencia en el procesamiento de imágenes son esenciales.

4.4. Aplicaciones y Desafíos de los Algoritmos Meméticos

A pesar de sus beneficios, la aplicación de **algoritmos meméticos** en modelos de aprendizaje profundo plantea desafíos. Uno de los principales es la necesidad de ajustar **parámetros complejos**, como el tamaño de la **población**, las tasas de **mutación** y los métodos de **selección** y **cruce**. Estos parámetros afectan de manera directa la **velocidad de convergencia** y la capacidad del algoritmo para encontrar soluciones **óptimas**. Además, los algoritmos meméticos suelen ser computacionalmente **exigentes**, lo que plantea una paradoja cuando se aplican en entornos con **recursos limitados** [23].

Para abordar estos desafíos, investigaciones recientes han explorado el desarrollo de **variantes de algoritmos meméticos adaptativos**, que ajustan automáticamente sus parámetros en función del desempeño durante el proceso de **entrenamiento**. Esta **adaptabilidad** representa una vía prometedora, ya que permite que los algoritmos se adapten dinámicamente a los cambios en el entorno de datos y en las necesidades del modelo, facilitando así su implementación en aplicaciones prácticas.

4.5. Perspectivas Futuras en la Optimización de CNN con Algoritmos Meméticos

La combinación de **CNN** y **algoritmos meméticos** sigue siendo un área emergente y prometedora en la investigación de **redes neuronales profundas**. A medida que la tecnología y la **capacidad de procesamiento** evolucionan, es probable que los algoritmos meméticos se integren de manera más fluida en **arquitecturas de aprendizaje profundo**, no solo para la selección de instancias, sino también para la **optimización de hiperparámetros** y **estructuras de red**. Asimismo, se anticipa que la investigación futura también explore la combinación de estos métodos con otros enfoques avanzados de **reducción de datos**, como las técnicas de

distilación de modelos y la poda de redes neuronales.

En conclusión, este capítulo ha establecido las bases teóricas que sustentan la selección de **algoritmos meméticos** para reducir conjuntos de datos en **redes profundas**, resaltando su relevancia en la optimización de **modelos CNN** en entornos de recursos limitados.

Capítulo 5

Descripción de los Algoritmos

En este capítulo se describen los diferentes algoritmos utilizados en el desarrollo de este trabajo. Todos ellos tienen como objetivo principal reducir el tamaño del conjunto de datos de entrenamiento utilizado en los modelos de aprendizaje profundo, con el fin de optimizar el rendimiento y reducir el costo computacional. El enfoque adoptado en este trabajo es la aplicación de algoritmos meméticos, los cuales combinan principios de algoritmos genéticos con estrategias de búsqueda local.

A continuación, se detallan los algoritmos principales implementados en este proyecto: el **algoritmo aleatorio**, el **algoritmo de búsqueda local**, el **algoritmo genético** y el **algoritmo memético**.

5.1. Algoritmo Aleatorio

El **algoritmo aleatorio** sirve como referencia básica para medir la efectividad de los algoritmos más avanzados. Este enfoque selecciona subconjuntos de datos de manera completamente aleatoria, sin aplicar ningún tipo de estrategia de optimización.

5.1.1. Descripción

El algoritmo comienza tomando el conjunto de datos completo y seleccionando una fracción de los ejemplos de entrenamiento de forma aleatoria. Esta selección se realiza sin ningún criterio basado en la relevancia de los datos, lo que implica que el conjunto de entrenamiento resultante puede no ser representativo o puede contener redundancias innecesarias.

5.1.2. Pseudocódigo

Algorithm 1 Algoritmo Aleatorio

```
1: Inicializar evaluaciones  $\leftarrow 0$ 
2: while evaluaciones < máximo do
3:   Generar una selección aleatoria
4:   Evaluar la selección
5:   if mejora el mejor fitness global then
6:     Guardar como mejor solución
7:     Reiniciar contador sin mejora
8:   else
9:     Incrementar contador sin mejora
10:  end if
11:  Incrementar número de evaluaciones
12:  if se supera el umbral de estancamiento then
13:    Terminar la búsqueda
14:  end if
15: end while
16: return mejor solución e historial
```

Para facilitar su lectura o replicación, se ha detallado el pseudocódigo del **Algorithm 1** Algoritmo Aleatorio.

5.1.3. Aplicación en la reducción de datos

A pesar de su simplicidad, el **algoritmo aleatorio** puede ser útil como método de comparación. En muchos casos, los algoritmos más complejos deben demostrar que pueden superar este enfoque básico en términos de precisión y eficiencia. Al seleccionar datos de manera aleatoria, este método a menudo produce conjuntos de entrenamiento subóptimos, lo que resulta en modelos menos precisos o con mayor varianza.

5.1.4. Resultados esperados

Debido a la naturaleza aleatoria del algoritmo, los resultados son altamente variables. Es probable que en muchas ejecuciones el rendimiento del modelo entrenado sea inferior al obtenido con métodos más estructurados. Este algoritmo proporciona una línea base importante para evaluar la efectividad de los algoritmos más avanzados.

5.2. Algoritmo Búsqueda Local

El **algoritmo de búsqueda local** es una técnica más sofisticada que explora el espacio de soluciones de manera más estructurada, buscando mejorar progresivamente una solución inicial.

5.2.1. Descripción

La búsqueda local se basa en la idea de comenzar con una solución inicial (un subconjunto de datos) y realizar pequeños cambios o ‘movimientos’ en esa solución para explorar otras soluciones cercanas. En este contexto, cada solución es un subconjunto de datos. El algoritmo evalúa diferentes subconjuntos de datos probando si estos mejoran el rendimiento del modelo de aprendizaje profundo al entrenarlo con ellos.

5.2.2. Pseudocódigo

Algorithm 2 Algoritmo de Búsqueda Local

```
1: Generar una solución inicial aleatoria
2: Evaluar su fitness como mejor_fitness
3: while evaluaciones < máximo do
4:   Generar un vecino de la solución actual
5:   Evaluar el fitness del vecino
6:   if el vecino mejora o iguala la solución actual then
7:     Reemplazar la solución actual
8:     if mejora el mejor_fitness then
9:       Actualizar mejor solución
10:    Reiniciar contador sin mejora
11:  else
12:    Incrementar contador sin mejora
13:  end if
14: else
15:   Incrementar contador sin mejora
16: end if
17: Incrementar evaluaciones
18: if hay estancamiento then
19:   Terminar búsqueda
20: end if
21: end while
22: return mejor solución
```

Para facilitar su lectura o replicación, se ha detallado el pseudocódigo

del **Algorithm 2** Algoritmo Búsqueda Local.

5.2.3. Aplicación en la reducción de datos

En el contexto de la reducción de datos, el objetivo de la búsqueda local es identificar un subconjunto más pequeño de ejemplos que sea suficiente para entrenar el modelo con un rendimiento similar al obtenido con el conjunto de datos completo. La búsqueda local explora el espacio de posibles subconjuntos, eliminando ejemplos redundantes o irrelevantes, y conservando solo aquellos que son cruciales para el rendimiento del modelo.

5.2.4. Ventajas y limitaciones

Ventajas: Este enfoque permite una exploración más exhaustiva del espacio de soluciones que un algoritmo aleatorio. Al hacer pequeños ajustes en cada iteración, el algoritmo puede encontrar mejores soluciones de manera eficiente.

Limitaciones: Sin embargo, la búsqueda local puede quedarse atrapada en **óptimos locales**, es decir, soluciones que parecen buenas en comparación con las cercanas, pero que no son globalmente óptimas.

5.3. Algoritmo Genético

5.3.1. Descripción

El algoritmo genético es una técnica de optimización inspirada en los principios de la evolución natural. Parte de una población inicial de soluciones candidatas (en este caso, subconjuntos de imágenes) que evoluciona generación tras generación mediante operadores de selección, cruce y mutación. Cada una de estas soluciones se evalúa con respecto a su aptitud (fitness), que se calcula según la métrica de rendimiento del modelo entrenado con dicho subconjunto (accuracy).

En esta versión inicial, el algoritmo utiliza una implementación sencilla pero funcional, diseñada para validar los beneficios de la evolución artificial sobre enfoques aleatorios o deterministas. A pesar de su simplicidad, esta versión ya logra una exploración significativa del espacio de soluciones, sirviendo como base para versiones posteriores más sofisticadas.

5.3.2. Componentes del Algoritmo

A continuación, se describen los principales componentes que conforman esta primera versión del algoritmo genético:

Población inicial

Se genera aleatoriamente una colección de subconjuntos de datos, donde cada individuo representa una selección binaria de imágenes. Este conjunto inicial sirve como punto de partida para la evolución del sistema.

Evaluación

Cada individuo se evalúa entrenando un modelo convolucional sobre el subconjunto de imágenes representado. El valor de *fitness* se calcula utilizando la métrica *accuracy*.

Selección por torneo

Se escogen dos padres de la población mediante el método de torneo. Para ello, se elige aleatoriamente un subconjunto de individuos y se selecciona el que tenga mayor *fitness*. Este proceso simula una competencia entre soluciones y prioriza las más prometedoras.

Cruce

Se aplica un operador de cruce básico que intercambia parte de las imágenes entre ambos padres. Este operador garantiza que el número total de imágenes seleccionadas en los hijos se mantenga constante, preservando la estructura del problema.

Mutación

Cada hijo generado puede sufrir mutaciones aleatorias. Con una probabilidad fija, algunas imágenes cambian su estado (de seleccionadas a no seleccionadas y viceversa). Este operador introduce diversidad genética en la población, evitando el estancamiento evolutivo.

Elitismo

Para asegurar que no se pierdan buenas soluciones, el mejor individuo de cada generación se conserva sin modificación. Esta estrategia protege el progreso evolutivo alcanzado hasta el momento.

Criterio de parada

El algoritmo se detiene cuando se alcanza un número máximo de evaluaciones o si no se ha logrado mejorar el mejor fitness durante un número consecutivo de iteraciones. Esto evita ciclos innecesarios y mejora la eficiencia computacional.

5.3.3. Pseudocódigo

Algorithm 3 Algoritmo Genético

```
1: Inicializar población aleatoria
2: Evaluar fitness de cada individuo
3: Guardar el mejor individuo
4: while no se alcance el número máximo de evaluaciones do
5:   Aplicar elitismo
6:   while población incompleta do
7:     Seleccionar dos padres por torneo
8:     Cruzar padres para obtener hijos
9:     Mutar hijos
10:    Evaluar hijos y añadirlos a la nueva población
11:  end while
12:  Reemplazar población
13:  Actualizar mejor individuo si mejora
14: end while
15: return mejor solución
```

Para facilitar su lectura o replicación, se ha detallado el pseudocódigo del **Algorithm 3** Algoritmo Genético.

5.3.4. Aplicación en la Reducción de Datos

El algoritmo genético se adapta de forma natural a la tarea de reducción de datos, ya que cada individuo puede representar explícitamente qué instancias se incluyen o excluyen del conjunto de entrenamiento. Esta representación binaria permite una manipulación sencilla de los subconjuntos y favorece la exploración combinatoria de múltiples configuraciones posibles.

Gracias a su enfoque evolutivo, este algoritmo puede descubrir subconjuntos de imágenes que, aunque significativamente más pequeños que el conjunto completo, mantienen un alto rendimiento del modelo. Además, el uso de operadores estocásticos contribuye a escapar de soluciones triviales o subóptimas.

5.3.5. Ventajas y Limitaciones

Ventajas:

- Flexible y fácilmente ajustable mediante cambios en los operadores.
- Capaz de encontrar buenas soluciones en espacios grandes y discretos.
- Evoluciona de forma progresiva, permitiendo un seguimiento del proceso.

Limitaciones:

- El operador de cruce simple puede resultar limitado, ya que no considera la calidad relativa de los padres.
- Sensible a la convergencia prematura si la diversidad de la población se pierde rápidamente.
- Aunque mejor que la búsqueda aleatoria, puede estancarse en óptimos locales sin mecanismos adicionales como reinicios o cruce ponderado.

5.4. Algoritmo Genético con Cruce Ponderado

5.4.1. Descripción

Esta versión introduce una mejora sobre el algoritmo genético básico al modificar el operador de cruce para que tenga en cuenta la calidad relativa de los padres. En lugar de combinar aleatoriamente los subconjuntos, el cruce ponderado asigna mayor peso al progenitor con mejor fitness, favoreciendo así la herencia de sus características más beneficiosas.

Además, en cada reproducción solo se selecciona el mejor hijo generado por la pareja de padres, descartando el otro. Esta estrategia reduce la generación de soluciones subóptimas, acelera la convergencia y mejora la estabilidad general del proceso evolutivo.

Se mantiene el uso de elitismo, selección por torneo y mutación con probabilidad fija, al igual que en la versión anterior.

5.4.2. Componentes del Algoritmo

A continuación, se describen los principales componentes añadidos o modificados para realizar esta versión con cruce ponderado.

Cruce ponderado

El operador de cruce se basa en el fitness de los padres. Se asigna un mayor número de imágenes al progenitor con mejor rendimiento, lo que incrementa la probabilidad de heredar características positivas. Esta estrategia mejora la calidad de los hijos generados.

Mutación

Se mantiene la mutación aleatoria con probabilidad fija. Su propósito es introducir variabilidad en la población y evitar la convergencia prematura. En este caso, la mutación se aplica después del cruce ponderado.

Selección de hijo único

A diferencia de versiones anteriores, esta implementación evalúa los dos hijos generados y conserva únicamente el mejor. Esta selección más estricta reduce la propagación de soluciones poco prometedoras y acelera la convergencia del algoritmo.

5.4.3. Pseudocódigo

Para facilitar su lectura o replicación, se ha detallado el pseudocódigo del **Algorithm 4** Algoritmo Genético con Cruce Ponderado.

5.4.4. Aplicación en la Reducción de Datos

Este algoritmo permite una selección más refinada de subconjuntos de imágenes representativas al priorizar soluciones de alta calidad durante el cruce. Al favorecer a los progenitores con mejor rendimiento, se acelera la convergencia hacia subconjuntos efectivos, o que resulta en una reducción más precisa y estable del conjunto de entrenamiento.

5.4.5. Ventajas y Limitaciones

Ventajas:

Algorithm 4 Algoritmo Genético con Cruce Ponderado

```
1: Inicializar población y evaluar
2: while no se alcance el máximo de evaluaciones do
3:   Conservar el mejor individuo (elitismo)
4:   while población incompleta do
5:     Seleccionar padres por torneo
6:     Aplicar cruce ponderado
7:     Mutar hijos
8:     Evaluar ambos hijos
9:     Conservar solo el mejor hijo
10:  end while
11:  Reemplazar población
12:  Actualizar mejor si mejora
13: end while
14: return mejor individuo encontrado
```

- Favorece la herencia de características beneficiosas.
- Elimina soluciones débiles al conservar solo el mejor hijo.
- Acelera la convergencia sin sacrificar la calidad de la solución.

Limitaciones:

- Puede reducir la diversidad de la población si las soluciones convergen demasiado rápido.
- Requiere una evaluación adicional por cada pareja de padres (evaluar dos hijos).

5.5. Algoritmo Genético con Mutación Adaptativa

5.5.1. Descripción

Esta versión introduce una mejora centrada en el operador de mutación. En lugar de aplicar un número fijo de modificaciones, se implementa una **mutación adaptativa** que ajusta dinámicamente el número de intercambios en función del tamaño del subconjunto seleccionado.

Esta estrategia permite una exploración más equilibrada del espacio de soluciones, adaptándose mejor a las características de cada individuo. De

este modo, se evita tanto una mutación excesiva que degrade soluciones prometedoras como una mutación insuficiente que impida escapar de óptimos locales.

El resto de componentes del algoritmo se mantienen: selección por torneo, cruce ponderado, elitismo y evaluación del mejor hijo generado. La mutación adaptativa se aplica tras el cruce y antes de la evaluación.

5.5.2. Componentes del Algoritmo

Mutación adaptativa

El operador de mutación ajusta el número de intercambios según el tamaño relativo del subconjunto. Esto se calcula usando la siguiente fórmula:

$$\text{numswaps} = \text{mín}(\text{longitud} \times 0.15, \text{longitud} \times \text{ratio} \times 0.8)$$

donde `longitud` representa el número de imágenes disponibles y `ratio` el porcentaje seleccionado. Así, se consigue una mutación proporcional y contextual.

Preservación de la estructura

A pesar de la variabilidad introducida, se conserva la estructura binaria de selección, evitando cambios abruptos que rompan la proporción de clases o introduzcan ruido excesivo.

5.5.3. Pseudocódigo

Para facilitar su lectura o replicación, se ha detallado el pseudocódigo del **Algorithm 5** Algoritmo Genético con Mutación Adaptativa.

5.5.4. Aplicación en la Reducción de Datos

La mutación adaptativa mejora la capacidad del algoritmo para explorar el espacio de soluciones sin comprometer la calidad de los individuos. Esta versión demostró un mejor equilibrio entre exploración y explotación, especialmente en subconjuntos de tamaño variable o intermedio.

Algorithm 5 Algoritmo Genético con Mutación Adaptativa

```
1: Inicializar población y evaluar
2: while no se alcance el máximo de evaluaciones do
3:   Conservar el mejor individuo (elitismo)
4:   while población incompleta do
5:     Seleccionar padres por torneo
6:     Aplicar cruce ponderado
7:     Aplicar Mutación Adaptativa a los hijos
8:     Evaluar ambos hijos
9:     Conservar solo el mejor hijo
10:  end while
11:  Reemplazar población
12:  Actualizar mejor si mejora
13: end while
14: return mejor individuo encontrado
```

5.5.5. Ventajas y Limitaciones

Ventajas:

- Introduce una mutación más flexible y ajustada al contexto de cada individuo.
- Mejora la estabilidad de las soluciones frente a mutaciones agresivas.
- Favorece la diversidad genética sin perder precisión.

Limitaciones:

- Puede requerir calibración del rango de mutación para casos extremos.
- En subconjuntos muy pequeños, la variabilidad puede verse limitada.

5.6. Algoritmo Genético con Reinicio Poblacional

5.6.1. Descripción

Esta versión del algoritmo genético introduce una estrategia para evitar el estancamiento evolutivo mediante reinicios poblacionales. Cuando el segundo mejor individuo de la población no mejora durante dos generaciones consecutivas, se considera que el algoritmo ha dejado de progresar de forma significativa. En ese caso, se reinicia la población manteniendo únicamente el mejor individuo, y se genera el resto de forma aleatoria.

Este mecanismo permite escapar de óptimos locales y explorar nuevas regiones del espacio de soluciones sin perder por completo los avances logrados. El algoritmo conserva los operadores de selección, cruce ponderado y mutación, así como el elitismo.

5.6.2. Componentes del Algoritmo

Ahora, se describen los principales componentes añadidos o modificados para añadirle el reinicio poblacional al algoritmo genético con cruce ponderado.

Criterio de reinicio

El algoritmo monitoriza el rendimiento del segundo mejor individuo en cada generación. Si este no mejora en dos iteraciones consecutivas, se considera que la población ha entrado en un estado de estancamiento evolutivo y se procede a un reinicio.

Reinicio selectivo

Durante el reinicio, solo se conserva el mejor individuo de la población actual. El resto de la población se regenera completamente mediante una nueva inicialización aleatoria. Esto permite explorar nuevas regiones del espacio de soluciones sin perder los avances obtenidos.

Evaluación post-reinicio

Tras el reinicio, todos los nuevos individuos se evalúan y se reorganiza la población según su fitness. Se actualizan los valores del mejor y segundo mejor individuo, y se reanuda el ciclo evolutivo.

Persistencia de historial

Para no perder información relevante, se mantiene un historial acumulativo del fitness a lo largo de todos los reinicios. Esto permite realizar un análisis completo del proceso evolutivo una vez finalizada la ejecución.

5.6.3. Pseudocódigo

Para facilitar su lectura o replicación, se ha detallado el pseudocódigo del **Algorithm 6** Algoritmo Genético con Reinicio Poblacional.

Algorithm 6 Algoritmo Genético con Reinicio Poblacional

```
1: Inicializar población y evaluar
2: Guardar mejor y segundo mejor individuos
3: while no se alcance el máximo de evaluaciones do
4:   Aplicar elitismo
5:   while población incompleta do
6:     Seleccionar padres por torneo
7:     Aplicar cruce ponderado y mutación adaptativa
8:     Evaluar hijos y conservar el mejor
9:   end while
10:  if segundo mejor no mejora en 2 generaciones then
11:    Reiniciar población conservando solo el mejor
12:  end if
13:  Actualizar mejores si hay mejora
14: end while
15: return mejor solución
```

5.6.4. Aplicación en la Reducción de Datos

Esta estrategia mejora la robustez del algoritmo en entornos donde hay múltiples óptimos locales. Al reintroducir diversidad controlada, incrementa la probabilidad de encontrar subconjuntos más eficientes para el entrenamiento del modelo, especialmente en datasets complejos o con clases desbalanceadas.

5.6.5. Ventajas y Limitaciones

Ventajas:

- Permite escapar de óptimos locales sin descartar los avances previos.
- Favorece una exploración más amplia del espacio de soluciones.
- Mejora la estabilidad y robustez del algoritmo frente al estancamiento.

Limitaciones:

- Incrementa la complejidad de implementación y control del flujo.
- El reinicio puede introducir ruido si se activa de forma prematura.

5.7. Algoritmo Memético

El **algoritmo memético** es una extensión del enfoque genético tradicional que incorpora técnicas de búsqueda local dentro del proceso evolutivo. Su objetivo es combinar la exploración global del espacio de soluciones (propia de los algoritmos evolutivos) con la explotación local de buenas soluciones (propia de la optimización heurística). En este trabajo, se implementó una versión adaptada del algoritmo memético orientada a la selección de subconjuntos óptimos de imágenes para el entrenamiento de modelos convolucionales.

5.7.1. Descripción

El funcionamiento general del algoritmo memético se basa en una estructura genética clásica: generación de población inicial, selección por torneo, cruce, mutación y reemplazo elitista. Sin embargo, introduce una novedad clave: la aplicación probabilística de una búsqueda local sobre los individuos recién generados. Este procedimiento local intenta mejorar los hijos generados por cruce, evaluando soluciones vecinas mediante pequeñas modificaciones (mutaciones controladas).

5.7.2. Pseudocódigo

Para facilitar su lectura o replicación, se ha detallado el pseudocódigo del **Algorithm 7** Algoritmo Memético.

5.7.3. Aplicación en la reducción de datos

El **algoritmo memético** fue diseñado para encontrar subconjuntos de datos que optimicen métricas como la *accuracy* o el *F1-score* en un número reducido de evaluaciones. Su capacidad para ajustar localmente los subconjuntos permite afinar las selecciones iniciales generadas por los operadores evolutivos, lo que se traduce en soluciones de mayor calidad con menor dispersión.

5.7.4. Ventajas y limitaciones

Ventajas:

- Mejora puntual de soluciones mediante refinamiento local.
- Reduce la probabilidad de estancamiento en óptimos locales.

Algorithm 7 Algoritmo Memético.

La búsqueda local está limitada por el número total de evaluaciones permitidas y permite realizar mejoras incrementales únicamente cuando resultan beneficiosas.

```
1: Inicializar población y evaluar
2: Guardar mejor individuo
3: while evaluaciones < máximo do
4:   Aplicar elitismo
5:   while nueva población incompleta do
6:     Seleccionar padres por torneo
7:     Aplicar cruce ponderado
8:     Aplicar mutación
9:     if se cumple probabilidad de búsqueda local then
10:       Ejecutar búsqueda local sobre el hijo
11:       Evaluar vecinos y conservar el mejor
12:     else
13:       Evaluar hijo directamente
14:     end if
15:     Añadir hijo resultante a la nueva población
16:   end while
17:   Reemplazar población
18:   Actualizar mejor si mejora
19:   if hay estancamiento then
20:     Terminar
21:   end if
22: end while
23: return mejor solución
```

- Genera soluciones más consistentes y robustas frente a la aleatoriedad.

Limitaciones:

- Aumenta ligeramente el coste computacional por las evaluaciones adicionales de la búsqueda local.
- Requiere calibración adicional de parámetros como la probabilidad de búsqueda local o el tamaño del vecindario.

5.8. Versiones Libres de los Algoritmos Evolutivos

En esta sección se describen las versiones **libres** de algunos algoritmos evolutivos, caracterizadas por permitir que el número de imágenes seleccionadas evolucione dinámicamente a lo largo del proceso. A diferencia de las versiones fijas, donde se impone un porcentaje constante de selección, estas variantes adaptan el tamaño de los subconjuntos seleccionados mediante operadores ajustables y estocásticos. Esta propiedad resulta útil cuando no se conoce a priori la proporción óptima de instancias para entrenar un modelo eficiente.

5.8.1. Algoritmo Genético con Cruce Ponderado (Versión Libre)

La versión libre del Algoritmo Genético con Cruce Ponderado incorpora un operador de cruce flexible que ajusta automáticamente el tamaño de los subconjuntos generados. Al activar el parámetro `adjust_size`, el número de imágenes seleccionadas en cada hijo se calcula aleatoriamente dentro de un rango proporcional al tamaño de los padres, en este caso, entre el 50 % y el 150 % del tamaño base.

Esto permite al algoritmo explorar configuraciones más amplias o más compactas en busca del subconjunto óptimo. Además, el operador de mutación también se adapta para permitir fluctuaciones en la cantidad total de imágenes activas, aumentando así la diversidad estructural de la población.

5.8.2. Algoritmo Genético con Reinicio Poblacional (Versión Libre)

La versión libre del Algoritmo Genético con Reinicio Poblacional conserva la lógica de reinicio para evitar estancamientos, pero permite que cada nuevo individuo generado tenga una cantidad de imágenes seleccionadas distinta.

Este comportamiento se gestiona mediante el mismo mecanismo de cruce adaptativo que en la versión libre del genético v2.

Gracias a esta propiedad, cada reinicio poblacional no solo introduce nueva diversidad en el contenido de los subconjuntos, sino también en su escala. Esto es especialmente útil para escapar de regiones del espacio de soluciones con un tamaño fijo subóptimo.

5.8.3. Algoritmo Memético (Versión Libre)

La versión libre del Algoritmo Memético combina tanto la flexibilidad del cruce ponderado como la búsqueda local ajustable. Al igual que en los algoritmos anteriores, el tamaño de los subconjuntos puede variar en cada generación mediante un parámetro `adjust_size` activado tanto en los operadores evolutivos como en la heurística local.

Durante la búsqueda local, esta versión permite que un vecino no solo cambie el contenido de las imágenes seleccionadas, sino también su número total. Esta dualidad entre exploración global y explotación local con tamaño dinámico convierte al algoritmo en una herramienta poderosa para adaptarse a distintas complejidades del conjunto de datos, logrando soluciones robustas sin una configuración de selección fija.

Capítulo 6

Implementación

En este capítulo se presenta en detalle la arquitectura técnica del sistema implementado, incluyendo los componentes y módulos principales, las herramientas específicas empleadas en la construcción del sistema, y los elementos clave para optimizar el rendimiento de los algoritmos y su evaluación.

6.1. Descripción del Sistema

La estructura del proyecto se organizó modularmente para facilitar el acceso, el mantenimiento y la extensibilidad del código fuente. La organización de carpetas es la siguiente:

- **data** – Conjunto de datos utilizados en los experimentos.
- **docs** – Documentación del proyecto en latex.
 - **bibliografia** – Archivos relacionados con las referencias bibliográficas.
 - **capitulos** – Archivos individuales para cada capítulo del documento.
 - **config** – Archivos de configuraciones de la documentación LaTeX.
 - **imagenes** – Imágenes utilizadas en la documentación.
 - **out** – Archivos generados por el compilador de LaTeX.
 - **portada** – Archivo portada del documento.
 - **prefacio** – Archivo prefacio del documento.
 - **proyecto.tex** – Archivo principal de LaTeX que compila el documento.

- `img` – Imágenes generadas automáticamente durante los experimentos.
- `LICENSE` – Términos de distribución del proyecto.
- `logs` – Registros de las ejecuciones, incluyendo tiempos de inicio, fin y resultados intermedios de los algoritmos.
- `README.md` – Descripción general.
- `requirements.txt` – Dependencias del proyecto.
- `results` – Resultados de los experimentos.
 - `csvs` – Resultados de las ejecuciones guardados en tablas.
 - `salidas` – Salidas en bruto de consola.
- `scripts` – Scripts de ejecución automática, comparación de experimentos y generación de gráficos finales.
- `src` – Código fuente principal del proyecto.
 - `algorithms` – Implementaciones de los algoritmos.
 - `main.py` – Módulo principal de ejecución individual.
- `tmp` – Ficheros temporales generados durante la ejecución.
- `utils` – Módulos de apoyo, como clases auxiliares, generación de gráficos y funciones utilitarias.

El código completo del proyecto, incluyendo los scripts, algoritmos, documentación y resultados, se encuentra disponible en el repositorio público de GitHub: <https://github.com/JoseRuizLopez/TFG>.

6.2. Herramientas y Lenguajes de Programación

El desarrollo del proyecto se ha llevado a cabo utilizando **Python 3.10** [24] como lenguaje principal, debido a su versatilidad y amplia adopción en el campo del **aprendizaje profundo** y la **manipulación de datos**. Python es conocido por su facilidad de uso, extensibilidad y la gran cantidad de librerías disponibles para el procesamiento de datos y la implementación de modelos de **machine learning**.

Las principales librerías empleadas durante el desarrollo son las siguientes:

- **PyTorch 2.3.1** [25, 26]: Para la construcción, entrenamiento y optimización de modelos de aprendizaje profundo. PyTorch fue elegido por su flexibilidad y capacidad para ejecutarse eficientemente en GPU.
- **Scikit-learn 1.5.2** [24]: Para la evaluación de los modelos se utilizaron métricas estándar [27]. Su API permite una integración fluida con PyTorch y otros módulos.
- **Numpy 2.0.0** [28]: Para operaciones matemáticas y manipulación de matrices, siendo una herramienta esencial en el procesamiento de datos.
- **Polars 1.9.0** [29]: Librería para manejar DataFrames de gran tamaño, elegida por su rendimiento superior en comparación con Pandas.
- **Pandas 2.2.3** [30]: Librería utilizada como apoyo para la generación de los gráficos, debido a las incompatibilidades detectadas con **polars**.
- **Matplotlib 3.9.2** [31]: Librería utilizada para la generación y visualización de gráficos.
- **Seaborn 0.13.2** [32]: Librería para la generación avanzada de gráficos estadísticos.
- **Openpyxl 3.1.5** [33]: Librería para la generación automática de archivos Excel a partir de resultados experimentales.

Cada una de estas herramientas fue seleccionada por su robustez y su idoneidad para cumplir con los requisitos específicos del proyecto, facilitando tanto la implementación de los algoritmos meméticos como la reducción y el análisis de los datos utilizados en los modelos de aprendizaje profundo.

6.3. Gestión de Dependencias

Para garantizar que el proyecto se ejecute correctamente y todas las librerías necesarias estén disponibles, se ha utilizado un archivo **requirements.txt**. Este archivo contiene una lista de todas las librerías y sus versiones específicas que el proyecto requiere.

Para el **desarrollo local**, se ha optado por crear un entorno virtual utilizando **venv** [34]. Esta práctica permite aislar las dependencias del proyecto de otros proyectos en la máquina, evitando conflictos entre versiones de librerías.

Para la **implementación en el servidor**, se ha utilizado **conda** [35] como gestor de paquetes y entornos. Conda facilita la gestión de entornos y la instalación de librerías, especialmente en configuraciones más complejas.

Esto facilita la reproducibilidad del proyecto y minimiza posibles conflictos de versión, lo que es fundamental para mantener la integridad del código y el rendimiento de las aplicaciones.

6.4. Arquitectura de la Implementación

La arquitectura de la implementación desarrollada para este trabajo está diseñada con un enfoque modular y escalable, que permite gestionar de forma eficiente las distintas fases del proceso de selección de instancias y evaluación de modelos. El sistema se compone de varios módulos interrelacionados que se encargan de la generación de subconjuntos, la ejecución de los algoritmos evolutivos y meméticos, la evaluación de las soluciones mediante modelos preentrenados, y la visualización y almacenamiento de resultados. Esta estructura facilita la incorporación de nuevos algoritmos o mejoras en los ya existentes, garantizando una mayor flexibilidad y mantenimiento del código.

6.4.1. Esquema General de Funcionamiento de los Algoritmos

El funcionamiento general de los algoritmos implementados en este trabajo sigue un flujo estructurado y modular que permite una ejecución ordenada de todas las etapas del proceso. La Figura 6.1 muestra el esquema general de funcionamiento, desde la configuración inicial hasta la evaluación final de resultados.

El proceso se organiza en los siguientes pasos principales:

1. **Ejecución del script de paralelización:** Se inicia el script que permite ejecutar múltiples experimentos de manera paralela, facilitando la obtención de resultados de diferentes semillas.
2. **Configuración inicial:** Se define el algoritmo, el modelo y el dataset a utilizar, a través de los parámetros en `generator.py`.
3. **Ejecución del núcleo principal:** El script `main.py` se encarga de orquestrar la ejecución, llamando a las funciones específicas según el algoritmo seleccionado.
4. **Generación y evaluación de subconjuntos:** Cada algoritmo genera subconjuntos de datos de entrenamiento a partir del conjunto original.

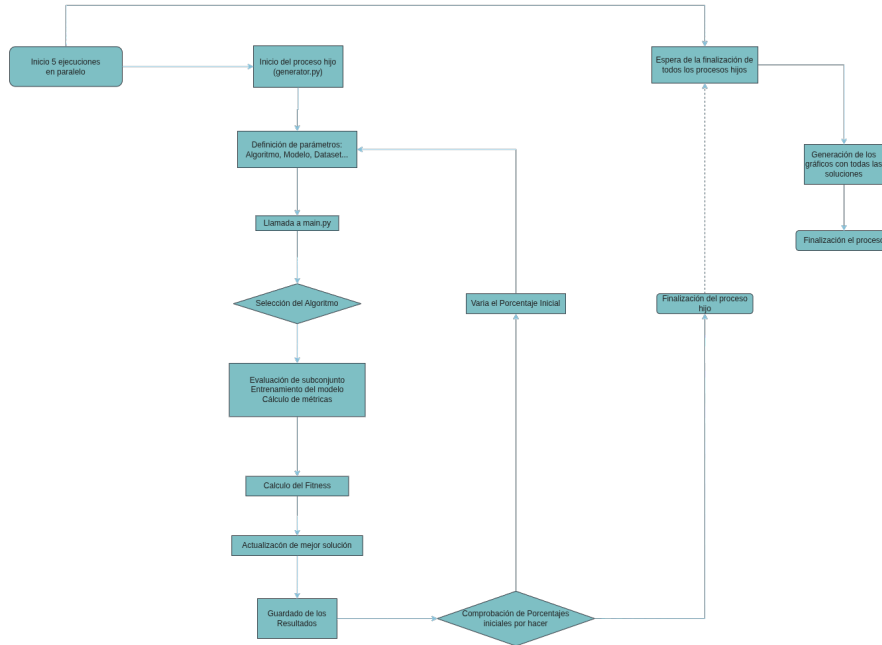


Figura 6.1: Esquema general del flujo de ejecución de los algoritmos y evaluación de subconjuntos.

5. **Entrenamiento del modelo:** Se realiza el entrenamiento del modelo utilizando los subconjuntos generados, aplicando técnicas de *transfer learning* con modelos preentrenados.
6. **Almacenamiento de resultados:** Cada evaluación almacena los resultados obtenidos (métricas y composición del subconjunto) y, tras finalizar la ejecución, se generan gráficos como boxplots y visualizaciones que permiten analizar el rendimiento de cada enfoque.

6.4.2. Configuración de Modelos Preentrenados

En este trabajo, se han utilizado modelos convolucionales preentrenados como base para la tarea de clasificación de imágenes. Concretamente, se han empleado las arquitecturas **ResNet50** y **MobileNetV2**, cargadas con pesos preentrenados sobre **ImageNet** o **CIFAR10**, según el dataset utilizado en cada experimento.

Para adaptar estos modelos a la tarea específica de clasificación de subconjuntos de datos, se ha seguido una estrategia de *transfer learning* que permite reducir significativamente el coste computacional del entrenamiento. Esta estrategia consiste en congelar todas las capas del modelo, excepto la

última, de manera que las representaciones aprendidas en la etapa preentrenada puedan ser reutilizadas como extractores de características generales.

Específicamente, la **última capa** de cada modelo preentrenado (la capa densa o *fully connected*) ha sido reemplazada por una nueva capa completamente conectada (**Linear**) adaptada al número de clases del dataset en uso. Esta capa es la única que se entrena durante la fase de evaluación de cada subconjunto generado por los algoritmos, permitiendo obtener métricas como *accuracy*, *precision*, *recall* y *F1-score* de forma eficiente.

El uso de pesos preentrenados y la congelación de capas proporcionan varias ventajas:

- Reducción del tiempo de entrenamiento en cada evaluación.
- Aprovechamiento de representaciones genéricas previamente aprendidas, lo que mejora la generalización.
- Adaptabilidad de los modelos a distintos datasets y tareas, modificando únicamente la capa final.

Este enfoque se implementa mediante las herramientas de PyTorch, congelando explícitamente los gradientes de todas las capas del modelo y definiendo la nueva capa final con el número adecuado de salidas para cada problema de clasificación.

6.4.3. Módulo de Algoritmos

Ubicado en `src/algorithms/` este módulo contiene las implementaciones principales de los algoritmos desarrollados en el proyecto.

Este módulo utiliza la arquitectura GPU para maximizar la velocidad de ejecución y está diseñado para ser escalable, permitiendo la inclusión de nuevos operadores meméticos si es necesario.

6.4.4. Núcleo de Ejecución

El módulo `main.py` centraliza la ejecución de un experimento individual, inicializando configuraciones, entrenando el modelo y generando gráficos.

Los pasos de la función principal de `main.py` es:

1. **Establece Configuración Inicial:** Configura una semilla, elige el dataset y prepara un archivo de log.

2. **Inicia el Proceso del Algoritmo:** Según el nombre del algoritmo (algoritmo) especificado, se llama a la función correspondiente (por ejemplo, `genetic_algorithm`, `memetic_algorithm`, etc.).
3. **Almacena Resultados:** Una vez que el algoritmo termina, registra la duración, los resultados y la métrica final en un archivo.
4. **Visualiza Resultados:** Si hay datos de fitness, genera una gráfica de la evolución del fitness a lo largo del proceso.
5. **Genera un Resumen:** Calcula estadísticas adicionales (como porcentaje de clases seleccionadas en Paper, Rock y Scissors), y devuelve estos resultados junto con el historial de fitness.

6.4.5. Módulo de Utilidades

La carpeta `utils/` contiene funciones auxiliares:

- **`utils_plot.py`:** Generación de gráficos.
- **`classes.py`:** Definición de enumeraciones para algoritmos, métricas, datasets y modelos.
- **`utils.py`:** Funciones de ayuda como el cálculo de métricas o la creación de diccionarios de selección de imágenes.

Estos módulos se encargan de generar gráficas comparativas entre distintos porcentajes o algoritmos y en generar un CSV con los datos finales para ser analizados.

6.4.6. Scripts de Ejecución en GPU

En scripts, se encuentran los programas necesarios para ejecutar los algoritmos en un servidor GPU, lo que permite maximizar la eficiencia en el entrenamiento y la evaluación de modelos.

1. **Configuración de GPU:** Los scripts están configurados para identificar y utilizar las GPU disponibles en el servidor, reduciendo los tiempos de entrenamiento de modelos.
2. **Optimización de Ejecución:** Se implementaron configuraciones de batch size y técnicas de procesamiento paralelo en PyTorch, aprovechando la memoria y el poder de procesamiento de las GPU.

Estos scripts están diseñados para ser ejecutados en un entorno de servidor, reduciendo los tiempos de prueba en el entorno local y permitiendo un análisis iterativo más rápido.

6.5. Consideraciones de Optimización

Durante el desarrollo, se optimizaron varios aspectos para mejorar el rendimiento del sistema:

1. **Aceleración en GPU:** Todas las operaciones de cálculo intensivo fueron migradas a la GPU mediante PyTorch.
2. **Uso Eficiente de Memoria:** Con Polars y Numpy, se optimizó el manejo de grandes volúmenes de datos, utilizando tipos de datos específicos para reducir el uso de memoria.
3. **Automatización de Evaluaciones:** Las pruebas de rendimiento se automatizaron, permitiendo una evaluación continua sin intervención manual.
4. **Control de reproducibilidad:** Se fijaron semillas aleatorias en todas las librerías involucradas (random, numpy, torch, cuda) y se desactivaron los algoritmos no deterministas de cuDNN [36]. Esta medida garantiza que las ejecuciones del sistema produzcan resultados consistentes entre sesiones, algo esencial en entornos de evaluación comparativa.
5. **Diagnóstico automático de GPU:** Implementación de un script (cuda-diagnostic.py) que comprueba disponibilidad de CUDA y dispositivos antes de lanzar experimentos, garantizando un entorno correcto.

Además, se implementó un mecanismo de **early stopping** basado en la ausencia de mejora del valor de fitness durante un número determinado de evaluaciones consecutivas. Aunque no se utiliza una pérdida de validación explícita como en enfoques tradicionales, este enfoque funcionalmente cumple el mismo propósito: detener el algoritmo cuando se detecta estancamiento, reduciendo así el coste computacional innecesario [37].

Gracias a estas optimizaciones, el sistema permite explorar un amplio abanico de configuraciones de manera eficiente, manteniendo la robustez y estabilidad de los resultados.

6.6. Métricas de Evaluación

Para evaluar el rendimiento de los modelos de clasificación utilizados en los experimentos, se han empleado cuatro métricas estándar en el ámbito del aprendizaje automático: **accuracy**, **precisión**, **recall** y **F1-score**. Dado que se trata de un problema multiclase, estas métricas se han calculado utilizando el promedio *macro*, lo que implica computarlas individualmente para cada clase y luego obtener la media aritmética.

A continuación se describen y formalizan cada una de estas métricas:

- **Accuracy** (exactitud): representa la proporción de predicciones correctas sobre el total de ejemplos.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

En el caso multiclase, se generaliza como:

$$Accuracy = \frac{\text{Número de predicciones correctas}}{\text{Total de predicciones}}$$

- **Precisión**: mide cuántas de las instancias clasificadas como positivas fueron realmente positivas. En el caso multiclase (macro), se calcula como:

$$Precisión_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i}$$

- **Recall**: indica cuántas de las instancias realmente positivas fueron correctamente identificadas por el modelo:

$$Recall_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FN_i}$$

- **F1-score**: es la media armónica entre precisión y recall, útil cuando se desea un equilibrio entre ambas:

$$F1\text{-score}_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \frac{2 \cdot Precisión_i \cdot Recall_i}{Precisión_i + Recall_i}$$

Donde:

- TP_i : verdaderos positivos de la clase i
- FP_i : falsos positivos de la clase i

- FN_i : falsos negativos de la clase i
- C : número total de clases

Para facilitar la comprensión de estos conceptos, en la Figura 6.2 se muestra una representación gráfica de una matriz de confusión.

A confusion matrix diagram. The horizontal axis is labeled 'True Class' with 'Positive' and 'Negative' categories. The vertical axis is labeled 'Predicated Class' with 'Positive' and 'Negative' categories. The matrix is a 2x2 grid of colored squares: top-left is green and labeled 'TP', top-right is light blue and labeled 'FP', bottom-left is light blue and labeled 'FN', and bottom-right is green and labeled 'TN'.

True Class		Positive	Negative
Predicated Class	Positive	TP	FP
	Negative	FN	TN

Figura 6.2: Representación gráfica de una matriz de confusión

Estas métricas fueron implementadas utilizando la librería `scikit-learn`, que permite su cálculo eficiente a partir de las predicciones del modelo y las etiquetas reales del conjunto de validación.

Capítulo 7

Entorno Experimental

En este capítulo se describe el entorno experimental empleado para evaluar los algoritmos propuestos. Se detallan los conjuntos de datos utilizados, el diseño de los experimentos, el procedimiento de ejecución y las métricas de evaluación. Asimismo, se explica cómo se presentan y visualizan los resultados obtenidos.

7.1. Datasets utilizados

En el aprendizaje profundo, los datasets son colecciones de datos etiquetados o no etiquetados que se utilizan para entrenar modelos. Estos conjuntos de datos contienen ejemplos organizados que representan la entrada para el modelo y, en muchos casos, también las etiquetas correspondientes que indican la salida deseada. Los datasets varían en tamaño, calidad y tipo, dependiendo de la tarea a resolver, como la clasificación de imágenes, el reconocimiento de patrones o la predicción de series temporales.

A continuación, se van a explicar cada uno de los Datasets que se han utilizado en el desarrollo del proyecto.

7.1.1. Rock, Paper, Scissors (Piedra, Papel, Tijera)

Rock, Paper, Scissors [38] es un conjunto de datos creado por Laurence Moroney que se utiliza para la clasificación de imágenes de manos representando los gestos de ‘piedra’, ‘papel’ y ‘tijeras’.

En la figura 7.1 se han mostrado una imagen de cada una de las clases del dataset Rock, Paper, Scissors, para que se pueda observar la similitud entre las distintas clases.

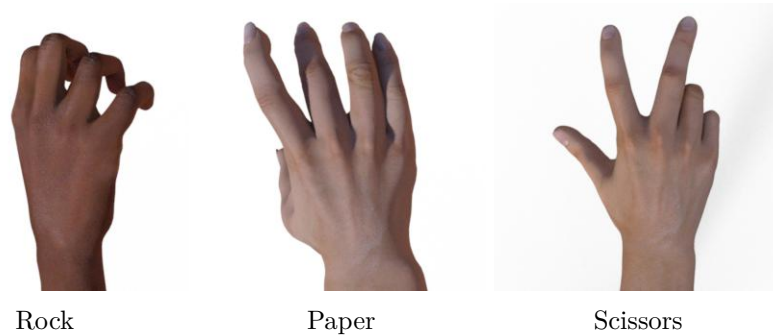


Figura 7.1: Ejemplos de imágenes del dataset Rock, Paper, Scissors

Estructura del Dataset

El conjunto de datos usado contiene 2,925 imágenes, distribuidas en tres categorías: piedra, papel y tijeras. Las imágenes están en color y tienen un tamaño de 300x300 píxeles.

Como se puede observar en la Figura 7.2, las imágenes están organizadas en directorios según su función en el entrenamiento (entrenamiento, validación o prueba) y, dentro de cada partición, se dividen a su vez por clases del dataset.

Formato de los Datos

Las imágenes están en formato JPEG (.jpg). Para su procesamiento, se han aplicado técnicas de preprocesamiento adaptadas a los requerimientos del modelo.

Uso del Dataset

Este dataset se ha utilizado para evaluar el rendimiento del modelo en un problema de clasificación de imágenes con múltiples clases, pero siendo un dataset sencillo y con un número de clases pequeño. Además, permite explorar la eficacia de los algoritmos meméticos en un entorno más cercano al reconocimiento de objetos.

Correcciones en la División de Datos

Según la nota observada en el README del dataset:

Note: in the source, Laurence calls “validation” as the “test”, and “test” the “validation”.

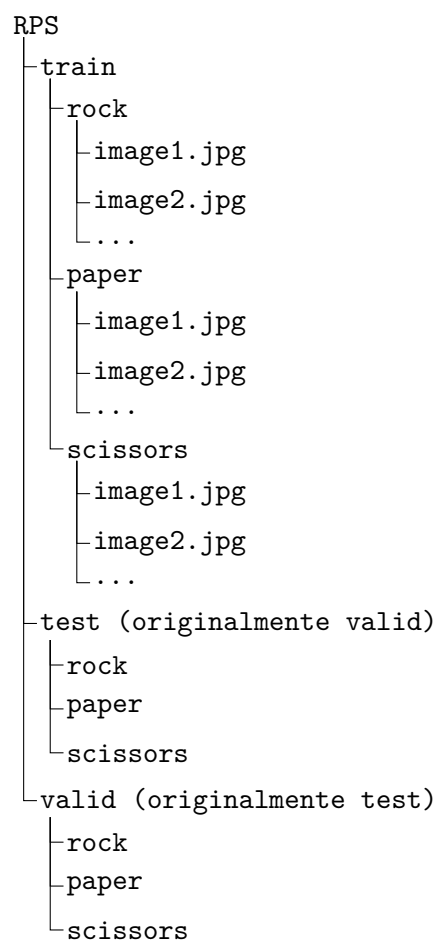


Figura 7.2: Estructura de carpetas del dataset Rock, Paper, Scissors

se han renombrado las particiones de `test` y `valid` para que correspondan correctamente con sus propósitos.

Licencia y uso

Este conjunto de datos se distribuye bajo la licencia **Creative Commons Attribution 4.0 International (CC BY 4.0)**, lo que permite su uso, modificación y distribución con la condición de otorgar el crédito adecuado a los creadores originales.

7.1.2. PAINTING (Art Images: Drawing/Painting/Sculptures/Engravings)

El dataset **Art Images: Drawing/Painting/Sculptures/Engravings** usado es una colección de 8.576 imágenes organizadas en cinco categorías de arte: dibujos, pinturas, esculturas, grabados y arte iconográfico.

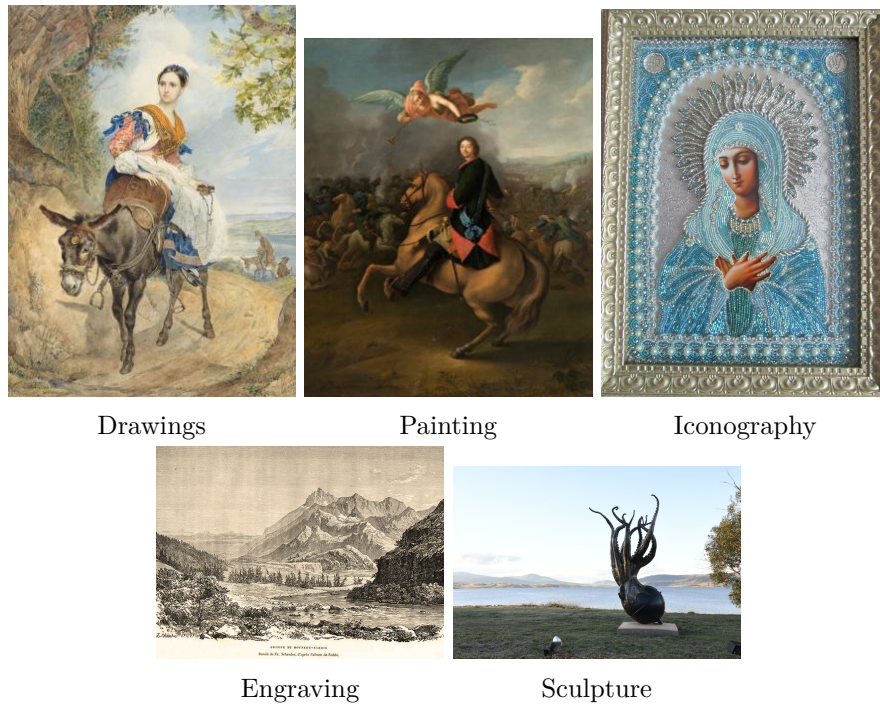


Figura 7.3: Ejemplos de clases en el dataset PAINTING

En la figura 7.3 se han mostrado una imagen de cada una de las clases del dataset PAINTING, para que se pueda observar la variabilidad de las imágenes.

Estructura del Dataset

Como se puede observar en la Figura 7.4, las imágenes están organizadas en directorios según su categoría artística, que en este caso corresponden a las distintas clases del dataset, previamente divididas en conjuntos de entrenamiento y prueba.

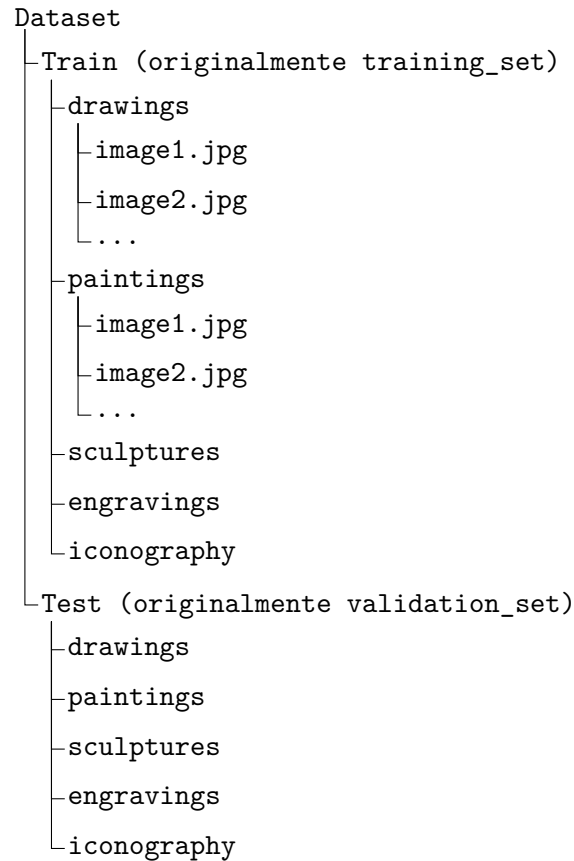


Figura 7.4: Estructura de carpetas del dataset PAINTING

Formato de los Datos

Todas las imágenes están en formato JPEG (.jpg) y presentan variaciones en resolución y dimensiones. Se han aplicado técnicas de preprocesamiento para homogenizar las características de las imágenes.

Uso del Dataset

Este dataset se ha utilizado para entrenar y evaluar modelos de clasificación de imágenes en un entorno diferente al RPS. Con este dataset, se ha comprobado el funcionamiento para evaluar los algoritmos con un dataset un poco mas complejo que el RPS, con un par de clases más y con un número mayor de imágenes.

Correcciones en la División de Datos

Observando los tamaños de la división de los datos, y teniendo en cuenta que la división de los datos suele ser en train y test, se ha decidido por renombrar las particiones de **valid** por **test** para que corresponda correctamente con su propósito. Y el set de validation lo he obtenido separando el set de train, normalmente haciendo una división 80 % test y 20 % valid.

Acceso al Dataset

Inicialmente, el dataset se descargó desde Kaggle [39]

Sin embargo, debido a la presencia de archivos innecesarios y algunas imágenes corruptas, se opta por una versión limpia disponible en Kaggle [40].

Licencia y Uso

Antes de su uso, se revisaron los términos y condiciones establecidos en la página de Kaggle para asegurar el cumplimiento con las licencias y restricciones aplicables.

7.1.3. Comparación entre datasets

Gracias a la tabla 7.1 que con las características más relevantes de los datasets utilizados, permite entender mejor la complejidad relativa de cada conjunto y cómo pueden influir en el comportamiento de los algoritmos:

Dataset	Nº Imágenes	Nº Clases	Formato	Tamaño Imagen
Rock, Paper, Scissors	2.925	3	JPG	300×300 px
PAINTING	8.576	5	JPG	Variable

Tabla 7.1: Resumen comparativo de los datasets utilizados

La selección de estos dos datasets responde a la necesidad de evaluar los algoritmos meméticos en distintos niveles de complejidad. El dataset **Rock, Paper, Scissors** se utilizó en las primeras fases del proyecto como punto de partida, ya que ofrecía un entorno sencillo y controlado, con un número reducido de clases y una estructura equilibrada. Esto permite desarrollar las bases del sistema y probar las primeras versiones de los algoritmos de manera más ágil y con menor complejidad computacional.

Por su parte, el dataset **PAINTING** se empleó posteriormente para validar el comportamiento de los algoritmos en un entorno más exigente. Al incluir cinco categorías de arte con distintos estilos visuales, este conjunto

introdujo una mayor variabilidad tanto semántica como estructural, lo que permite evaluar la robustez y capacidad de generalización de las soluciones desarrolladas.

7.2. Diseño de los experimentos

La fase experimental se organiza en varias etapas. Inicialmente se opta por un dataset simple (*Rock, Paper, Scissors*) para validar el funcionamiento general del sistema. Posteriormente, se realizan pruebas con datasets más exigentes. Los experimentos se repiten utilizando diferentes porcentajes iniciales de datos (10 %, 25 %, 50 % y 75 %) para estudiar cómo afecta la cantidad de datos seleccionados al rendimiento del modelo.

Con el fin de asegurar la consistencia entre ejecuciones experimentales, se aplican las medidas de control de reproducibilidad descritas en el Apartado 6.5. Esto permite comparar algoritmos en condiciones homogéneas, evitando variaciones indeseadas causadas por componentes aleatorios del entorno de ejecución.

En cada prueba, se realizan 5 ejecuciones en paralelo, cada una utilizando una semilla distinta. Esta estrategia permite obtener resultados promedio más robustos frente a la aleatoriedad del proceso evolutivo, asegurando una mayor fiabilidad estadística.

Los apartados siguientes explican con mayor detalle el procedimiento adoptado para llevar a cabo dichas ejecuciones.

7.3. Procedimiento de Ejecución y Evaluación

Para garantizar la consistencia y objetividad en la comparación entre algoritmos, se diseñan un procedimiento experimental sistemático y replicable. Cada ejecución se realizó bajo las mismas condiciones computacionales y utilizando los mismos parámetros base, salvo en aquellos casos en que se deseaba estudiar una variación concreta, como los distintos porcentajes iniciales o el uso de metaheurísticas con porcentajes libres.

7.3.1. Métricas de Evaluación

Para evaluar el rendimiento de los modelos se utilizaron métricas estándar como **accuracy**, **precisión**, **recall** y **F1-score**, calculadas sobre el conjunto de validación tras cada evaluación. Para una definición formal de estas métricas, véase el Apartado 6.6.

Estas métricas se calcularon utilizando las funciones de `scikit-learn`, a partir de las predicciones del modelo y las etiquetas reales correspondientes a los subconjuntos de imágenes seleccionados por cada algoritmo.

7.3.2. Evaluaciones por Ejecución

Se buscó un equilibrio entre la cantidad de evaluaciones y el tiempo de ejecución, permitiendo una exploración suficiente del espacio de soluciones sin comprometer la eficiencia computacional. Por ello, cada algoritmo se configura para realizar un máximo de 100 evaluaciones por ejecución, independientemente del tipo de algoritmo utilizado, con el fin de mantener la equidad comparativa.

Cada evaluación consiste en generar un subconjunto de datos, entrenar el modelo correspondiente (ResNet50 o MobileNetV2), y calcular su *fitness* de acuerdo con las métricas mencionadas.

El número de evaluaciones sin mejora también se monitoriza para aplicar criterios de parada anticipada, explicados previamente en el Apartado 6.5, reduciendo así el tiempo computacional en caso de estancamiento.

7.3.3. Repeticiones y Semillas

Con el objetivo de obtener resultados estadísticamente significativos y reducir el efecto de la aleatoriedad, cada configuración experimental se ejecuta 5 veces, utilizando 5 semillas distintas. Los resultados presentados en las tablas y gráficos corresponden a la media de esas ejecuciones, junto con medidas de dispersión cuando procede, como los boxplots.

Cabe destacar que, en el caso de los boxplots, en lugar de representar la media de las 5 ejecuciones por configuración, se opta por incluir todos los valores individuales obtenidos con las distintas semillas. Esta decisión permite visualizar una distribución más realista del comportamiento de cada algoritmo, resaltando mejor la mediana, así como los valores máximos y mínimos alcanzados durante las ejecuciones.

7.3.4. Tiempos de Ejecución

Cada evaluación implica entrenar un modelo desde cero, por lo que los tiempos de ejecución son considerables. Por ejemplo, una ejecución completa con 100 evaluaciones puede tardar entre 30 minutos y 2 horas, dependiendo del algoritmo y del modelo utilizado.

Los algoritmos más complejos, como el memético o las versiones con

reinicio poblacional, requieren un mayor tiempo de ejecución debido a las operaciones adicionales de mejora local o regeneración de población.

7.4. Visualización de resultados

Para facilitar la comparación entre algoritmos, en el resto del trabajo se incluyen representaciones gráficas de los resultados obtenidos mediante **boxplots** y **barplots**. Ambos tipos de gráficos permiten visualizar el comportamiento global de cada algoritmo a partir de múltiples ejecuciones con distintas semillas.

Los boxplots (diagramas de caja) muestran la distribución estadística de los valores obtenidos. En la Figura 7.5 se presenta un ejemplo anotado que ilustra las distintas partes de este tipo de gráfico. La línea central de la caja representa la **mediana**, mientras que los bordes inferior y superior corresponden al **primer cuartil** (Q1) y **tercer cuartil** (Q3), respectivamente. La diferencia entre ellos define el **rango intercuartílico** (*IQR*), que contiene el 50 % central de los valores. Las líneas que se extienden desde la caja (conocidas como *bigotes*) alcanzan típicamente hasta 1.5 veces el *IQR*. Los puntos que quedan fuera de ese rango se consideran **valores atípicos**, lo que permite detectar ejecuciones excepcionales. Esta representación es especialmente útil para comparar la tendencia central, la dispersión y la estabilidad de los resultados obtenidos por los distintos algoritmos.

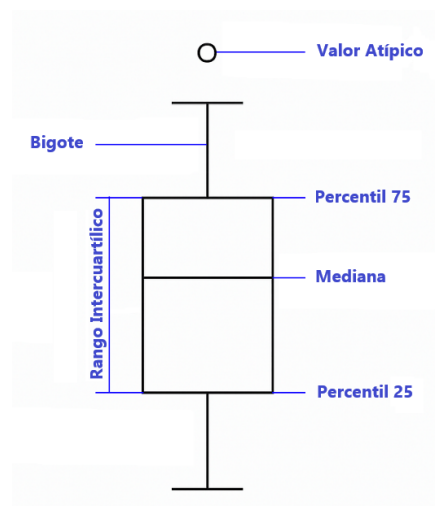


Figura 7.5: Ejemplo de boxplot explicado.

Los barplots (gráficos de barras), por su parte, se emplean para representar valores agregados como medias o proporciones, y son útiles para

observar cómo varía una métrica concreta entre distintos algoritmos, modelos o configuraciones. Además, los barplots muestran unas barras de error que indica la desviación estándar de los valores, lo que permite apreciar la variabilidad de los resultados obtenidos en las distintas ejecuciones.

Los scatter plots (diagramas de dispersión) permiten visualizar de forma conjunta la relación entre el porcentaje inicial de datos seleccionados y el porcentaje final alcanzado tras aplicar los algoritmos. En estos gráficos, cada punto representa una ejecución concreta, donde:

- El eje **X** indica el **Porcentaje Inicial** de datos usados.
- El eje **Y** muestra el **Porcentaje Final** de datos seleccionados por el algoritmo tras la ejecución.
- El **color** y el **tamaño** de los puntos reflejan el **accuracy** alcanzado: los puntos más grandes y con tonos más intensos corresponden a ejecuciones con mayor precisión.

Además, cada punto se conecta mediante líneas de guía horizontales y verticales a los ejes, lo que facilita la interpretación de su posición exacta. Estas líneas ayudan a identificar rápidamente los valores individuales en ambos ejes, mejorando la legibilidad del gráfico.

Este tipo de visualización es especialmente útil para analizar cómo los algoritmos ajustan el tamaño del subconjunto de datos en función de las condiciones iniciales, y cómo esta selección influye en la precisión del modelo. Permite detectar patrones como agrupamientos de soluciones o dispersiones, y observar cómo ciertas configuraciones iniciales tienden a producir mejores o peores resultados.

7.5. Nomenclatura simplificada de los algoritmos

Para facilitar la lectura de las tablas y gráficos presentados a lo largo de este trabajo, se opta por utilizar abreviaciones consistentes para referirse a los algoritmos implementados. Estas abreviaciones permiten condensar la información visualmente sin perder la claridad en la interpretación de los resultados.

La correspondencia entre el nombre completo de cada algoritmo y su abreviatura se muestra en la Tabla 7.2.

A partir de este punto, todas las tablas y gráficos del documento utilizarán estas abreviaturas para referirse a los algoritmos, mejorando la legibilidad de los resultados y facilitando la interpretación de las comparativas entre métodos.

Nombre Completo del Algoritmo	Abreviatura
Aleatorio (Random Search)	RS
Búsqueda Local (Local Search)	LS
Búsqueda Local Libre	LS-F
Algoritmo Genético	GA
Genético con Cruce Ponderado (Weighted Crossover)	GA-WC
Genético con Mutación Adaptativa (Adaptive Mutation)	AM
Genético con Mutación Adaptativa Libre	AM-F
Genético con Reinicio Poblacional (Population Restart)	PR
Algoritmo Memético	MA
Algoritmo Memético Libre	MA-F

Tabla 7.2: Nomenclatura simplificada de los algoritmos.

7.6. Estructura de las Tablas de Resultados

En los siguientes capítulos, algunos resultados obtenidos a partir de los experimentos se presentan de forma tabulada para facilitar la comprensión y el análisis comparativo entre algoritmos, modelos y configuraciones. Estas tablas recogen las métricas de rendimiento clave obtenidas a lo largo de las distintas ejecuciones.

Las columnas principales que pueden aparecer en las tablas son las siguientes:

- **Porcentaje Inicial:** Indica el porcentaje inicial de datos seleccionados por el algoritmo para cada ejecución. Este valor permite analizar cómo varía el rendimiento del modelo en función de la cantidad de datos utilizados.
- **Evaluaciones Realizadas:** Número total de evaluaciones efectuadas por el algoritmo en cada configuración. Generalmente, este valor se fija en 100, salvo en casos concretos como la evaluación con el 100 % del conjunto de datos.
- **Duración Total:** Tiempo total requerido para completar todas las evaluaciones de una configuración, expresado en formato *horas:minutos:segundos*. Este campo ayuda a comparar la eficiencia computacional de los algoritmos. Se obtiene a partir de las distintas ejecuciones del algoritmo con diferentes semillas.
- **Duración por Evaluación:** Tiempo necesario para realizar un único entrenamiento, permitiendo evaluar la rapidez de cada modelo de

forma precisa. Este valor se calcula dividiendo la duración total entre el número de evaluaciones realizadas.

- **Accuracy (Avg)**: Precisión media alcanzada, expresada en porcentaje, calculada sobre el conjunto de validación. Se obtiene a partir de las distintas ejecuciones del algoritmo con diferentes semillas.
- **Precision (Avg)**: Media de la precisión por clase (macro promedio), que refleja la proporción de verdaderos positivos entre todas las predicciones positivas para cada clase.
- **Recall (Avg)**: Media del *recall* por clase, que indica la proporción de verdaderos positivos entre todas las instancias reales de cada clase.
- **F1-score (Avg)**: Media armónica entre precisión y *recall*, calculada por clase y luego promediada (macro promedio). Es una métrica clave para problemas multiclase.

En algunas tablas, los resultados se agrupan por **modelo de red neuronal** o por **porcentaje inicial**, permitiendo comparar directamente su rendimiento bajo las mismas condiciones experimentales y facilitando el análisis del impacto de la cantidad de datos seleccionados.

Por último, aclarar que las abreviaturas de los algoritmos empleadas en las tablas corresponden a las definidas en la Tabla 7.5, para mejorar la legibilidad de los resultados y simplificar las comparaciones.

Las tablas están diseñadas para ofrecer una visión clara, precisa y estructurada del impacto de cada configuración experimental en el rendimiento final de los modelos. Esta información es fundamental para evaluar la eficacia de las estrategias de reducción de datos implementadas en este trabajo.

7.7. Parámetros de los Algoritmos y del Entrenamiento

En este apartado se describen los principales parámetros utilizados en el desarrollo de los algoritmos de selección de instancias y en el proceso de entrenamiento de los modelos de clasificación. La correcta configuración de estos parámetros es fundamental para asegurar un equilibrio entre la calidad de los resultados y la eficiencia computacional.

7.7.1. Parámetros Generales de los Algoritmos

- **Porcentaje inicial de selección (`initial_percentage`)**: Define el porcentaje inicial de datos seleccionados para cada ejecución. Se han

evaluado valores como 10 %, 25 %, 50 % y 75 %.

- **Número máximo de evaluaciones** (`max_evaluations`): Límite de iteraciones para cada algoritmo, generalmente fijado en 100.
- **Número máximo de evaluaciones sin mejora** (`max_evaluations_without_improvement`): Controla la parada anticipada si no se mejora el mejor resultado durante un número consecutivo de iteraciones.
- **Métrica de optimización** (`metric`): Métrica utilizada para calcular el fitness, siendo `accuracy` la más habitual, aunque también se soportan `f1-score` y otras métricas.
- **Modelo de red neuronal** (`model_name`): Arquitectura utilizada para la clasificación, como `ResNet50` o `MobileNetV2`.

7.7.2. Parámetros del Entrenamiento de Modelos

- **Tamaño del batch** (`batch_size`): Número de imágenes procesadas simultáneamente durante el entrenamiento. Se fija en 32 para equilibrar tiempo y estabilidad.
- **Número de épocas** (`num_epochs`): Número de veces que el modelo se entrena sobre el subconjunto seleccionado. Se utiliza 10 para cada evaluación.
- **Tasa de aprendizaje** (`learning_rate`): Controla la magnitud de los ajustes de pesos durante el entrenamiento. Se establece en 0.001 para la capa final, manteniendo el resto de la red congelada.
- **Dispositivo de ejecución** (`device`): Determina si el entrenamiento se realiza en CPU o GPU. Se prioriza el uso de GPU cuando está disponible.

7.7.3. Parámetros Específicos por Algoritmo

Random Search:

- No utiliza parámetros adicionales, ya que la selección de instancias es completamente aleatoria.

Búsqueda Local:

- **Tamaño del vecindario** (`neighbor_size`): Número de imágenes modificadas en cada generación de vecino, fijado en 10.

- **Ajuste dinámico del tamaño del subconjunto** (`adjust_size`): Permite variar el tamaño total de las soluciones en las versiones libres.

Algoritmos Genéticos:

- **Tamaño de la población** (`population_size`): Número de individuos en cada generación, generalmente 10.
- **Tamaño del torneo** (`tournament_size`): Número de soluciones evaluadas para seleccionar padres en el cruce, fijado en 3.
- **Tasa de mutación** (`mutation_rate`): Probabilidad de modificar aleatoriamente las soluciones, fijado en 0.1.
- **Ajuste dinámico del tamaño del subconjunto** (`adjust_size`): Permite modificar el tamaño del subconjunto durante la ejecución.

Algoritmo Memético:

- **Probabilidad de búsqueda local** (`local_search_probability`): Probabilidad de aplicar una búsqueda local a un individuo, fijada en 0.2.
- **Número de evaluaciones de búsqueda local** (`local_search_evaluations`): Máximo de evaluaciones locales por individuo, fijado en 10.
- **Tamaño del vecindario en búsqueda local** (`local_search_neighbor_size`): Número de cambios permitidos en cada iteración de búsqueda local. Fijado en 5.
- **Ajuste dinámico del tamaño del subconjunto** (`adjust_size`): Permite modificar el tamaño del subconjunto durante la ejecución.

La combinación de estos parámetros permite un control detallado del proceso de selección de instancias y del entrenamiento de los modelos, asegurando resultados reproducibles, eficientes y de calidad.

Capítulo 8

Resultados y Análisis

En este capítulo se exponen los experimentos realizados y los resultados obtenidos en los distintos escenarios evaluados. El objetivo principal fue analizar el rendimiento de los modelos entrenados con conjuntos de datos reducidos, seleccionados mediante los distintos algoritmos desarrollados a lo largo del proyecto.

8.1. Resultados con el conjunto completo

Como punto de partida, se evalúa el rendimiento de los modelos convolucionales entrenados con el **100 %** del conjunto de datos. Esta prueba sirve como referencia para contrastar los resultados obtenidos mediante las técnicas de reducción aplicadas posteriormente. Se emplean los modelos **ResNet50** y **MobileNetV2**, y se miden las métricas de *accuracy*, *precision*, *recall* y *F1-score* sobre el conjunto de validación.

Duración por Eval.	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)
Modelo ResNet50				
00:02:42	87,90 %	88,96 %	87,90 %	87,81 %
Modelo MobileNet				
00:03:16	78,60 %	81,55 %	78,60 %	77,68 %

Tabla 8.1: Comparativa de resultados del **100 %** con los modelos **ResNet50** y **MobileNet**.

Los resultados presentados en la tabla 8.1, y visualizados en la figura 8.1, representan el rendimiento máximo alcanzable en condiciones ideales, es

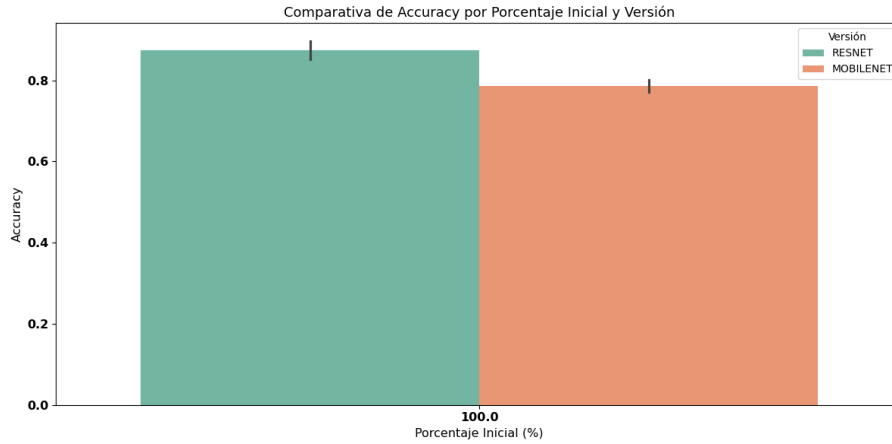


Figura 8.1: Comparación de *accuracy* de los modelos ResNet50 y MobileNetV2 usando el 100 % del conjunto de datos.

decir, utilizando la totalidad del conjunto de datos sin aplicar técnicas de reducción. Este escenario establece un techo de rendimiento que se utiliza como referencia para evaluar la eficacia de los algoritmos de reducción de datos implementados en el resto del estudio.

En particular, se observa que **ResNet50** obtiene mejores resultados en todas las métricas evaluadas, destacando especialmente en *accuracy* y *precision*, donde supera por más de 9 puntos porcentuales a **MobileNetV2**. Esta superioridad en el rendimiento viene acompañada de un tiempo por evaluación ligeramente menor (**00:02:42** frente a **00:03:16**), lo que indica una mayor eficiencia en la etapa de predicción.

Esta comparativa inicial permite establecer las bases para el análisis de los resultados obtenidos mediante las técnicas de reducción de datos, sirviendo como referencia para valorar el impacto de las estrategias aplicadas en las secciones posteriores.

8.2. Comparativa inicial de modelos

Antes de aplicar los algoritmos propuestos de reducción de datos, se considera fundamental realizar una comparativa inicial entre distintos modelos de redes neuronales convolucionales para determinar cuál de ellos es el más adecuado para los experimentos. Esta comparación permite identificar la arquitectura que ofrece un mejor equilibrio entre rendimiento y eficiencia computacional, estableciendo una base sólida sobre la que construir los siguientes análisis.

Para llevar a cabo esta comparativa, se utiliza el enfoque aleatorio (RS) como estrategia de referencia. Al seleccionar subconjuntos de datos de manera aleatoria, sin ninguna optimización, se obtiene una línea base que permite evaluar el comportamiento de cada modelo en condiciones controladas. Esta línea base es especialmente valiosa, ya que ofrece una perspectiva realista del rendimiento mínimo esperable sin aplicar técnicas avanzadas de reducción de datos, sirviendo como punto de partida para comparar las mejoras introducidas por los algoritmos posteriores.

Porcentaje Inicial	Evaluaciones Realizadas	Duración Total	Duración por Eval.	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)
Modelo ResNet50							
10 %	100	01:21:37	00:00:48	85,16 %	86,30 %	85,16 %	85,02 %
25 %	100	01:28:14	00:00:52	87,26 %	88,13 %	87,26 %	87,02 %
50 %	100	02:27:17	00:01:28	88,49 %	89,51 %	88,49 %	88,30 %
75 %	100	03:52:41	00:02:19	89,89 %	90,53 %	89,89 %	89,70 %
100 %	1	-	00:02:55	87,42 %	88,73 %	87,42 %	87,18 %
Modelo MobileNet							
10 %	100	00:38:16	00:00:22	81,34 %	82,46 %	81,34 %	80,52 %
25 %	100	01:18:22	00:00:47	80,70 %	81,90 %	80,70 %	79,90 %
50 %	100	02:26:40	00:01:28	80,86 %	82,65 %	80,86 %	80,22 %
75 %	100	03:05:48	00:01:51	82,26 %	84,16 %	82,26 %	81,67 %
100 %	1	-	00:03:16	78,60 %	81,55 %	78,60 %	77,68 %

Tabla 8.2: Comparativa de resultados de la generación inicial utilizando el RS y el 100 % con los modelos ResNet50 y MobileNet.

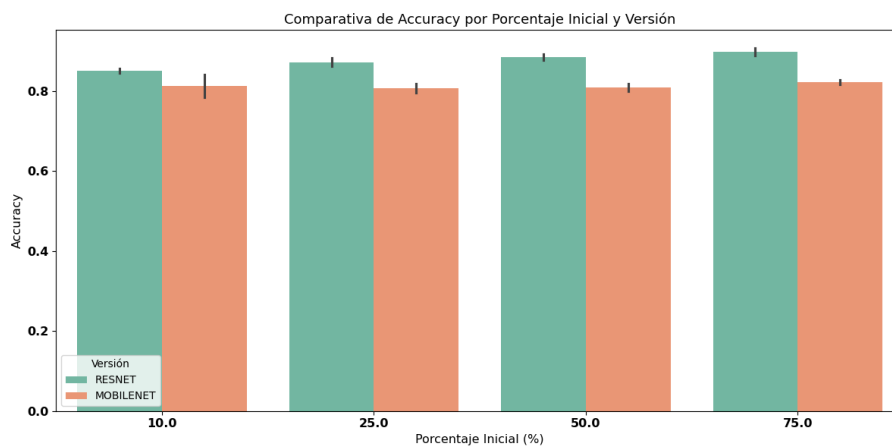


Figura 8.2: Diagrama de barras para comparar los modelos usando el *accuracy* alcanzado por cada porcentaje inicial.

Volver a ejecutar los resultados para comprobar tiempos. Se realizan pruebas con distintos porcentajes iniciales de datos seleccionados aleatoriamente (10 %, 25 %, 50 %, 75 % y 100 %), utilizando tanto ResNet50 como MobileNetV2. Los resultados obtenidos se presentan en la Tabla 8.2, que resume las métricas alcanzadas por cada modelo, y en la Figura 8.2, que muestra la

comparación de los valores de *accuracy* mediante barras separados por los porcentajes iniciales.

Los resultados muestran que ResNet50 logra un mejor rendimiento en términos de *accuracy*, *precision*, *recall* y *F1-score*. Sin embargo, este mejor rendimiento viene acompañado de un tiempo de entrenamiento considerablemente mayor. Por su parte, MobileNetV2 ofrece una solución más eficiente en cuanto a tiempos de ejecución, a costa de una ligera pérdida de precisión, lo que la convierte en una opción atractiva para entornos con recursos computacionales limitados.

En conjunto, esta comparativa inicial permite elegir MobileNetV2 como modelo principal para el resto de experimentos. Aunque ResNet50 ofrece una precisión ligeramente superior, su elevado coste computacional no justifica su uso en las pruebas de reducción de datos, donde la eficiencia es un factor clave. Además, el uso del RS como referencia permite establecer un punto de comparación para evaluar las mejoras que los algoritmos propuestos aportarían sobre esta línea base.

8.3. Resultados de la búsqueda local

Como se describe en el apartado correspondiente (ver Sección 5.2), la búsqueda local permite mejorar progresivamente una solución inicial mediante pequeñas modificaciones guiadas por el rendimiento. En este apartado se evalúa su efectividad como alternativa más estructurada frente al enfoque aleatorio, pero sin llegar a la complejidad de los algoritmos evolutivos. Su inclusión busca analizar hasta qué punto una estrategia simple pero guiada puede generar subconjuntos de datos más representativos y consistentes.

En la Figura 8.3 se muestran los resultados mediante un boxplot que permite observar la distribución completa de valores obtenidos en las distintas ejecuciones.

Se puede apreciar que el algoritmo de búsqueda local (LS) mejora claramente la mediana del *accuracy* respecto al enfoque aleatorio (RS). Mientras que el RS se sitúa en torno a una mediana de **0.787**, la LS alcanza una mediana superior, próxima a **0.818**. Esta diferencia refleja una mayor capacidad del algoritmo local para generar subconjuntos más representativos y eficaces.

Además, los valores máximos que alcanzan ambos algoritmos son similares (en torno a **0.858-0.860**), pero la LS muestra una dispersión más acotada hacia valores altos, lo que sugiere mayor estabilidad en sus resultados. En cambio, el RS presenta una mayor dispersión hacia valores bajos y aunque presente una mayor sensibilidad a la aleatoriedad de las selecciones, puede

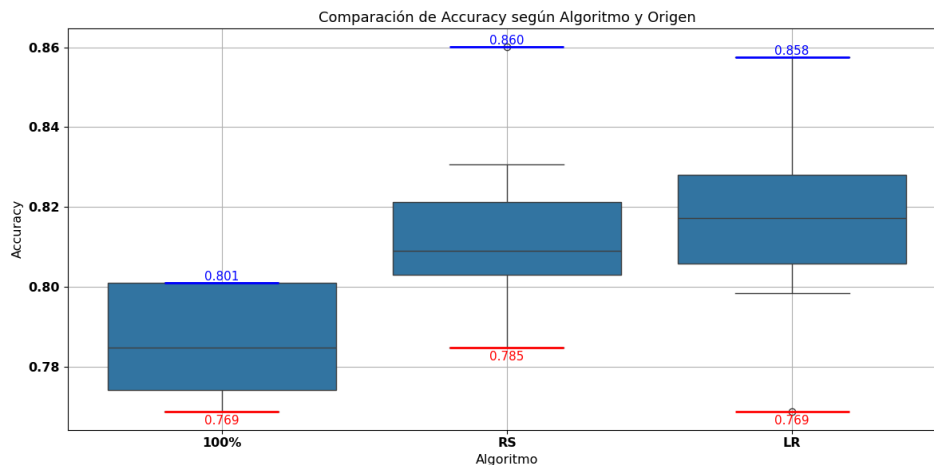


Figura 8.3: Boxplot comparando el RS con la LS usando *accuracy*.

llegar a tener mejores valores mínimos, como se evidencia en su menor valor mínimo (**0.785** frente a **0.769** en LS), pero siendo el de la LS un valor atípico.

Esto pone de manifiesto que, aunque el RS puede ocasionalmente alcanzar buenos resultados, la LS ofrece una mejor consistencia y fiabilidad, con menos varianza entre ejecuciones y una tendencia general a obtener subconjuntos de entrenamiento más efectivos.

8.4. Resultados del Algoritmo Genético

Con el objetivo de superar las limitaciones observadas (como el riesgo de estancamiento o la exploración poco estructurada del espacio de soluciones) se incorpora un enfoque evolutivo más completo: el algoritmo genético (GA), descrito en la Sección 5.3, el cual sirve como punto de partida para explorar la aplicación de estrategias metaheurísticas en la selección de subconjuntos representativos de imágenes. Su estructura evolutiva, basada en selección por torneo, cruce e incorporación de mutación, ofrece ya desde sus primeras versiones una capacidad superior para generalizar, en comparación con métodos más simples como la LS.

Tal como se observa en la Figura 8.4, el algoritmo genético (GA) consigue una **mediana de *accuracy*** más alta que la búsqueda local (LS), reflejando un rendimiento medio más consistente. Además, presenta un valor máximo superior (alcanza hasta **0.876**), lo que evidencia su mayor potencial para encontrar soluciones de alta calidad.

No obstante, también se aprecia una ligera mayor dispersión en los re-

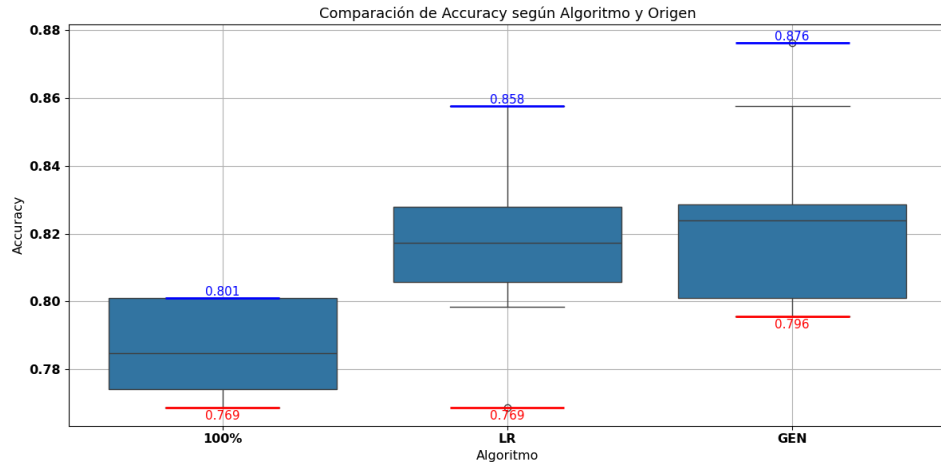


Figura 8.4: Boxplot comparando LS con GA usando *accuracy*.

sultados del GA, particularmente hacia los valores bajos. Esto indica que, pese a su capacidad exploratoria, el GA puede generar soluciones poco efectivas si no se controlan adecuadamente ciertos operadores como el cruce o la mutación. De hecho, su valor mínimo (**0.796**) es superior al de la LS en esta comparativa, pero deja margen para mejoras en la presión selectiva o en mecanismos que eviten estancamientos.

La LS, por su parte, mantiene un comportamiento más estable, aunque con una mediana ligeramente inferior. Su distribución es más concentrada y limitada en el extremo superior, lo que evidencia su carácter más explotador pero con menor capacidad para alcanzar soluciones óptimas globales.

Estos resultados sirven como evidencia empírica para continuar desarrollando nuevas versiones del GA, incorporando mejoras específicas en sus operadores con el fin de aprovechar su capacidad exploratoria y, al mismo tiempo, mitigar sus limitaciones.

8.5. Mejorando el operador de cruce

La primera mejora introducida al GA consiste en reemplazar el cruce aleatorio por un cruce ponderado, donde se prioriza la contribución del progenitor con mayor *fitness*. Además, se incorpora una estrategia selectiva que conserva únicamente el mejor de los dos hijos generados en cada cruce. Ambos cambios, explicados en detalle en la Sección 5.4, buscan aumentar la presión evolutiva y acelerar la convergencia hacia soluciones de mayor calidad, evitando así que soluciones mediocres se propaguen innecesariamente en la población.

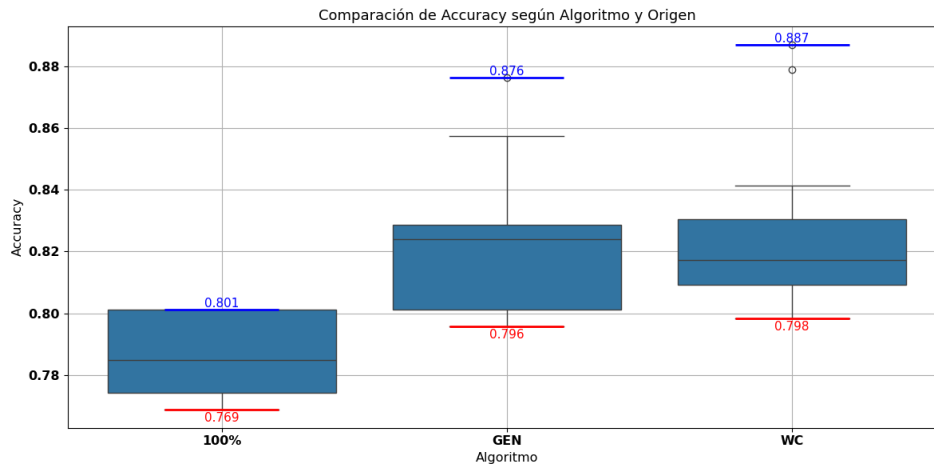


Figura 8.5: Boxplot de *accuracy* comparando el GA y el GA-WC.

Tal como se observa en la Figura 8.5, esta modificación produce una mejora clara en la calidad y estabilidad de los resultados. La versión con cruce ponderado (GA-WC) alcanza un valor máximo atípico superior (**0.887**), aunque presenta una mediana más baja que la versión básica (GA).

Pero por otra parte, se aprecia una ligera reducción en la dispersión de los valores inferiores, con un mínimo de **0.798** frente al **0.796** en la versión anterior, lo que sugiere una mayor consistencia. Aunque el IQR (Rango Intercuartílico) sigue siendo amplio, la acumulación de valores más cercanos al rango superior refleja una convergencia evolutiva más enfocada y menos dependiente del azar.

En conjunto, esta mejora en el cruce no solo permite una transferencia más eficiente de características ventajosas, sino que también incrementa la presión selectiva sobre la calidad de las soluciones. Esto se traduce en un comportamiento más robusto, menos propenso a resultados erráticos y con una mayor capacidad de exploración dirigida del espacio de soluciones.

8.6. Mejorando el operador de mutación

A partir de los resultados obtenidos con el GA-WC, se evalúa una nueva versión en la que se introdujo una **mutación adaptativa** (ver Sección 5.5). A diferencia de la versión original con tasa fija, esta estrategia ajusta el número de intercambios en función del tamaño del subconjunto mutado, lo que permite una mayor flexibilidad en escenarios de diferente escala.

Como se aprecia en la Figura 8.6, ambos algoritmos presentan distribuciones de *accuracy* muy similares en términos de mediana e IQR. El GA-WC

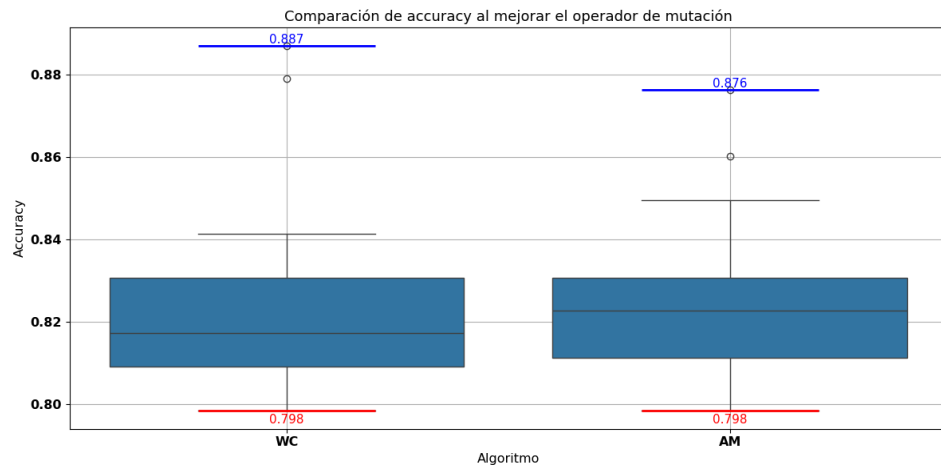


Figura 8.6: Comparación de *accuracy* entre el GA-WC y el GA-AM.

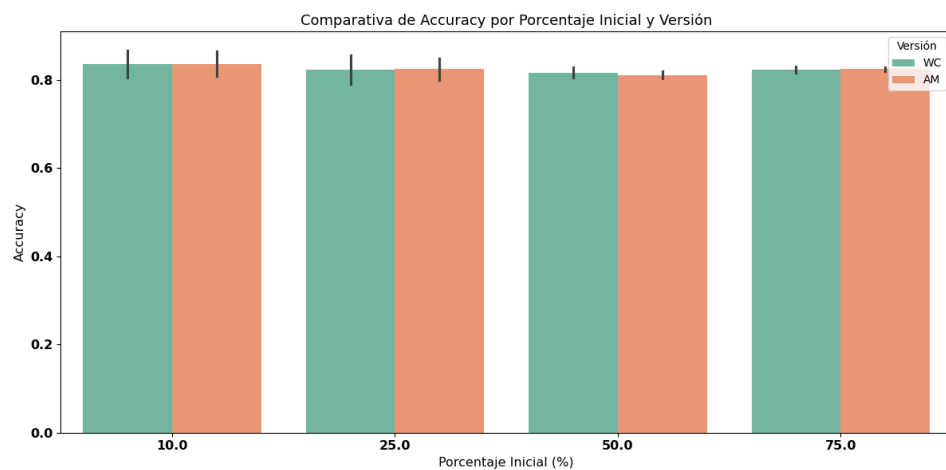


Figura 8.7: Diagrama de comparación usando el *accuracy* entre el GA-WC Y el GA-AM, separados por porcentaje inicial.

mantiene un ligero máximo superior, alcanzando un **0.887** frente al **0.876** del GA-AM. Sin embargo, el algoritmo con mutación adaptativa muestra una distribución más compacta en la parte central, con menor dispersión hacia valores bajos, lo que sugiere una mayor consistencia entre ejecuciones.

La principal diferencia se observa en la robustez de los resultados: el GA-AM presenta una distribución más estable, con menos valores atípicos y una mayor concentración de ejecuciones cerca del cuartil superior. Este comportamiento indica una menor propensión a caídas abruptas de rendimiento y refuerza la idea de que la mutación adaptativa mejora la estabilidad del proceso evolutivo.

Al analizar los resultados por porcentaje inicial (Figura 8.7), se confirma que el GA-AM mantiene una estabilidad más uniforme en todos los escenarios, mientras que el GA-WC muestra una mayor variabilidad, especialmente en configuraciones con menor cantidad de datos. Aunque las diferencias en *accuracy* medio no son significativas, la menor dispersión y la reducción de valores extremos justifican la preferencia por la versión adaptativa en contextos donde la fiabilidad es prioritaria.

En conjunto, la mutación adaptativa no genera una mejora radical en precisión media, pero sí contribuye a una evolución más controlada y menos susceptible a degradaciones, favoreciendo una mayor consistencia en los resultados. Además, al adaptarse al tamaño del subconjunto, evita configuraciones subóptimas que podrían surgir con una tasa fija de mutación, lo que la hace especialmente adecuada para escenarios con escalas variables.

Esta mejora consolida la capacidad del algoritmo para equilibrar exploración y explotación, y lo posiciona como una alternativa más robusta y estable en tareas de reducción de datos para aprendizaje profundo.

8.7. Resultados del reinicio poblacional

La última mejora realizada de los algoritmos genéticos (ver Sección 5.6) introduce una lógica de reinicio poblacional diseñada para evitar estancamientos evolutivos. El algoritmo monitoriza el rendimiento del segundo mejor individuo, y si este no mejora durante dos generaciones consecutivas, se aplica un reinicio parcial que conserva únicamente al mejor individuo de la población.

Los resultados, mostrados en la Figura 8.8 y la Tabla 8.3, permiten extraer varias conclusiones relevantes. En primer lugar, el **valor máximo de accuracy** es idéntico para GA-AM y PR, alcanzando ambos un **0.876**, lo que sugiere que el reinicio no favorece la aparición de soluciones de mayor calidad. Por el contrario, el **valor mínimo de accuracy** observado en PR

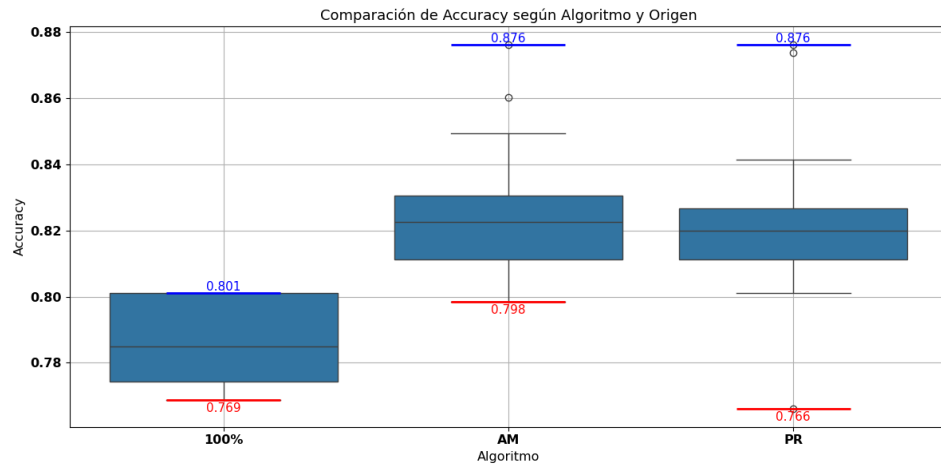


Figura 8.8: Comparación de *accuracy* entre el GA-AM y con PR.

Algoritmo	Duración Total	Duración por Eval.	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)	Evaluaciones	Porc.Paper	Porc.Rock	Porc.Scissors
10 %										
GA-AM	00:27:39	00:00:17	83,66 %	84,24 %	83,66 %	83,09 %	100	35,64 %	30,87 %	33,49 %
PR	00:29:02	00:00:17	81,93 %	83,07 %	81,93 %	81,19 %	100	34,52 %	32,14 %	33,33 %
25 %										
GA-AM	00:57:29	00:00:34	82,42 %	83,68 %	82,42 %	81,78 %	100	33,33 %	32,54 %	34,13 %
PR	01:00:22	00:00:36	82,64 %	83,78 %	82,64 %	81,96 %	100	33,62 %	33,27 %	33,11 %
50 %										
GA-AM	01:47:31	00:01:05	81,13 %	82,62 %	81,13 %	80,41 %	100	33,94 %	32,76 %	33,30 %
PR	01:53:13	00:01:08	81,24 %	83,09 %	81,24 %	80,59 %	100	33,21 %	33,20 %	33,59 %
75 %										
GA-AM	02:36:41	00:01:34	82,42 %	83,94 %	82,42 %	81,85 %	100	33,67 %	33,11 %	33,22 %
PR	02:41:24	00:01:37	82,69 %	84,33 %	82,69 %	82,06 %	100	33,52 %	33,44 %	33,04 %
100 %										
100 %	-	00:03:16	78,60 %	81,55 %	78,60 %	77,68 %	1	33,33 %	33,33 %	33,33 %

Tabla 8.3: Resultados de los algoritmos GA-AM y PR por porcentaje inicial, incluyendo distribución de clases y duración total y por evaluación.

(**0.766**) es notablemente más bajo que el de **GA-AM** (**0.798**), lo que indica una mayor propensión a obtener ejecuciones de bajo rendimiento cuando se utiliza la estrategia de reinicio.

La distribución general de resultados revela que el PR no logra una reducción significativa en la dispersión de los valores: aunque los cuartiles y medianas de **GA-AM** y **PR** son similares, se observa un leve desplazamiento hacia valores ligeramente más bajos en **PR**, lo cual es especialmente evidente en el escenario de menor porcentaje inicial (10 %). Esta tendencia sugiere que el reinicio poblacional, al reintroducir diversidad de forma abrupta, puede generar soluciones subóptimas que no contribuyen de manera sustancial a mejorar la población.

La tabla de resultados 8.3 respalda estas observaciones: en promedio, el algoritmo **GA-AM** presenta una ligera superioridad en precisión media (**83,66 %** frente a **81,93 %** en el 10 %), así como en F1-score y recall. Estas diferencias, aunque no drásticas, son consistentes en la mayoría de los escenarios. Además, las métricas de distribución de clases y las duraciones de ejecución son prácticamente equivalentes entre **GA-AM** y **PR**, lo que refuerza la idea de que el impacto del reinicio poblacional no justifica su complejidad añadida.

En resumen, aunque el reinicio poblacional busca aumentar la exploración y evitar el estancamiento, en esta implementación concreta no aporta mejoras tangibles en términos de precisión ni estabilidad, e incluso puede introducir soluciones más erráticas en ciertas ejecuciones.

8.8. Resultados con versiones libres

Como parte de la evolución de los algoritmos desarrollados, se propone la creación de **versiones libres**, en las que el tamaño del subconjunto seleccionado no permanece fijo durante la ejecución, sino que puede ajustarse de forma dinámica en función de las decisiones evolutivas. Esta flexibilidad permite que los algoritmos modifiquen el número de datos utilizados a medida que avanzan las generaciones, adaptándose de manera más natural a las características del problema.

Para evaluar el impacto de esta modificación, se generan versiones libres tanto para el algoritmo de búsqueda local (**LS**) como para el algoritmo con mutación adaptativa (**GA-AM**). En el caso de la **LS**, además, se incorpora una variación adicional: el porcentaje inicial de imágenes no es fijo, sino que se selecciona aleatoriamente en cada ejecución, introduciendo así un mayor grado de aleatoriedad y diversidad en el proceso.

En la Figura 8.9 se observa que las versiones libres mantienen un ren-

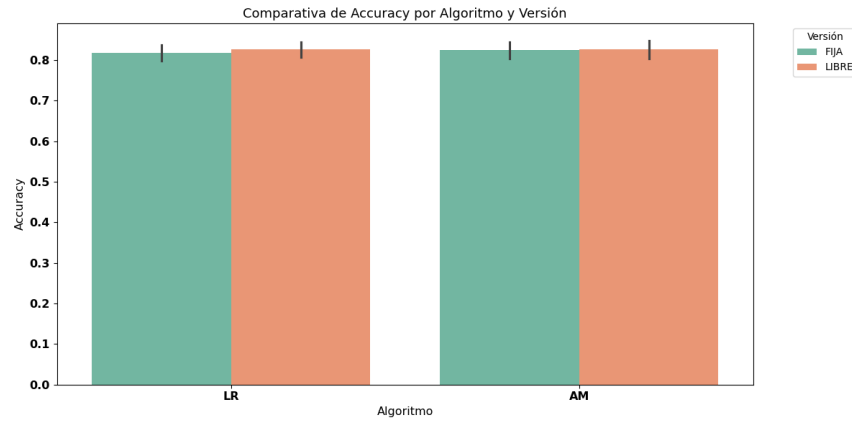


Figura 8.9: Comparación de *accuracy* entre el LS, GA-AM y sus versiones Libres (formato BARPLOT).

dimiento promedio muy similar al de las versiones originales. La precisión media (*accuracy*) de las versiones libres es ligeramente superior en algunos casos, lo que indica que la capacidad de adaptación no introduce pérdidas de rendimiento e incluso puede aportar pequeñas mejoras en escenarios específicos. Las barras de error sugieren que la variabilidad entre ejecuciones se mantiene controlada, lo que refuerza la idea de que la flexibilidad no compromete la estabilidad de los resultados.

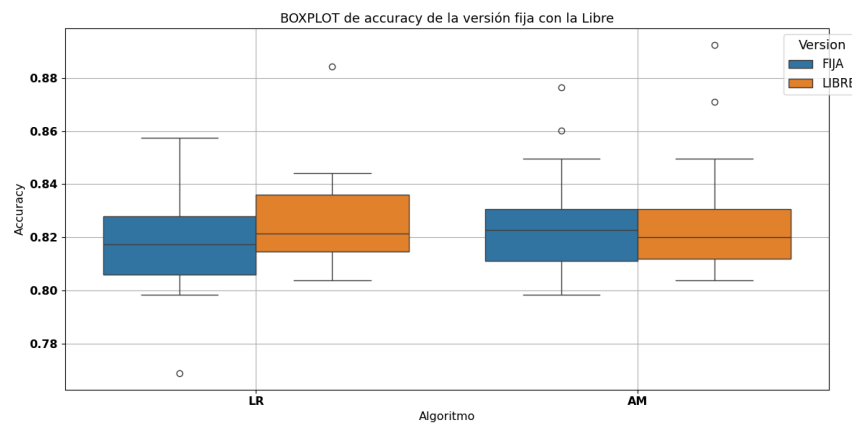


Figura 8.10: Comparación de *accuracy* entre el LS, GA-AM y sus versiones Libres en formato BOXPLOT.

Los diagramas de caja en la Figura 8.10 permiten una comparación más detallada de la dispersión de los resultados. En el caso de la LS, la versión libre muestra una ligera mejora en la mediana y un rango intercuartílico (IQR) más compacto, lo que indica una mayor estabilidad. Además, se observan

valores atípicos superiores que sugieren la posibilidad de obtener ejecuciones con precisión especialmente alta. En el caso del **GA-AM**, las diferencias entre las versiones fija y libre son más sutiles: aunque las medianas son prácticamente idénticas, la versión libre presenta una ligera reducción en los valores mínimos, lo que podría reflejar una mayor adaptabilidad frente a escenarios más difíciles.

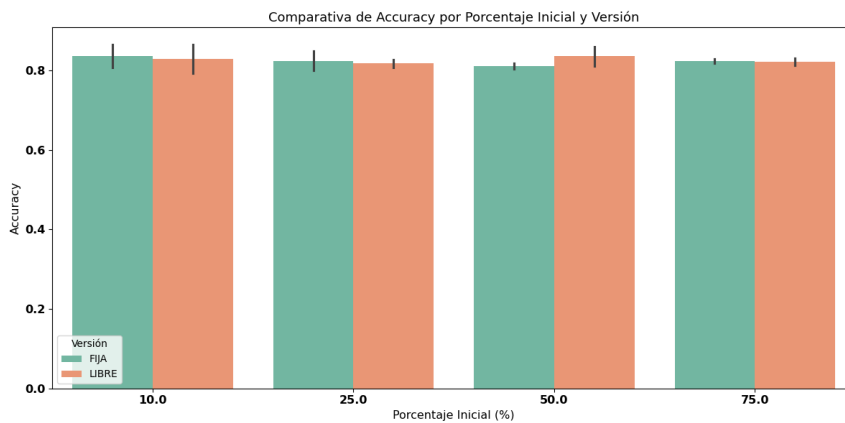


Figura 8.11: Comparación de *accuracy* en función del porcentaje inicial para el algoritmo **GA-AM**.

La Figura 8.11 muestra la evolución del *accuracy* en función del porcentaje inicial de datos utilizados en el algoritmo **GA-AM**. Se observa que la precisión se mantiene estable a lo largo de los distintos puntos de partida, con ligeras variaciones que no afectan de manera significativa al comportamiento general del algoritmo. Este resultado confirma que la flexibilidad en el tamaño del subconjunto no introduce un sesgo negativo en la calidad de las soluciones encontradas, sino que permite adaptarse a diferentes condiciones iniciales sin comprometer el rendimiento.

Por otro lado, la Figura 8.12 demuestra que tanto las versiones fijas como las libres preservan una distribución equilibrada de las clases **Rock**, **Paper** y **Scissors**. Las proporciones entre clases se mantienen estables, sin que la flexibilidad en el tamaño del subconjunto genere desbalances significativos. Este aspecto es crucial, ya que garantiza la validez de los experimentos y asegura que las mejoras observadas no son producto de un sesgo de clase inadvertido.

Finalmente, las Figuras 8.13 y 8.14 permiten visualizar la relación entre el *accuracy* y el porcentaje final de datos seleccionados en las versiones libres. En el caso de la búsqueda local libre (**LS-F**), se observa una tendencia a incrementar ligeramente el tamaño del subconjunto final, especialmente en ejecuciones con mejor precisión. Por el contrario, el algoritmo **GA-AM-F** muestra una distribución más equilibrada, con una mayor variabilidad en

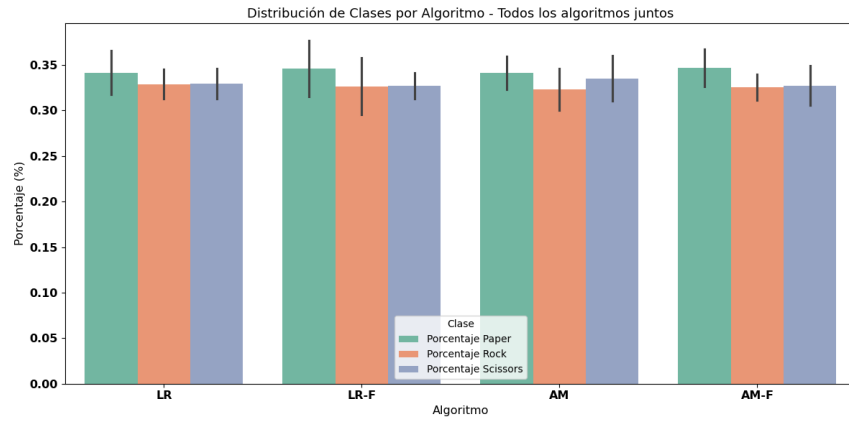


Figura 8.12: Distribución de clases en los subconjuntos generados por los algoritmos estándar y libres.

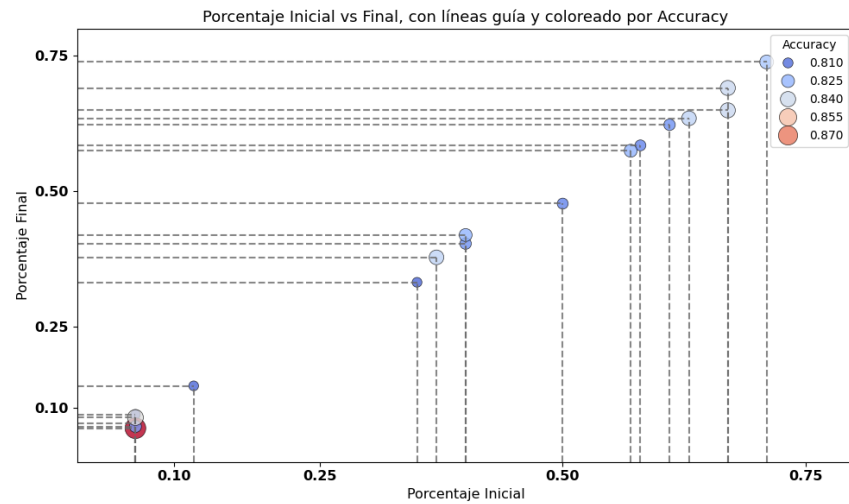


Figura 8.13: Relación entre *accuracy* y porcentaje final de datos seleccionados por el algoritmo LS-F.

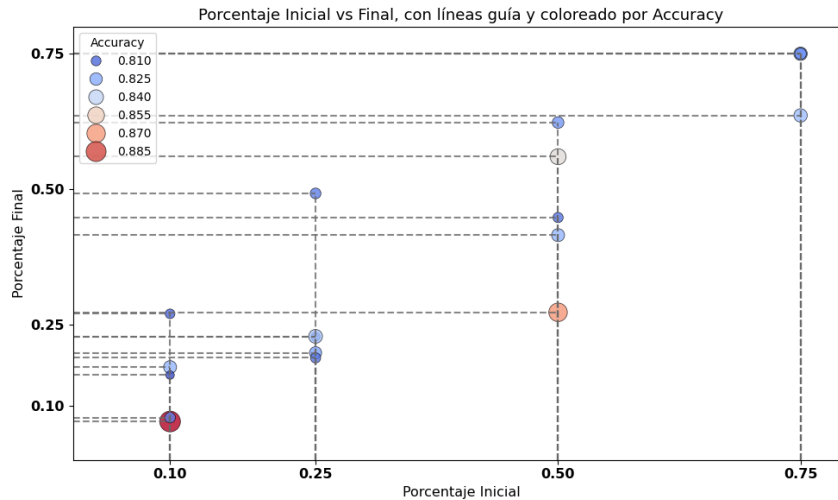


Figura 8.14: Relación entre *accuracy* y porcentaje final de datos seleccionados por el algoritmo GA-AM-F.

los tamaños finales, lo que refleja su capacidad adaptativa para ajustarse a diferentes condiciones de búsqueda.

Este comportamiento evidencia que los algoritmos libres no solo adaptan la composición del subconjunto, sino también su escala, ajustando dinámicamente la cantidad de datos utilizados según las necesidades del proceso evolutivo. En conjunto, estas versiones demuestran ser una alternativa más versátil y robusta, capaz de mantener la calidad del rendimiento incluso en escenarios con alta variabilidad en los datos o restricciones de tamaño no conocidas de antemano.

8.9. Resultados del algoritmo memético

Finalmente, se evalúa el algoritmo memético (MA) (ver Sección 5.7), el cual combina la evolución genética con una búsqueda local aplicada de forma probabilística sobre ciertos individuos seleccionados. Este enfoque híbrido busca equilibrar la exploración del espacio de soluciones con una intensificación localizada, ofreciendo mejoras tanto en precisión como en estabilidad.

Tal como se aprecia en la Figura 8.15, el algoritmo memético (MA) supera claramente al mejor de los enfoques genéticos (GA-AM), alcanzando una mediana más elevada y un valor máximo de *accuracy* de hasta **0.903**. Su distribución es más compacta, con menor dispersión hacia los valores bajos, lo que refleja una mayor consistencia entre ejecuciones.

La versión libre del memético, que incorpora también un ajuste dinámico

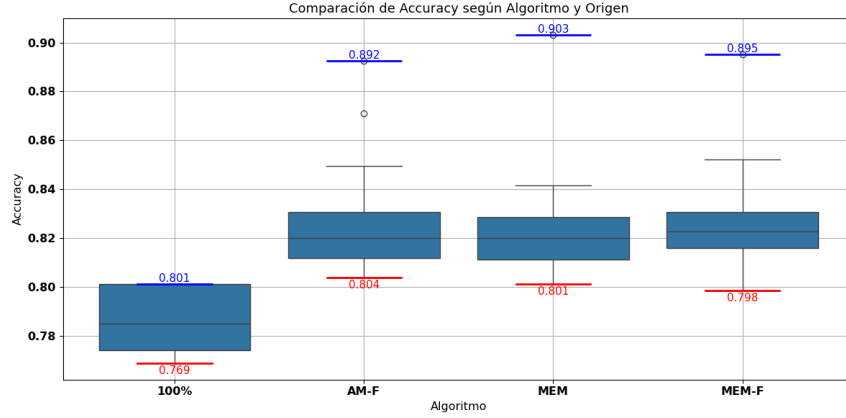


Figura 8.15: Comparación de *accuracy* entre el algoritmo memético (MA) y su versión libre.

del tamaño del subconjunto (ver Apartado 5.8.3), muestra un rendimiento muy similar al estándar, con una ligera reducción en el valor máximo pero una estabilidad comparable. Esto sugiere que el componente adaptativo no penaliza la calidad de las soluciones y puede incluso aportar mayor flexibilidad en entornos más inciertos.

Algoritmo	Duración Total	Duración por Eval.	Porc. Final	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)	Evaluaciones
10 %								
GA-AM-F	00:34:33	00:00:21	14,94 %	82,96 %	84,07 %	82,96 %	82,12 %	100
MA	00:27:43	00:00:17	10,00 %	82,85 %	84,44 %	82,85 %	81,91 %	100
MA-F	00:32:39	00:00:20	9,62 %	83,28 %	84,70 %	83,28 %	82,46 %	100
25 %								
GA-AM-F	01:07:13	00:00:40	26,68 %	81,77 %	82,66 %	81,77 %	81,18 %	100
MA	00:57:12	00:00:34	25,00 %	81,99 %	83,11 %	81,99 %	81,21 %	100
MA-F	01:08:54	00:00:41	27,18 %	82,80 %	83,77 %	82,80 %	82,17 %	100
50 %								
GA-AM-F	01:51:24	00:01:07	46,36 %	83,60 %	85,22 %	83,60 %	83,03 %	100
MA	01:47:32	00:01:05	50,00 %	81,51 %	83,36 %	81,51 %	80,89 %	100
MA-F	02:16:51	00:01:22	67,07 %	81,61 %	83,09 %	81,61 %	80,96 %	100
75 %								
GA-AM-F	02:33:25	00:01:32	72,72 %	82,20 %	83,94 %	82,20 %	81,47 %	100
MA	02:37:43	00:01:35	75,00 %	82,74 %	84,33 %	82,74 %	82,14 %	100
MA-F	03:08:21	00:01:53	71,16 %	82,69 %	84,18 %	82,69 %	82,10 %	100
100 %								
100 %	-	00:03:16	100,00 %	78,60 %	81,55 %	78,60 %	77,68 %	1

Tabla 8.4: Resultados de los MA y del GA-AM por porcentaje inicial.

El análisis detallado de la Tabla 8.4 confirma estas observaciones y permite extraer conclusiones más matizadas sobre el comportamiento de los algoritmos. En primer lugar, el MA-F mantiene una precisión media comparable o superior a la del MA en la mayoría de los escenarios, especialmente en configuraciones con porcentajes iniciales más bajos (10 % y 25 %). Esto indica que la capacidad de ajuste dinámico no solo no perjudica el rendimiento,

sino que puede ofrecer ventajas en términos de adaptabilidad, permitiendo al algoritmo optimizar la selección de datos sin depender de un tamaño fijo preestablecido.

Además, el porcentaje final alcanzado por MA-F muestra una notable variabilidad según el escenario: en los casos con menor porcentaje inicial, tiende a mantenerse cercano al valor de partida (por ejemplo, un 9,62 % en el escenario del 10 %), mientras que en configuraciones más amplias (50 % y 75 %), el tamaño final puede incrementarse significativamente (hasta un 67,07 % en el caso del 50 %). Este comportamiento adaptativo refleja que el algoritmo es capaz de ajustar la escala de la solución en función de las características del espacio de búsqueda, evitando tanto la sobrerrepresentación como la selección excesiva de datos innecesarios.

Por otro lado, la duración total y el tiempo medio por evaluación aumentan con el porcentaje inicial, pero esto no compromete la eficiencia general del MA-F, que mantiene un equilibrio adecuado entre precisión y coste computacional.

De esta forma, tanto el MA como su variante libre, MA-F, se consolidan como las estrategias más eficaces del estudio, ofreciendo una combinación robusta de rendimiento, adaptabilidad y estabilidad.

8.10. Resultados finales entre enfoques

Tras analizar el rendimiento individual de cada enfoque a lo largo de las secciones anteriores, en esta sección se realiza una síntesis comparativa entre los principales algoritmos desarrollados: el algoritmo genético con cruce ponderado (GA-WC), el algoritmo genético con mutación adaptativa (GA-AM), el algoritmo memético (MA) y su versión libre (MA-F). Se incluyen además las referencias al 100 % del conjunto de datos y a la selección aleatoria (RS) como líneas base para contextualizar los resultados.

El objetivo es identificar cuál de estos enfoques logra el mejor compromiso entre precisión (*accuracy*), estabilidad de resultados, y eficiencia en la reducción de datos, así como validar la hipótesis de que una selección inteligente de ejemplos puede superar incluso al uso del 100 % del conjunto de datos.

8.10.1. Análisis comparativo de *accuracy*

El análisis comparativo de *accuracy* entre algoritmos confirma de manera general las conclusiones obtenidas en los análisis de las secciones anteriores. Como muestra el boxplot 8.16, los algoritmos meméticos, especialmente la

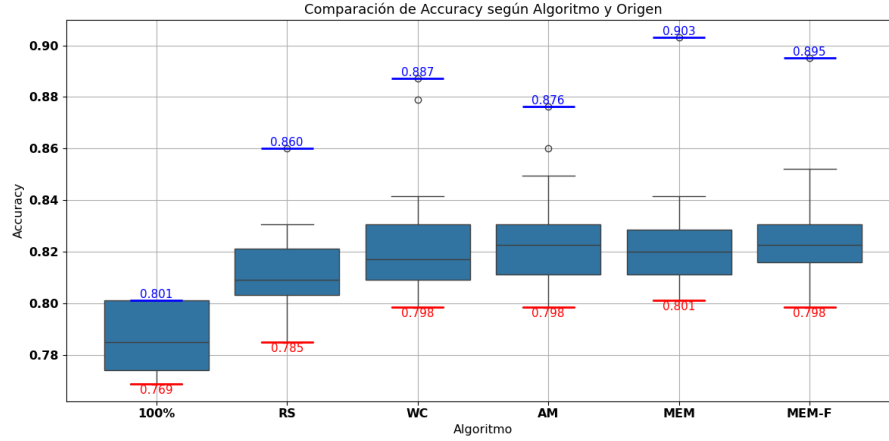


Figura 8.16: Boxplot de *accuracy* para los algoritmos RS, GA-WC, GA-AM, MA y MA-F.

versión libre MA-F, presentan las mejores métricas globales en términos de precisión y estabilidad. MA-F alcanza las medianas más altas y muestra una distribución más compacta, lo que refleja una mayor consistencia entre ejecuciones. Además, la dispersión de resultados en MA-F es menor, lo que sugiere una robustez notable frente a la aleatoriedad de las ejecuciones y una mayor fiabilidad en su rendimiento.

Por otro lado, el algoritmo memético estándar MA también destaca por su rendimiento superior, aunque presenta una ligera mayor variabilidad en algunos casos. Aun así, se posiciona como una solución altamente efectiva, confirmando que la combinación de evolución global y búsqueda local permite generar subconjuntos más representativos y eficientes.

En contraste, los algoritmos RS y GA-WC muestran una mayor dispersión en los resultados, lo que indica una menor estabilidad y una mayor dependencia de la aleatoriedad en la selección de datos. RS, como se ha observado a lo largo de todo el análisis, tiende a obtener resultados más erráticos, mientras que GA-WC, aunque introduce mejoras respecto al RS, sigue sin alcanzar la solidez de los algoritmos meméticos.

Por su parte, el algoritmo GA-AM presenta un rendimiento intermedio: mejora las métricas obtenidas por RS y GA-WC, pero no logra superar la precisión ni la estabilidad de los algoritmos meméticos. Esto refuerza la hipótesis planteada a lo largo del análisis de los resultados: los enfoques basados en algoritmos meméticos, especialmente en su versión libre, permiten alcanzar una combinación óptima de precisión, reducción de datos y estabilidad en problemas de selección de instancias para modelos de aprendizaje profundo.

8.10.2. Porcentaje final de datos seleccionados

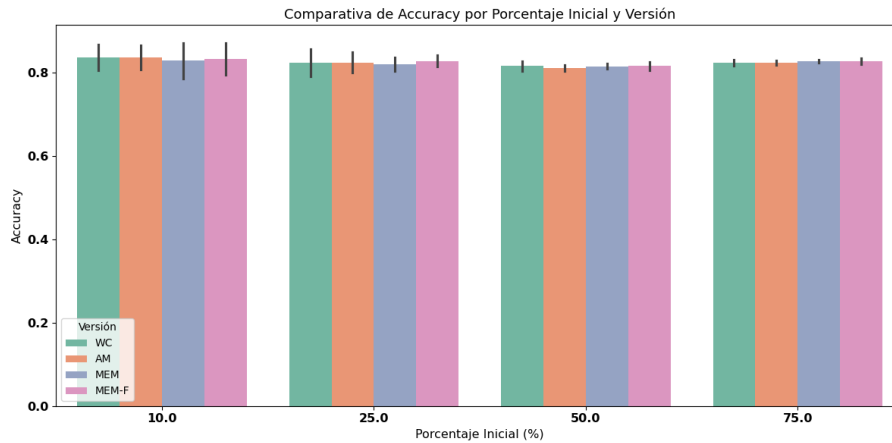


Figura 8.17: Porcentaje final de datos seleccionados por cada algoritmo.

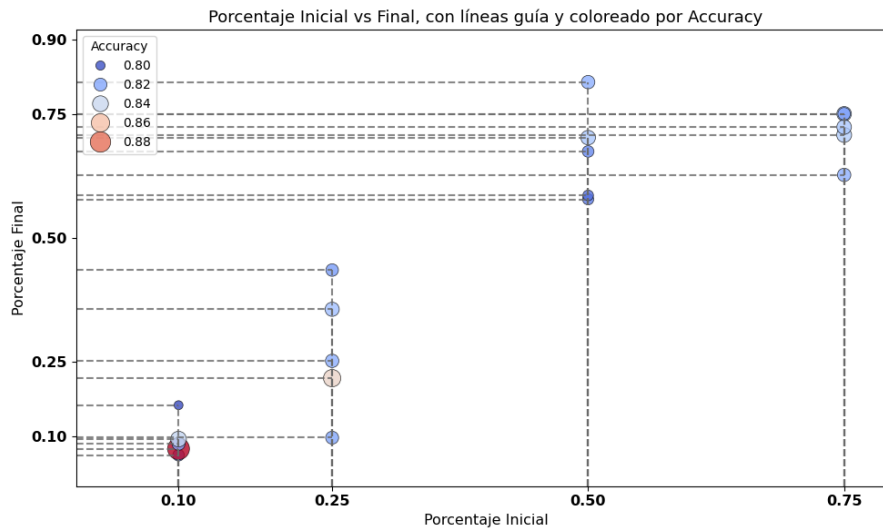


Figura 8.18: Relación entre *accuracy* y porcentaje final de datos seleccionados por cada algoritmo.

El análisis del porcentaje final de datos seleccionados permite evaluar la eficiencia de cada algoritmo, entendida como la capacidad para mantener una precisión elevada reduciendo el volumen de datos utilizados. Como se observa en la Figura 8.17, los algoritmos meméticos, y en particular MA-F, logran mantener o incluso reducir de forma significativa el tamaño del subconjunto de entrenamiento en comparación con los enfoques basados en algoritmos genéticos como GA-WC y GA-AM.

En concreto, **MA-F** destaca por su capacidad para obtener subconjuntos más compactos sin comprometer la calidad del modelo, seleccionando en promedio un 70 % o menos del conjunto original de datos. Este comportamiento es consistente a lo largo de las distintas configuraciones, y sugiere que **MA-F** es capaz de adaptarse dinámicamente al tamaño de la solución en función de la complejidad del problema. Esta flexibilidad resulta especialmente valiosa, ya que permite optimizar el uso de recursos computacionales sin sacrificar rendimiento.

Por otro lado, los algoritmos **GA-WC** y **GA-AM** requieren porcentajes más elevados para alcanzar resultados competitivos, lo que refleja una menor eficiencia en la reducción de datos. El algoritmo **RS**, al no aplicar estrategias de optimización específicas, presenta una mayor dispersión y una dependencia directa de la aleatoriedad en la selección de instancias.

La Figura 8.18 complementa este análisis mostrando la relación entre *accuracy* y el porcentaje final de datos seleccionados. Se observa que **MA** y **MA-F** consiguen altos valores de *accuracy* utilizando porcentajes de datos más reducidos, validando la hipótesis de que es posible mejorar la calidad del modelo mediante una selección optimizada. En contraste, los enfoques menos sofisticados requieren porcentajes más altos para alcanzar precisiones similares, lo que demuestra la ventaja competitiva de los algoritmos meméticos.

8.10.3. Estabilidad en la distribución de clases

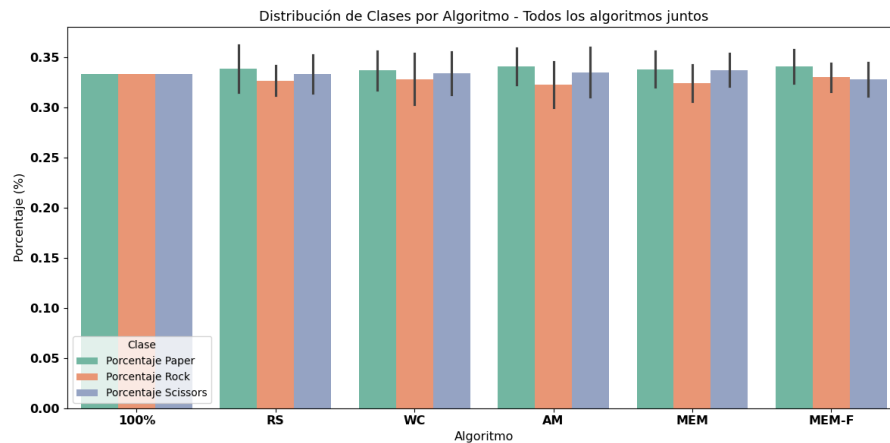


Figura 8.19: Distribución de clases seleccionadas por cada algoritmo en el conjunto reducido final.

El análisis de la distribución de clases en los subconjuntos finales permite evaluar si los algoritmos de reducción mantienen un equilibrio adecuado en-

tre las diferentes clases presentes en el dataset original. Como se observa en la Figura 8.19, todos los algoritmos tienden a preservar un reparto razonablemente balanceado entre las clases *Paper*, *Rock* y *Scissors*, con porcentajes cercanos al 33 %.

Sin embargo, se aprecian ligeras variaciones en la estabilidad de la distribución según el algoritmo utilizado. Los algoritmos **MA** y **MA-F** presentan una mayor consistencia en el reparto de clases, mostrando menores desviaciones y barras de error más reducidas. Esto indica que las soluciones generadas por los enfoques meméticos tienden a conservar de forma más homogénea la diversidad del conjunto original, evitando la infrarepresentación o sobrecarga de clases específicas.

En contraste, los algoritmos **RS**, **GA-WC** y **GA-AM** muestran una mayor dispersión en los porcentajes de las clases, especialmente en el caso de **RS**, lo que refleja una mayor influencia de la aleatoriedad y una menor capacidad para garantizar la estabilidad en la distribución. Esta variabilidad podría generar sesgos no deseados en el modelo final, afectando su rendimiento en tareas de clasificación.

En resumen, los resultados confirman que los algoritmos meméticos, y en particular **MA-F**, no solo son eficaces en términos de precisión y reducción de datos, sino que también logran mantener una distribución de clases equilibrada. Esta propiedad es fundamental para asegurar la generalización del modelo, ya que evita la introducción de desequilibrios que podrían comprometer la calidad del aprendizaje.

8.10.4. Síntesis final

Los resultados globales obtenidos para el conjunto de datos *Rock*, *Paper*, *Scissors* (véase Tabla 8.5) permiten extraer conclusiones claras sobre el rendimiento de los algoritmos evaluados.

En primer lugar, el **MA-F** se consolida como la solución más robusta y eficaz: no solo alcanza la mayor precisión en términos de *accuracy*, sino que también logra mantener un porcentaje reducido de datos seleccionados, optimizando la eficiencia del modelo sin comprometer su rendimiento. Además, **MA-F** presenta una notable estabilidad entre ejecuciones, reflejada en su baja dispersión y en la conservación de una distribución equilibrada de clases en los subconjuntos generados. Estos resultados demuestran su capacidad para superar incluso a la referencia del 100 % de datos, logrando un mejor rendimiento con menos instancias.

El **MA** también ofrece resultados sobresalientes, manteniendo altos niveles de precisión y una considerable reducción de datos. Sin embargo, presenta una ligera mayor variabilidad en comparación con **MA-F**, lo que sugiere una

Algoritmo	Duración Total	Duración por Eval.	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)	Evaluaciones	Porc. Final
10 %								
RS	00:38:17	00:00:22	81,34 %	82,46 %	81,34 %	80,52 %	100	10,00 %
GA-WC	00:34:03	00:00:21	83,60 %	84,38 %	83,60 %	82,82 %	100	10,00 %
GA-AM	00:27:39	00:00:17	83,66 %	84,24 %	83,66 %	83,09 %	100	10,00 %
MA	00:27:43	00:00:17	82,85 %	84,44 %	82,85 %	81,91 %	100	10,00 %
MA-F	00:32:39	00:00:20	83,28 %	84,70 %	83,28 %	82,46 %	100	9,62 %
25 %								
RS	01:18:22	00:00:47	80,70 %	81,90 %	80,70 %	79,90 %	100	25,00 %
GA-WC	01:10:21	00:00:43	82,37 %	83,86 %	82,37 %	81,57 %	100	25,00 %
GA-AM	00:57:29	00:00:34	82,42 %	83,68 %	82,42 %	81,78 %	100	25,00 %
MA	00:57:12	00:00:34	81,99 %	83,11 %	81,99 %	81,21 %	100	25,00 %
MA-F	01:08:54	00:00:41	82,80 %	83,77 %	82,80 %	82,17 %	100	27,18 %
50 %								
RS	02:26:40	00:01:28	80,86 %	82,65 %	80,86 %	80,22 %	100	50,00 %
GA-WC	02:15:43	00:01:21	81,61 %	83,78 %	81,61 %	81,06 %	100	50,00 %
GA-AM	01:47:31	00:01:05	81,13 %	82,62 %	81,13 %	80,41 %	100	50,00 %
MA	01:47:32	00:01:05	81,51 %	83,36 %	81,51 %	80,89 %	100	50,00 %
MA-F	02:16:51	00:01:22	81,61 %	83,09 %	81,61 %	80,96 %	100	67,07 %
75 %								
RS	03:05:49	00:01:51	82,26 %	84,16 %	82,26 %	81,67 %	100	75,00 %
GA-WC	03:33:06	00:02:08	82,31 %	83,91 %	82,31 %	81,59 %	100	75,00 %
GA-AM	02:36:41	00:01:34	82,42 %	83,94 %	82,42 %	81,85 %	100	75,00 %
MA	02:37:43	00:01:35	82,74 %	84,33 %	82,74 %	82,14 %	100	75,00 %
MA-F	03:08:21	00:01:53	82,69 %	84,18 %	82,69 %	82,10 %	100	71,16 %
100 %								
100 %	-	00:03:15	78,60 %	81,55 %	78,60 %	77,68 %	1	100,00 %

Tabla 8.5: Resultados detallados por porcentaje inicial y algoritmo: precisión, evaluaciones y duración.

estabilidad algo inferior en ciertos escenarios.

Por otro lado, el **GA-AM** mejora respecto a enfoques más simples como **RS** y **GA-WC**, alcanzando resultados intermedios en términos de precisión y reducción de datos. Sin embargo, no logra superar la consistencia ni la eficiencia de los enfoques meméticos, mostrando una estabilidad moderada en la distribución de clases.

El **GA-WC** ofrece ligeras mejoras frente a la búsqueda aleatoria, pero mantiene una mayor dispersión de resultados y requiere porcentajes de datos más elevados para alcanzar precisiones competitivas. Esto limita su eficacia en escenarios donde la reducción de datos es prioritaria.

Finalmente, la **RS**, utilizada como referencia, queda claramente superada por todos los enfoques evolutivos y meméticos. Sus resultados muestran una alta variabilidad, una menor precisión global y una mayor propensión a generar subconjuntos con distribuciones de clases menos estables, lo que pone de manifiesto la importancia de utilizar estrategias de selección guiada para optimizar tanto el rendimiento como la estabilidad del modelo.

En conjunto, este análisis confirma que los **MA**, especialmente en su versión libre, representan la mejor alternativa para abordar la selección de subconjuntos de datos en tareas de aprendizaje profundo. Logran un equilibrio

óptimo entre precisión, eficiencia y estabilidad, superando ampliamente a las técnicas basadas en selección aleatoria o enfoques genéticos más sencillos.

8.11. Validación con el dataset PAINTING

Para comprobar la robustez de los algoritmos desarrollados, se validan los experimentos con el dataset **PAINTING**, caracterizado por una mayor complejidad visual y un número superior de clases respecto al conjunto **RPS**. Con el fin de mantener un equilibrio entre complejidad y coste computacional, se limitan los experimentos a los porcentajes iniciales del 25 % y 50 %, evaluando únicamente los algoritmos más representativos.

En particular, se selecciona el **MA**, junto con su versión libre, ya que ambos han demostrado ser los más eficaces en las pruebas anteriores. Para establecer una referencia clara, se incluyen también los resultados obtenidos usando el 100 % del conjunto de datos, y para la comparación de *accuracy* entre algoritmos también se añade el resultado del **RS**.

8.11.1. Comparación de *accuracy* entre algoritmos

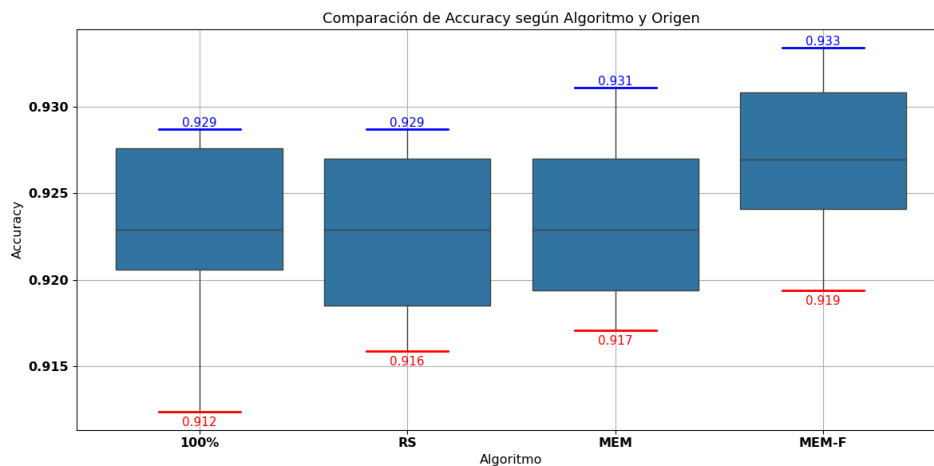


Figura 8.20: Boxplot comparando resultados con el dataset **PAINTING** usando *accuracy*.

En la Figura 8.20 se comparan los valores de *accuracy* obtenidos al aplicar distintos algoritmos de selección de subconjuntos en el conjunto de datos **PAINTING**. A simple vista, destaca el buen rendimiento de los tres enfoques meméticos frente al uso directo del 100 % del conjunto o la selección aleatoria, lo cual es especialmente significativo al tratarse de un conjunto con alta

complejidad visual y estructural.

El MA-F obtiene el mejor rendimiento general, con una mediana que ronda el **0.927** y un valor máximo de **0.933**, superando incluso la ejecución con el 100 % de los datos, cuyo máximo se sitúa en **0.929**. Esto no solo refuerza su capacidad de generalización, sino que demuestra que una selección optimizada puede superar a la totalidad del conjunto original, posiblemente por eliminar ejemplos redundantes o incluso perjudiciales.

El MA también muestra una precisión elevada, con una mediana apenas inferior al libre, y un máximo de **0.931**. Sin embargo, su varianza es ligeramente mayor, lo que sugiere que la falta de ajuste dinámico del tamaño del subconjunto puede limitar su adaptabilidad en ciertas ejecuciones.

Por otro lado, la selección aleatoria, aunque muestra valores aceptables, vuelve a confirmar su principal debilidad: la alta dispersión. Con un mínimo de **0.916** y una mediana prácticamente idéntica a la obtenida con el uso completo del dataset, su comportamiento se sitúa como una referencia básica pero poco fiable. Puede alcanzar buenos resultados, pero lo hace de forma inconsistente y sin mecanismos que garanticen estabilidad.

La ejecución con el **100 %** de los datos, utilizada como referencia absoluta, queda por debajo de las soluciones obtenidas por los algoritmos meméticos. Esto valida empíricamente que no es la cantidad de datos, sino su calidad y representatividad, lo que determina la eficacia del entrenamiento en entornos complejos.

En resumen, los algoritmos meméticos, especialmente en su versión libre, muestran el mejor rendimiento en términos de precisión y estabilidad, superando al uso del 100 % de los datos y al enfoque aleatorio.

8.11.2. Impacto del porcentaje inicial

Algoritmo	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)	Evaluaciones	Porc. Final
25 %						
MA	91,89 %	91,63 %	91,89 %	91,67 %	100	25,00 %
MA-F	92,24 %	92,06 %	92,24 %	92,09 %	100	41,35 %
50 %						
MA	92,73 %	92,61 %	92,73 %	92,63 %	100	50,00 %
MA-F	93,11 %	93,00 %	93,11 %	93,01 %	100	68,69 %
100 %						
100 %	92,24 %	92,05 %	92,24 %	92,06 %	1	100,00 %

Tabla 8.6: Resultados de los algoritmos meméticos y uso del 100 % en el dataset PAINTING.

FALTA añadir los resultados del 75 %.

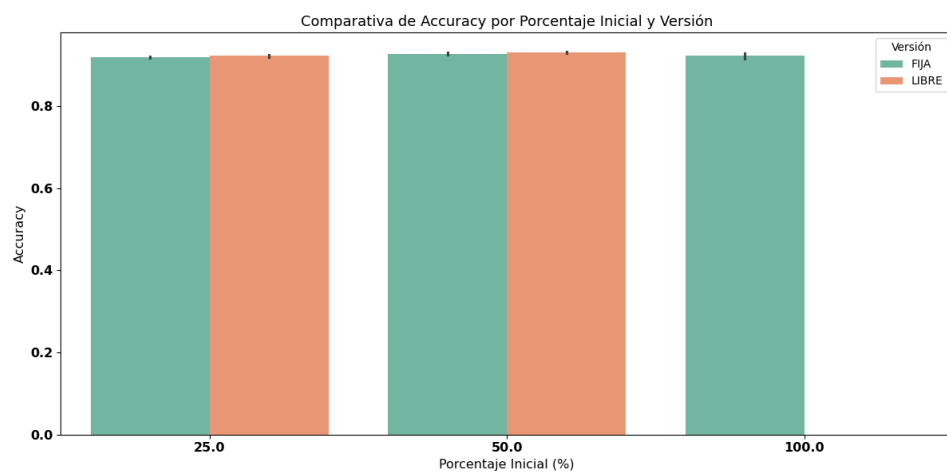


Figura 8.21: Diagrama de barras de *accuracy* según el porcentaje inicial de datos del MA, MA-F y el 100 %.

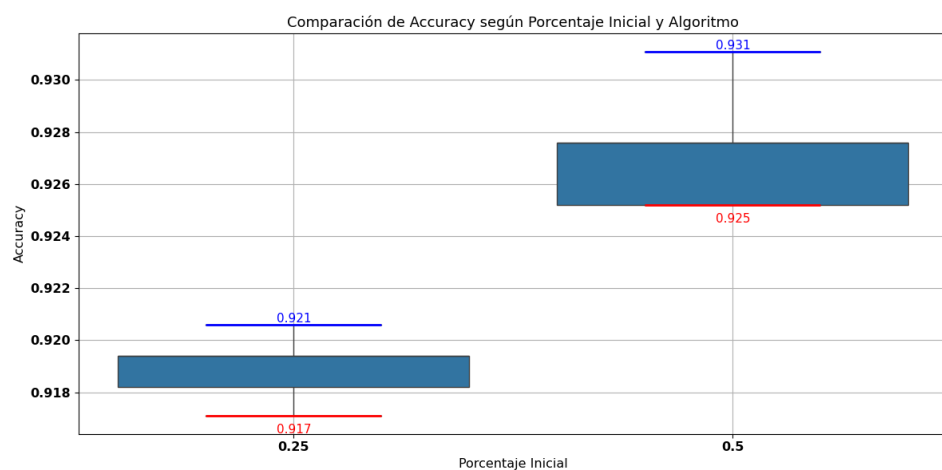


Figura 8.22: Boxplot de *accuracy* para el algoritmo memético (MA) según el porcentaje inicial de datos en el dataset PAINTING.

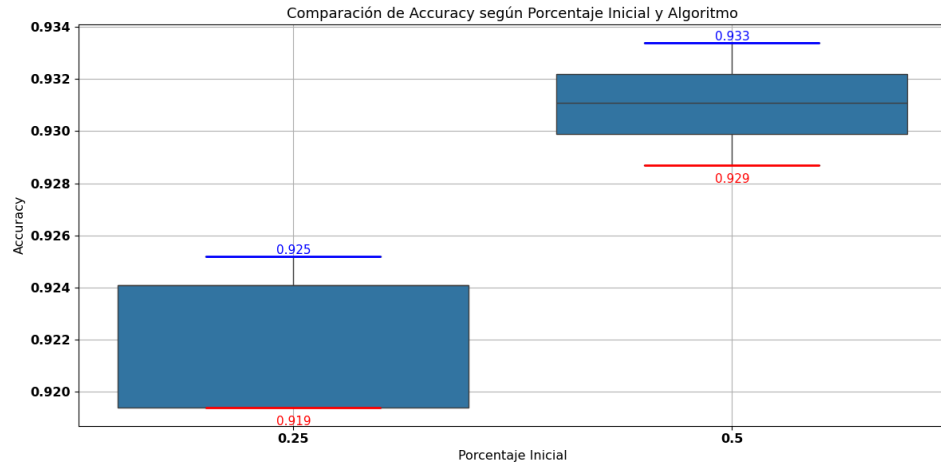


Figura 8.23: Boxplot de *accuracy* para el algoritmo memético libre (MA-F) según el porcentaje inicial de datos en el dataset PAINTING.

FALTA ajustar mejor el análisis.

La Figura 8.21 muestra cómo varía el rendimiento del modelo (medido en términos de *accuracy*) al utilizar diferentes porcentajes iniciales de datos para los algoritmos meméticos en el dataset PAINTING. Lo primero que destaca es que el uso del 50 % del conjunto original no solo logra un rendimiento equiparable, sino incluso superior al uso del 100 %. Con una mediana cercana a **0.930** y un máximo de **0.933**, este porcentaje logra un equilibrio ideal entre compresión de datos y preservación de información relevante.

Este resultado sugiere que, a partir de cierto umbral, añadir más datos no solo deja de aportar valor, sino que puede introducir ruido o redundancia. En este caso, el entrenamiento con el 100 % de los datos exhibe mayor dispersión y un mínimo significativamente más bajo (**0.912**), lo cual indica una mayor variabilidad entre ejecuciones y una menor estabilidad general.

Por otro lado, el uso del 25 % inicial también demuestra un rendimiento sorprendentemente competitivo, alcanzando un máximo de **0.925**. Sin embargo, su rango intercuartílico más estrecho (menor dispersión de los resultados) y su menor valor mínimo (**0.917**) evidencian una cierta fragilidad ante la reducción excesiva: puede funcionar bien si la selección es óptima, pero el margen de error es menor.

Este comportamiento ilustra un fenómeno interesante en la reducción de datos: no existe una relación lineal entre cantidad de datos y precisión, sino que el valor está en la calidad y representatividad del subconjunto. El resultado más robusto refuerza la hipótesis de que existe un punto de saturación a partir del cual la adición de ejemplos tiene un efecto marginal

o incluso contraproducente en modelos convolucionales.

En resumen, los resultados respaldan la viabilidad de utilizar una fracción cuidadosamente seleccionada de datos para obtener resultados comparables o incluso superiores al uso del conjunto completo.

8.11.3. Equilibrio en la distribución de clases

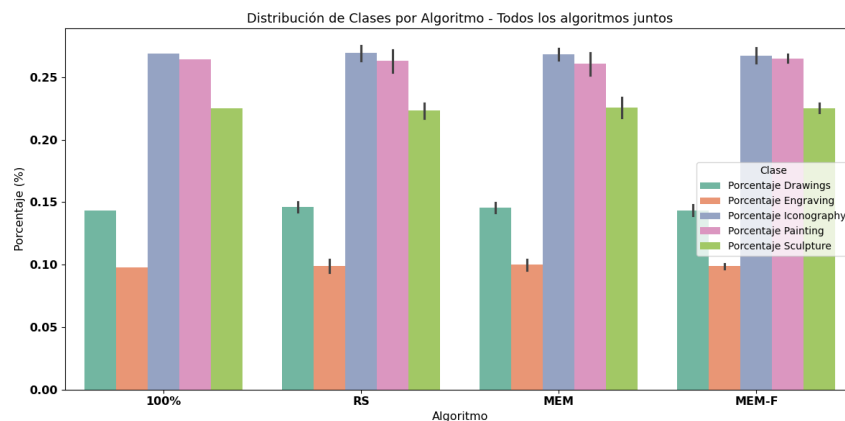


Figura 8.24: Distribución de clases seleccionadas por cada algoritmo.

La Figura 8.24 muestra la proporción de clases preservada por cada algoritmo durante el proceso de reducción. Se observa que tanto el MA como el MA-F mantienen una distribución prácticamente idéntica a la original, sin introducir sesgos estructurales.

Las clases mayoritarias (Iconography, Painting y Sculpture) se mantienen consistentes en su proporción, al igual que las clases minoritarias (Drawings y Engraving). Esta conservación sugiere que el proceso de selección no actúa de forma aleatoria ni ciega, sino que incorpora implícitamente una presión hacia la diversidad, lo cual es clave en problemas multiclase.

8.11.4. Evolución del tamaño del subconjunto

Volver a mencionar la tabla de resultados.

En la Figura 8.25 se analiza cómo evoluciona el tamaño de los subconjuntos seleccionados por los algoritmos libres respecto al valor inicial. Se observa que el MA-F tiende a incrementar el porcentaje de imágenes seleccionadas durante el proceso evolutivo.

Este comportamiento adaptativo pone de manifiesto la capacidad del algoritmo para ajustar dinámicamente la escala de la solución en función de

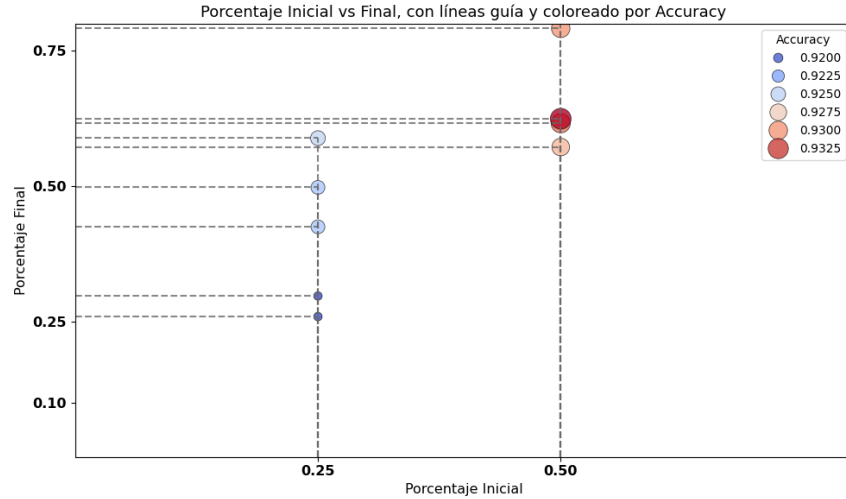


Figura 8.25: Evolución del tamaño del subconjunto seleccionado por los algoritmos libres según el porcentaje inicial.

la complejidad del problema. A diferencia de enfoques con tamaño fijo, el algoritmo libre no solo decide qué ejemplos seleccionar, sino también cuántos, ampliando el conjunto cuando detecta que puede mejorar el rendimiento sin incurrir en sobreajuste.

Además, se observa que al partir del 25 % inicial, el algoritmo tiende a aumentar el tamaño del subconjunto hasta un 16,35 % adicional, en cambio, al partir del 50 % inicial, el crecimiento es mayor, alcanzando un 18,69 % adicional. Este hecho puede sugerir una cierta prudencia evolutiva: el algoritmo no se expande indiscriminadamente, sino que responde a señales de mejora en el proceso de evaluación. Este mecanismo emergente refuerza la versatilidad del enfoque libre, que no solo busca soluciones de alta calidad, sino que lo hace optimizando también el volumen de datos utilizados.

Síntesis final de la validación con PAINTING

Los resultados obtenidos con el dataset **PAINTING** confirman la solidez y capacidad de generalización de los algoritmos desarrollados. Tanto el enfoque memético estándar como, especialmente, su versión libre, lograron mantener un rendimiento competitivo en un entorno más complejo y con mayor número de clases.

Se evidencia que es posible superar el rendimiento del conjunto completo de entrenamiento utilizando únicamente un subconjunto bien seleccionado, reduciendo significativamente el volumen de datos sin comprometer la pre-

cisión. Además, los algoritmos meméticos demuestra preservar el equilibrio entre clases y adaptarse dinámicamente al tamaño del subconjunto, lo que constituye una ventaja clave en tareas reales donde no siempre se dispone de datasets perfectamente balanceados ni completos.

En conjunto, esta validación externa no solo refuerza las conclusiones obtenidas con RPS, sino que demuestra que el uso de técnicas evolutivas, y en particular las variantes libres, puede ser una alternativa eficaz, escalable y controlable frente a la selección masiva o aleatoria de datos en procesos de entrenamiento profundo.

Capítulo 9

Conclusiones

Este Trabajo de Fin de Grado aborda el problema de la reducción de conjuntos de datos en el contexto del aprendizaje profundo, proponiendo el uso de algoritmos meméticos como técnica de selección de instancias. A lo largo del desarrollo se diseñan, implementan y evalúan diferentes enfoques evolutivos (genéticos y meméticos), con el objetivo de identificar subconjuntos representativos de imágenes que permitan entrenar modelos convolucionales de forma eficiente y sin pérdida significativa de rendimiento.

El estudio se centra en evaluar cómo diferentes variantes algorítmicas impactan sobre la precisión, la estabilidad y el tamaño final del conjunto de entrenamiento. Para ello, se utilizan dos modelos de referencia ampliamente extendidos (ResNet50 y MobileNetV2) y dos conjuntos de datos con distintos niveles de complejidad: Rock, Paper, Scissors y PAINTING.

Los resultados experimentales permiten extraer varias conclusiones clave:

- El uso de algoritmos evolutivos mejora notablemente la calidad de los subconjuntos generados en comparación con estrategias aleatorias o deterministas. Incluso las versiones básicas, como el algoritmo genético clásico, superan con claridad al enfoque aleatorio.
- Las mejoras introducidas en los operadores de cruce y mutación (ponderación basada en fitness y mutación adaptativa) incrementan la estabilidad y precisión de los resultados, especialmente en escenarios con porcentajes bajos de datos iniciales.
- La inclusión del reinicio poblacional ofrece cierto beneficio en términos de diversidad, aunque no aporta mejoras significativas respecto a las versiones mejoradas del genético en cuanto a precisión media.
- Las versiones *libres* de los algoritmos, donde el tamaño del subconjunto se ajusta dinámicamente, permiten alcanzar un equilibrio más eficaz

entre precisión y tamaño final. Este enfoque demuestra ser especialmente útil cuando no se conoce a priori la fracción óptima de datos para entrenar.

- El algoritmo memético, y en particular su versión libre (MA-F), alcanza los mejores resultados globales. Este enfoque logra una elevada precisión, una notable reducción del conjunto de datos y una gran estabilidad entre ejecuciones, consolidándose como la solución más eficaz del estudio.

La validación con un segundo conjunto de datos (PAINTING), más complejo y con mayor número de clases, permite comprobar la capacidad de generalización de los enfoques propuestos. El algoritmo memético libre no solo mantiene su rendimiento, sino que en algunos casos supera a la precisión obtenida con el 100 % del conjunto completo. Además, se confirma que las soluciones generadas preservan una distribución equilibrada de clases y ajustan el tamaño del subconjunto según la dificultad del problema.

Los resultados obtenidos tienen implicaciones prácticas significativas. Queda demostrado que es posible reducir de forma significativa los conjuntos de datos para entrenar modelos convolucionales sin comprometer el rendimiento, siempre que se utilicen estrategias de selección guiadas y adaptativas. Esto tiene implicaciones relevantes en escenarios donde los recursos computacionales son limitados o donde la auditoría y trazabilidad de los datos es una exigencia normativa.

Como líneas futuras, se plantea:

- Aplicar las técnicas propuestas en tareas más complejas, como segmentación semántica o clasificación multietiqueta.
- Investigar la incorporación de criterios adicionales en la función de fitness, como el tiempo de entrenamiento o el consumo energético.
- Explorar la integración de los algoritmos meméticos con otras técnicas de reducción de datos como destilación de conocimiento o poda de redes.

En resumen, los hallazgos de este estudio ponen de manifiesto el potencial de los algoritmos meméticos en la reducción eficiente de conjuntos de datos para aprendizaje profundo. Este trabajo demuestra que la combinación de técnicas evolutivas y búsqueda local, aplicada de forma flexible y adaptativa, constituye una herramienta poderosa para optimizar el entrenamiento de modelos de aprendizaje profundo. Más allá de la mejora en precisión o eficiencia, esta línea de investigación apunta hacia un aprendizaje más sostenible, trazable y accesible, en línea con las necesidades actuales de la inteligencia artificial moderna.

Desde un punto de vista personal, este proyecto ha representado un desafío enriquecedor tanto a nivel técnico como organizativo. La implementación completa de los algoritmos, la gestión eficiente del entorno experimental y la interpretación rigurosa de los resultados me han permitido consolidar conocimientos clave en aprendizaje automático y optimización. Asimismo, la dificultad de compaginar este trabajo con responsabilidades laborales ha reforzado mi compromiso con el rigor y la constancia, valores que considero fundamentales en mi futuro profesional.

Capítulo 10

Bibliografía

- [1] *Scrum Guide*, <https://scrumguides.org/scrum-guide.html>. visitado 25 de feb. de 2025.
- [2] *Notion - Gestión de Tareas*, <https://www.notion.com/es-es/help/guides/personal-work-dashboard>. visitado 27 de feb. de 2025.
- [3] *Enhancing collaboration in project-based organizations with it*, <https://www.pmi.org/learning/library/enhancing-collaboration-project-based-organizations-7141>. visitado 11 de mayo de 2025.
- [4] S. Chacon y B. Straub, *Pro Git*, Second. New York: Apress, 2014, <https://git-scm.com/book/en/v2>, ISBN: 978-1-4842-0076-6. DOI: 10.1007/978-1-4842-0076-6.
- [5] *Salario para Data Scientist en España - Salario Medio*, <https://es.talent.com/salary>. visitado 24 de feb. de 2025.
- [6] *Overview Google Cloud*, <https://cloud.google.com/docs/overview?hl=es-419>. visitado 25 de feb. de 2025.
- [7] *What is cloud run / cloud run documentation*, <https://cloud.google.com/run/docs/overview/what-is-cloud-run>. visitado 25 de feb. de 2025.
- [8] S. Weidman, *Deep Learning from Scratch: Building with Python from First Principles*. O'Reilly Media, Incorporated, 2019, <https://books.google.es/books?id=PRSCwwEACAAJ>, ISBN: 978-1-4920-4141-2.
- [9] P. H. Sydenham y R. Thorn, eds., *Handbook of measuring system design*. Chichester: Wiley, 2005, ISBN: 978-0-470-02143-9.
- [10] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci y M. Parmar, “A review of convolutional neural networks in computer vision,” *Artificial Intelligence Review*, vol. 57, n.º 4, pág. 99, mar. de 2024, <https://doi.org/10.1007/s10462-024-10721-6>, ISSN: 1573-7462. DOI: 10.1007/s10462-024-10721-6.

- [11] *Resnet50*, https://pytorch.org/hub/nvidia_deeplearningexamples_resnet50/. visitado 24 de feb. de 2025.
- [12] K. He, X. Zhang, S. Ren y J. Sun, “Deep Residual Learning for Image Recognition,” en *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, págs. 770-778. DOI: 10.1109/CVPR.2016.90.
- [13] A. Alnuaim, M. Zakariah, W. A. Hatamleh, H. Tarazi, V. Tripathi y E. T. Amoatey, “Human-computer interaction with hand gesture recognition using resnet and mobilenet,” *Computational Intelligence and Neuroscience*, vol. 2022, n.º 1, pág. 8 777 355, 2022, <https://onlinelibrary.wiley.com/doi/abs/10.1155/2022/8777355>, ISSN: 1687-5273. DOI: 10.1155/2022/8777355.
- [14] A. G. Howard et al., *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*, <http://arxiv.org/abs/1704.04861>, abr. de 2017. DOI: 10.48550/arXiv.1704.04861. arXiv: 1704.04861 [cs].
- [15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov y L.-C. Chen, *MobileNetV2: Inverted Residuals and Linear Bottlenecks*, <http://arxiv.org/abs/1801.04381>, mar. de 2019. DOI: 10.48550/arXiv.1801.04381. arXiv: 1801.04381 [cs].
- [16] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*. MIT Press, 2016.
- [17] Y. LeCun, Y. Bengio y G. Hinton, “Deep learning,” *Nature*, vol. 521, n.º 7553, págs. 436-444, mayo de 2015, <https://doi.org/10.1038/nature14539>, ISSN: 1476-4687. DOI: 10.1038/nature14539.
- [18] C. Shorten y T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, vol. 6, n.º 1, pág. 60, jul. de 2019, <https://doi.org/10.1186/s40537-019-0197-0>, ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0.
- [19] J. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, 1975, <https://books.google.es/books?id=YE5RAAAAMAAJ>, ISBN: 978-0-472-08460-9.
- [20] P. Moscato, “On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts - Towards Memetic Algorithms,” *Caltech Concurrent Computation Program*, oct. de 2000.
- [21] F. Neri, C. Cotta, P. Moscato y J. Kacprzyk, eds., *Handbook of Memetic Algorithms* (Studies in Computational Intelligence). Berlin, Heidelberg: Springer, 2012, vol. 379, <http://link.springer.com/10.1007/978-3-642-23247-3>, ISBN: 978-3-642-23246-6 978-3-642-23247-3. DOI: 10.1007/978-3-642-23247-3.

- [22] J. Dong, L. Zhang, B. Hou y L. Feng, “A Memetic Algorithm for Evolving Deep Convolutional Neural Network in Image Classification,” dic. de 2020, págs. 2663-2669. DOI: 10.1109/SSCI47803.2020.9308162.
- [23] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison Wesley series in artificial intelligence). Addison-Wesley, 1989, <https://books.google.es/books?id=2IIJAAAACAAJ>, ISBN: 978-0-201-15767-3.
- [24] J. VanderPlas, *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media, Inc., nov. de 2016, <https://books.google.es/books?hl=es&lr=&id=xYmNDQAAQBAJ&oi=fnd&pg=PR2&dq=Python+Data+Science+Handbook&ots=Xs9Rj3qj-M&sig=f5K5ixzKjH7pc2Uo2IYW9jrPNI8#v=onepage&q=Python%20Data%20Science%20Handbook&f=false>, ISBN: 978-1-4919-1214-0.
- [25] N. Ketkar y J. Moolayil, “Introduction to PyTorch,” en *Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch*, https://doi.org/10.1007/978-1-4842-5364-9_2, Berkeley, CA: Apress, 2021, págs. 27-91, ISBN: 978-1-4842-5364-9. DOI: 10.1007/978-1-4842-5364-9_2.
- [26] *torch.cuda — PyTorch 2.4 documentation*, <https://pytorch.org/docs/stable/cuda.html>. visitado 8 de oct. de 2024.
- [27] O. Kramer, “Scikit-Learn,” en *Machine Learning for Evolution Strategies*, https://doi.org/10.1007/978-3-319-33383-0_5, Cham: Springer International Publishing, 2016, págs. 45-53, ISBN: 978-3-319-33383-0. DOI: 10.1007/978-3-319-33383-0_5.
- [28] *NumPy v2.0 Manual*, <https://numpy.org/doc/2.0/index.html>. visitado 24 de feb. de 2025.
- [29] *Polars — Python API reference*, <https://docs.pola.rs/api/python/stable/reference/>. visitado 24 de feb. de 2025.
- [30] *Pandas 2.2.3 documentation*, <https://pandas.pydata.org/pandas-docs/version/2.2/index.html>. visitado 8 de jun. de 2025.
- [31] *Matplotlib 3.9.3 documentation*, <https://matplotlib.org/3.9.3/index.html>. visitado 24 de feb. de 2025.
- [32] *Seaborn 0.13.2 documentation*, <https://seaborn.pydata.org/tutorial.html>. visitado 3 de mayo de 2025.
- [33] *Openpyxl 3.1.3 documentation*, <https://openpyxl.readthedocs.io/en/stable/>. visitado 3 de mayo de 2025.
- [34] *Creation of virtual environments*, <https://docs.python.org/3/library/venv.html>. visitado 24 de feb. de 2025.

- [35] *Conda Documentation*, <https://docs.conda.io/en/latest/>. visitado 24 de feb. de 2025.
- [36] *cuBLAS Deterministic Algorithms*, https://docs.nvidia.com/cuda/cublas/index.html#cublasApi_reproducibility. visitado 24 de feb. de 2025.
- [37] *Early Stopping Discussion*, <https://www.geeksforgeeks.org/early-stopping-on-validation-loss-or-on-accuracy/>, feb. de 2024. visitado 4 de mayo de 2025.
- [38] *Rock Paper Scissors Dataset*, <https://public.roboflow.com/classification/rock-paper-scissors>. visitado 24 de feb. de 2025.
- [39] *(original) art images: Drawing/painting/sculptures/engravings*, <https://www.kaggle.com/datasets/thedownhill/art-images-drawings-painting-sculpture-engraving>. visitado 24 de feb. de 2025.
- [40] *(cleaned) art images: Drawing/painting/sculptures/engravings*, <https://www.kaggle.com/datasets/moosecat/art-images-drawings-painting-sculpture-engraving>. visitado 24 de feb. de 2025.

