



UNIVERSIDAD
DE GRANADA

TRABAJO FIN DE GRADO
INGENIERÍA INFORMATICA

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** intentando 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 INFORMATICA 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 14 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 14 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, extiendo 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 desarrolle sin barreras tecnológicas, facilitando el aprendizaje y la innovación.

Índice general

1. Introducción	1
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. Importancia de la Reducción de Datos en Deep Learning	21
4.2. Selección de Instancias en Aprendizaje Profundo	22
4.3. Algoritmos Evolutivos y Meméticos para la Selección de Instancias en CNN	23
4.3.1. Componentes Avanzados y Optimización Multiobjetivo	23
4.3.2. Aplicación en Redes Convolucionales (CNN)	24
4.3.3. Selección Multi-Dominio y Cross-Task	24
4.4. Destilación de Datos (Dataset Distillation)	25
4.5. Extracción de Conocimiento y Destilación de Modelos	26
4.5.1. Compresión de Modelos como Pipeline Completo	26
4.6. Desafíos y Perspectivas Futuras	27
5. Descripción de los Algoritmos	29
5.1. Algoritmo Aleatorio	29
5.1.1. Descripción	29
5.1.2. Pseudocódigo	30
5.1.3. Aplicación en la reducción de datos	30
5.1.4. Resultados esperados	30
5.2. Algoritmo Búsqueda Local	31
5.2.1. Descripción	31
5.2.2. Pseudocódigo	31
5.2.3. Aplicación en la reducción de datos	32
5.2.4. Ventajas y limitaciones	32
5.3. Algoritmo Genético	32
5.3.1. Descripción	32
5.3.2. Componentes del Algoritmo	33

5.3.3. Pseudocódigo	34
5.3.4. Aplicación en la Reducción de Datos	34
5.3.5. Ventajas y Limitaciones	35
5.4. Algoritmo Genético con Cruce Ponderado	35
5.4.1. Descripción	35
5.4.2. Componentes del Algoritmo	36
5.4.3. Pseudocódigo	36
5.4.4. Aplicación en la Reducción de Datos	36
5.4.5. Ventajas y Limitaciones	36
5.5. Algoritmo Genético con Mutación Adaptativa	37
5.5.1. Descripción	37
5.5.2. Componentes del Algoritmo	38
5.5.3. Pseudocódigo	38
5.5.4. Aplicación en la Reducción de Datos	38
5.5.5. Ventajas y Limitaciones	39
5.6. Algoritmo Genético con Reinicio Poblacional	39
5.6.1. Descripción	39
5.6.2. Componentes del Algoritmo	40
5.6.3. Pseudocódigo	40
5.6.4. Aplicación en la Reducción de Datos	41
5.6.5. Ventajas y Limitaciones	41
5.7. Algoritmo Memético	42
5.7.1. Descripción	42
5.7.2. Pseudocódigo	42
5.7.3. Aplicación en la reducción de datos	42
5.7.4. Ventajas y limitaciones	42
5.8. Versiones Libres de los Algoritmos Evolutivos	44
5.8.1. Algoritmo Genético con Cruce Ponderado (Versión Libre)	44
5.8.2. Algoritmo Genético con Reinicio Poblacional (Versión Libre)	44

5.8.3. Algoritmo Memético (Versión Libre)	45
6. Implementación	47
6.1. Descripción del Sistema	47
6.2. Herramientas y Lenguajes de Programación	48
6.3. Gestión de Dependencias	49
6.4. Arquitectura de la Implementación	50
6.4.1. Esquema General de Funcionamiento de los Algoritmos	50
6.4.2. Configuración de Modelos Preentrenados	52
6.4.3. Módulo de Algoritmos	52
6.4.4. Núcleo de Ejecución	53
6.4.5. Módulo de Utilidades	53
6.4.6. Scripts de Ejecución en GPU	53
6.5. Consideraciones de Optimización	54
6.6. Métricas de Evaluación	55
7. Entorno Experimental	57
7.1. Datasets utilizados	57
7.1.1. Rock, Paper, Scissors (Piedra, Papel, Tijera)	57
7.1.2. PAINTING (Art Images: Drawing/Painting/Sculptures/Engravings)	60
7.1.3. CIFAR-10	62
7.1.4. Comparación entre datasets	64
7.2. Diseño de los experimentos	65
7.3. Procedimiento de Ejecución y Evaluación	65
7.3.1. Métricas de Evaluación	65
7.3.2. Evaluaciones por Ejecución	66
7.3.3. Repeticiones y Semillas	66
7.3.4. Tiempos de Ejecución	66
7.4. Visualización de resultados	67
7.5. Nomenclatura simplificada de los algoritmos	68
7.6. Estructura de las Tablas de Resultados	69

7.7.	Parámetros de los Algoritmos y del Entrenamiento	70
7.7.1.	Parámetros Generales de los Algoritmos	71
7.7.2.	Parámetros del Entrenamiento de Modelos	71
7.7.3.	Parámetros Específicos por Algoritmo	71
8.	Resultados y Análisis	73
8.1.	Resultados con el conjunto completo	73
8.2.	Comparativa inicial de modelos	74
8.3.	Resultados de la búsqueda local	76
8.4.	Resultados del Algoritmo Genético	78
8.5.	Mejorando el operador de cruce	79
8.6.	Mejorando el operador de mutación	80
8.7.	Resultados del reinicio poblacional	81
8.8.	Resultados con versiones libres	83
8.9.	Resultados del algoritmo memético	89
8.10.	Resultados finales entre enfoques	91
8.10.1.	Análisis comparativo de accuracy	91
8.10.2.	Porcentaje final de datos seleccionados	92
8.10.3.	Estabilidad en la distribución de clases	94
8.10.4.	Síntesis final	95
8.11.	Validación con el dataset PAINTING	98
8.11.1.	Comparación entre algoritmos	98
8.11.2.	Evolución del tamaño del subconjunto	99
8.11.3.	Impacto del porcentaje inicial	99
8.11.4.	Distribución y equilibrio de clases	101
8.11.5.	Síntesis final	101
8.12.	Validación con el dataset CIFAR10	102
8.12.1.	Evaluación general del rendimiento	102
8.12.2.	Comparativa con las versiones libres	103
8.12.3.	Consistencia frente al porcentaje inicial	104
8.12.4.	Distribución de clases	105

8.12.5. Síntesis final de la validación con CIFAR10	106
9. Conclusiones	107
10. Bibliografía	111

Capítulo 1

Introducción

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.

En consecuencia, surge la necesidad de **reducir los conjuntos de datos de entrenamiento** para 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 **restrictiones 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.

Gracias a estas características, 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.

En este contexto, el presente 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**.

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**¹ como repositorio central, facilitando así la colaboración con el tutor y permitiendo una trazabilidad clara de todas las modificaciones.

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 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.

¹El código fuente completo y los recursos del proyecto están disponibles en el repositorio: <https://github.com/JoseRuizLopez/TFG>.

2.1.4. Seguimiento del Progreso

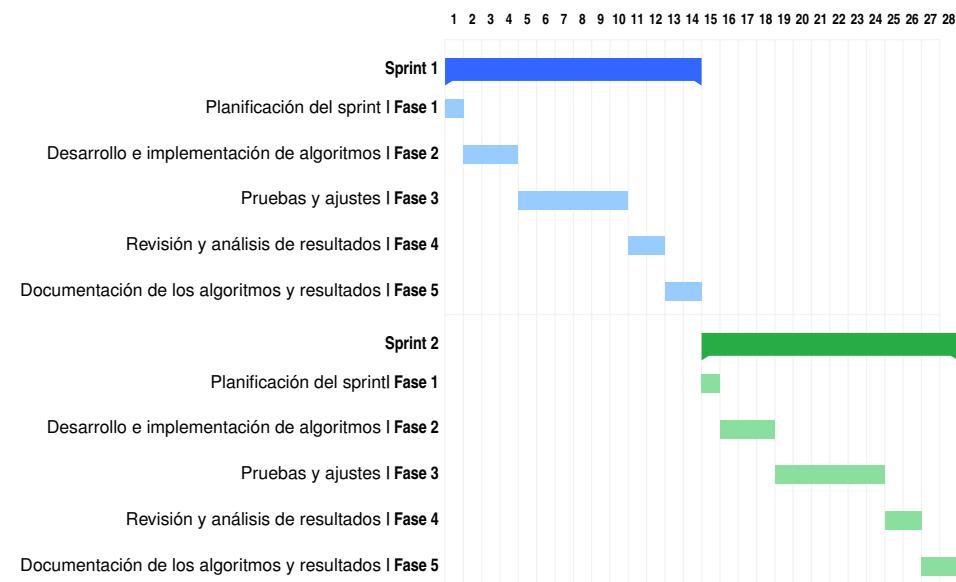


Figura 2.1: Diagrama de Gantt que muestra la planificación de dos *sprints* del proyecto, con sus respectivas fases: planificación, desarrollo, pruebas, revisión y documentación. Este esquema representa la estructura iterativa adoptada a lo largo del desarrollo

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 preprocessado:** 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	10	20	200
Repaso bibliográfico	34	20	680
Comprensión del problema y formulación	25	20	500
Selección de datasets y preprocesado inicial	18	20	360
Implementación inicial de algoritmos	43	20	860
Desarrollo de propuestas y versiones mejoradas	50	20	1.000
Evaluación experimental y análisis de resultados	62	20	1.240
Generación de gráficos	30	20	600
Escritura de la memoria del TFG	80	20	1.600
Revisión y corrección final del documento	48	20	960
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:

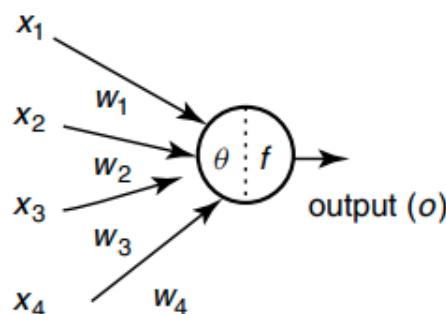


Figura 3.1

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:

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

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:

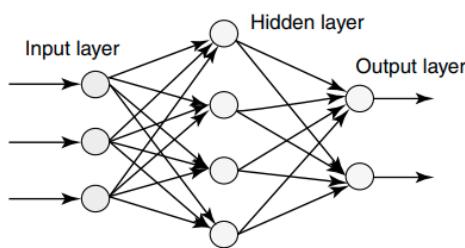


Figura 3.2

- **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 biases 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

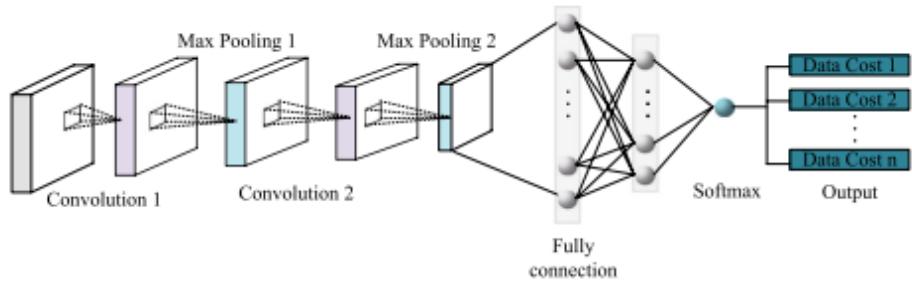


Figura 3.3: En esta imagen extraída de [10] puede observarse la estructura de una red neuronal convolucional. Junto a sus capas convolucionales.

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 además de representar avances significativos en la eficiencia y efectividad del aprendizaje profundo, son los modelos utilizados en este proyecto.

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 competiciones 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

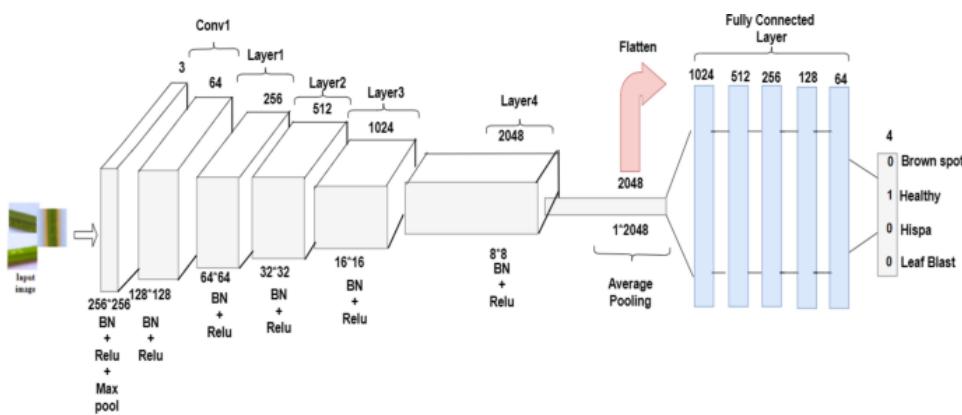


Figura 3.4: Diagrama de la Arquitectura de ResNet50

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

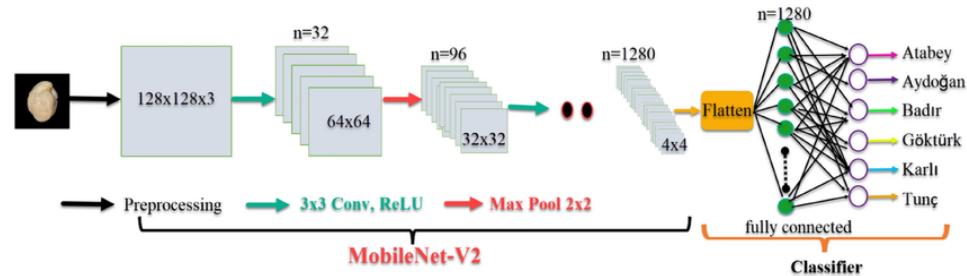


Figura 3.5: Diagrama de la Arquitectura de MobileNetV2

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

Este capítulo presenta un panorama integral de las técnicas y avances recientes en la reducción del volumen de datos necesarios para entrenar modelos de aprendizaje profundo, con especial énfasis en la selección de instancias asistida por algoritmos evolutivos y meméticos, así como la destilación de datos. La síntesis combina la revisión que aparece en el Trabajo Fin de Grado (TFG) —centrada en la optimización memética de conjuntos de imágenes para CNN— con el marco conceptual, motivación y tendencias identificados en el documento «Contexto general y motivación».

4.1. Importancia de la Reducción de Datos en Deep Learning

El auge del «big data» y la explosión de arquitecturas masivas de deep learning han disparado los requisitos de computación, energía y huella de carbono [<https://arxiv.org/abs/2104.10350>]. Estudios de life-cycle assessment estiman que entrenar un modelo transformer a gran escala puede emitir tantas toneladas de

$$CO_2$$

como cinco coches a lo largo de su vida útil [<https://aclanthology.org/P19-1355/>]. En este contexto, la filosofía data-centric AI propone desplazar el foco de la complejidad del modelo hacia la calidad del dato, argumentando que la mejora marginal al añadir más parámetros se agota rápidamente si el dataset contiene ruido o redundancia [<https://arxiv.org/abs/2303.10158>].

En lugar de seguir acumulando datos de forma indiscriminada, cada vez más estudios abogan por estrategias que prioricen la representatividad y relevancia de las instancias. Sorscher et al. (2022) matizan las tradicionales leyes de escalado —según las cuales la pérdida decrece como una potencia

del tamaño del conjunto— al demostrar que es posible “podar” hasta un 28 % del entrenamiento sin degradar la precisión de modelos como ResNet-50 [**sorscher2022beyond**]. Este hallazgo desafía la idea de que «más datos» es siempre mejor y refuerza la utilidad de identificar subconjuntos óptimos que reduzcan el coste computacional, aceleren el entrenamiento y faciliten la trazabilidad de los datos. Este último aspecto cobra especial relevancia ante normativas como la AI Act europea, que exige una gobernanza estricta del dato, incluyendo el origen y balance demográfico de los ejemplos utilizados [**aiAct**]. Así, la reducción de datos no solo aporta eficiencia, sino que se convierte en una condición necesaria para garantizar transparencia, equidad y sostenibilidad en los sistemas de IA.

4.2. Selección de Instancias en Aprendizaje Profundo

La **selección de instancias** (Instance Selection, IS) consiste en identificar un subconjunto representativo del conjunto original, eliminando ejemplos redundantes, ruidosos o poco informativos. Aunque esta técnica surgió en los años 70 como un método de edición y condensación para clasificadores *lazy* —como el algoritmo *Condensed Nearest Neighbor* (CNN) de Hart [**hart1968cnn**] o el *Edited Nearest Neighbor* (ENN) de Wilson [**wilson1972enn**]—, ha cobrado nueva relevancia en el contexto del aprendizaje profundo, especialmente ante las limitaciones de memoria de GPU y TPU modernas.

Desde una perspectiva teórica, la selección de instancias puede considerarse un problema de selección de características en el espacio de datos, y es intrínsecamente NP-duro. Por ello, se recurre a funciones objetivo bi-criterio que equilibran el tamaño del subconjunto seleccionado con el error de validación obtenido. La frontera de Pareto entre ambos criterios suele guiar la búsqueda de soluciones óptimas [**derrac2010evolutionary**].

En la práctica, IS se ha adaptado con éxito al entrenamiento de redes neuronales profundas. Por ejemplo, Albelwi y Mahmood (2016) demuestran que filtrar imágenes redundantes acelera significativamente la convergencia de las CNN sin pérdida de precisión, e incluso puede mejorarla en entornos ruidosos [**albelwi2016redundant**]. Investigaciones más recientes confirman que una selección previa de datos puede suavizar las fronteras de decisión, mitigar el sobreajuste y reducir considerablemente el tiempo de entrenamiento. En tareas de visión por computador con datasets como ImageNet-1k, CIFAR o Camelyon16, técnicas como el grad-norm filtering o el coresset pruning han mostrado reducciones de hasta un 35 % en el tiempo de convergencia y una mayor robustez frente al label noise [**nguyen2021coreset**,

[rolnick2018scale]. Además, IS resulta especialmente útil en entornos de continual learning, ya que permite almacenar conjuntos compactos en la memoria de repetición (replay) sin incurrir en catastrophic forgetting.

4.3. Algoritmos Evolutivos y Meméticos para la Selección de Instancias en CNN

Determinar el subconjunto óptimo de instancias para entrenar un modelo de aprendizaje profundo es un problema combinatorio de tamaño 2^N , donde N es el número total de ejemplos disponibles. Esta complejidad hace que los enfoques exactos sean inviables en la práctica, dando lugar a la adopción de metodologías metaheurísticas. Entre ellas, los **algoritmos evolutivos** (AE) y especialmente los **algoritmos meméticos** (MA) se consolidan como herramientas eficaces para abordar la selección de instancias en grandes datasets.

Los AE representan subconjuntos mediante cromosomas binarios, optimizados con operadores de cruce y mutación. Su principal ventaja radica en la capacidad de realizar una búsqueda global en espacios no derivables. Sin embargo, a medida que el tamaño del conjunto crece, el coste de evaluar el fitness —que implica entrenar modelos en cada generación— se vuelve prohibitivo.

Para afrontar esta limitación, los algoritmos meméticos incorporan **estrategias de mejora local** tras cada generación evolutiva. Esto permite una convergencia más rápida hacia soluciones de alta calidad al equilibrar exploración global y explotación local. Por ejemplo, García et al. (2008) presentan un algoritmo memético que mejora la precisión de k-NN utilizando solo una fracción de los ejemplos [garcia2008memetic], mientras que Derrac et al. (2010) recogen un amplio estudio que valida la eficacia de estos métodos en tareas de reducción de datos [derrac2010evolutionary].

4.3.1. Componentes Avanzados y Optimización Multiobjetivo

Los desarrollos recientes en MAs incluyen componentes clave como:

- **Representación jerárquica:** codificar subconjuntos a nivel de clústeres en lugar de instancias individuales reduce la dimensionalidad del espacio de búsqueda [bostrom2020hierarchical].
- **Optimización multiobjetivo:** funciones que equilibran tamaño del conjunto, error de validación, *Expected Calibration Error* y *Brier score*

[karimi2022multiobjective].

- **Paralelismo masivo:** evaluación de múltiples soluciones en paralelo mediante GPUs o frameworks distribuidos como MapReduce.
- **Aprendizaje de operadores:** enfoques como *Deep Neuro-Evolution* o D-NEAT entranen redes generativas (GANs) para proponer mutaciones más efectivas [real2019regularized].

4.3.2. Aplicación en Redes Convolucionales (CNN)

La integración de algoritmos meméticos en el entrenamiento de **redes convolucionales** (CNN) ha demostrado beneficios sustanciales. En lugar de entrenar sobre datasets completos, los MAs permiten identificar subconjuntos más informativos que capturan patrones visuales clave y eliminan redundancias intra-clase (por ejemplo, imágenes casi idénticas). Esto se traduce en:

- **Reducción del tiempo de entrenamiento,** al disminuir la cantidad de datos procesados por iteración y las operaciones de E/S.
- **Menor sobreajuste,** al eliminar ejemplos ruidosos o borderline que podrían distorsionar la superficie de error.
- **Facilidad de auditoría y trazabilidad,** gracias a conjuntos más compactos y manejables.

Dong et al. (2020) muestran que un algoritmo memético puede seleccionar solo el 42 % de CIFAR-100 y alcanzar la misma precisión que el conjunto completo, ahorrando hasta 2.4x en horas de GPU y mejorando la F1-score en 1.7 puntos [dong2020memetic].

4.3.3. Selección Multi-Dominio y Cross-Task

Una línea emergente en la investigación consiste en la **selección de subconjuntos transferibles** entre tareas y dominios. Esta idea, conocida como *cross-task subset selection*, permite reutilizar subconjuntos seleccionados para tareas de aprendizaje con pocos ejemplos (few-shot learning), especialmente útiles en entornos médicos. Wang y Hu (2024) aplican esta técnica en dominios como la radiología torácica y las pruebas RT-PCR por colonoscopia, logrando hasta un 50 % de ahorro en anotaciones manuales sin comprometer el rendimiento [wang2024cross].

4.4. Destilación de Datos (Dataset Distillation)

La **destilación de datos** (Dataset Distillation, DD) es una técnica emergente que busca sintetizar un conjunto reducido y artificial de ejemplos que, al ser usados para entrenar un modelo, inducen un comportamiento similar al que se obtendría entrenando con el conjunto completo de datos reales [**wang2018dataset**]. Estos ejemplos no necesariamente existen en el mundo real, pero están optimizados para encapsular la información esencial del dataset original.

Existen dos grandes enfoques dominantes en DD:

- **Meta-aprendizaje por gradientes de segundo orden:** Introducido por Wang et al. (2018), este método optimiza directamente los píxeles de las imágenes sintéticas utilizando técnicas de unrolled optimization e hipergradientes, logrando, por ejemplo, generar tan solo 10 imágenes por clase en MNIST con una precisión del 99 % [**wang2018dataset**].
- **Emparejamiento de gradientes (*gradient matching*):** Zhao et al. (2021) proponen minimizar la distancia entre los gradientes obtenidos al entrenar con los datos reales y los sintéticos. Este enfoque escala con éxito a dominios más complejos como Imagenette o Tiny-ImageNet, logrando conjuntos con solo el 1 % del tamaño original y menos del 2 % de pérdida relativa [**zhao2021gradient**].

Estudios recientes han demostrado que es posible condensar conjuntos como CIFAR-10 a menos del 0.5 % de su tamaño original sin pérdidas significativas de rendimiento [**lei2023survey**]. Además, técnicas como la **generative distillation** combinan modelos generativos avanzados (por ejemplo, StyleGAN-XL) con la destilación de datos, generando ejemplos sintéticos adaptativos que además pueden preservar la privacidad. Este enfoque permite trabajar con datos encriptados o agregados, una ventaja notable en entornos regulados como la medicina o la seguridad [**yu2024generative**].

A pesar de su potencial, la destilación de datos presenta limitaciones importantes. Los métodos actuales muestran dificultades al aplicarse a datos de alta resolución (mayores de 512x512 píxeles) o a secuencias temporales, donde los gradientes tienden a volverse inestables. Además, los modelos destilados pueden verse afectados por cambios en la distribución de datos (distribution shift), como ocurre en condiciones meteorológicas variables. Para mitigar este problema, se están explorando técnicas de **test-time subset adaptation**, que reoptimizan los ejemplos sintéticos en línea durante la inferencia [**metz2023testtime**].

4.5. Extracción de Conocimiento y Destilación de Modelos

Tanto la selección de instancias como la destilación de datos forman parte del paradigma general de **extracción de conocimiento desde los datos**, que busca maximizar la eficiencia de los sistemas de IA sin comprometer su rendimiento. Dentro de este marco, la **destilación de modelos** —proposta por Hinton et al. (2015)— transfiere el conocimiento de un modelo complejo (teacher) a otro más ligero (student) mediante el uso de *soft targets*, es decir, distribuciones de probabilidad en lugar de etiquetas duras [[hinton2015distilling](#)]. Esta técnica permite reducir significativamente la complejidad del modelo manteniendo su precisión, como sucede, por ejemplo, al transferir un BERT-Large a un BERT-Tiny en tareas de clasificación de sentimientos.

De forma análoga, la **destilación de datos** puede interpretarse como una transferencia del conocimiento, no en el espacio de los parámetros, sino en el espacio de los datos. En este contexto, un conjunto sintético actúa como un contenedor comprimido del conocimiento del conjunto original. La combinación de ambas estrategias —destilar los datos y luego destilar el modelo— está emergiendo como una línea de investigación prometedora, particularmente en entornos con restricciones de computación o privacidad.

Esta sinergia resulta especialmente útil en escenarios como el **federated learning**, donde enviar datasets sintéticos destilados puede reemplazar el intercambio directo de gradientes crudos, preservando así la privacidad de los usuarios [[awasthi2023federated](#)].

4.5.1. Compresión de Modelos como Pipeline Completo

Además, la combinación de técnicas como selección de instancias, destilación de datos y destilación de modelos se está integrando con otras estrategias de compresión como el **pruning** (poda de pesos) y la **quantization** (reducción de precisión numérica). Esto da lugar a un pipeline de compresión completo y eficiente compuesto por las siguientes fases:

1. Seleccionar el subconjunto más informativo del conjunto original.
2. Sintetizar un conjunto destilado artificial optimizado para entrenamiento rápido.
3. Destilar el modelo hacia una arquitectura más ligera.
4. Aplicar técnicas de pruning y quantization para su ejecución eficiente.

Casos industriales como los de Google o Tencent demuestran que este enfoque integrado permite desplegar modelos en dispositivos con recursos limitados, alcanzando hasta 16x de reducción en el consumo energético sin pérdida apreciable de rendimiento [google2025efficient].

4.6. Desafíos y Perspectivas Futuras

A pesar de los avances, persisten retos clave que condicionan la eficacia y adopción generalizada de las técnicas de selección de instancias y destilación de datos:

- **Optimización de hiperparámetros:** ajustar población, tasas de cruce y operadores en algoritmos meméticos aún requiere validación intensiva. Se investiga el uso de *meta-optimizadores* basados en *reinforcement learning*, aunque su coste computacional inicial es elevado.
- **Escalabilidad a dominios complejos:** imágenes de alta resolución, tareas auto-supervisadas o escenarios multimodales (texto, audio, vídeo) exigen nuevas estrategias, como destilación jerárquica o dual [dong2020memetic, yu2024generative].
- **Robustez frente a ataques:** los subconjuntos o datos sintéticos pueden ser manipulados mediante *poisoning attacks*. Se proponen defensas como la selección certificada de subconjuntos con garantías teóricas [chiang2022certified].
- **Bias y equidad:** reducir datos sin control puede eliminar ejemplos de minorías. Se plantean métricas demográficas y restricciones de diversidad para evitar sesgos [buolamwini2018gender].
- **Evaluación estandarizada:** hace falta un benchmark público que mida precisión, coste energético, huella de carbono y robustez adversaria. Iniciativas como *Open-DataShrink* apuntan en esa dirección.
- **Regulación y auditabilidad:** el AI Act europeo (2025) exige declarar criterios de selección y garantizar trazabilidad y equilibrio cultural. Las técnicas revisadas ofrecen una base sólida para generar datasets auditables [aiAct].
- **Integración de técnicas:** se prevé la combinación de selección memética adaptativa, destilación de datos condicionada al dominio, *self-supervised learning* y *active learning* para construir subconjuntos refinables en línea.

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 ← 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 adecuada 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 degrada 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:

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

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 explotar 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 proyecto¹ 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.

¹El proyecto se encuentra disponible en el repositorio público de GitHub: <https://github.com/JoseRuizLopez/TFG>

- `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.

6.2. Herramientas y Lenguajes de Programación

El desarrollo del proyecto se ha llevado a cabo utilizando **Python 3.10** [16] 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** [17, 18]: 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** [16]: Para la evaluación de los modelos se utilizaron métricas estándar [19]. Su API permite una integración fluida con PyTorch y otros módulos.
- **Numpy 2.0.0** [20]: Para operaciones matemáticas y manipulación de matrices, siendo una herramienta esencial en el procesamiento de datos.
- **Polars 1.9.0** [21]: Librería para manejar DataFrames de gran tamaño, elegida por su rendimiento superior en comparación con Pandas.
- **Pandas 2.2.3** [22]: Librería utilizada como apoyo para la generación de los gráficos, debido a las incompatibilidades detectadas con polars.
- **Matplotlib 3.9.2** [23]: Librería utilizada para la generación y visualización de gráficas.
- **Seaborn 0.13.2** [24]: Librería para la generación avanzada de gráficos estadísticos.
- **Openpyxl 3.1.5** [25]: 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` [26]. 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` [27] 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 representa este flujo, destacando especialmente el ciclo interno de optimización mediante metaheurísticas.

Como se observa en el diagrama, el sistema parte de un conjunto de datos que se divide en entrenamiento y prueba. El núcleo del proceso está constituido por un ciclo de optimización en el que una metaheurística genera subconjuntos candidatos del conjunto de entrenamiento. Cada subconjunto es evaluado mediante un modelo de aprendizaje automático (*ML*), que es entrenado brevemente para obtener métricas como la *accuracy*, utilizadas como valor de fitness para guiar la búsqueda.

Este ciclo se repite hasta alcanzar un número máximo de evaluaciones o convergencia. Una vez identificado el mejor subconjunto, se entrena un modelo final con este conjunto reducido, el cual se evalúa con el conjunto de test para obtener las métricas definitivas. Finalmente, se generan los resultados y se almacenan para su análisis posterior.

De forma complementaria, el proceso se organiza en los siguientes pasos principales:

1. **Ejecución del script de paralelización:** Se lanza el script principal que permite ejecutar múltiples tareas en paralelo, facilitando la obtención de resultados para diferentes configuraciones y semillas aleatorias.

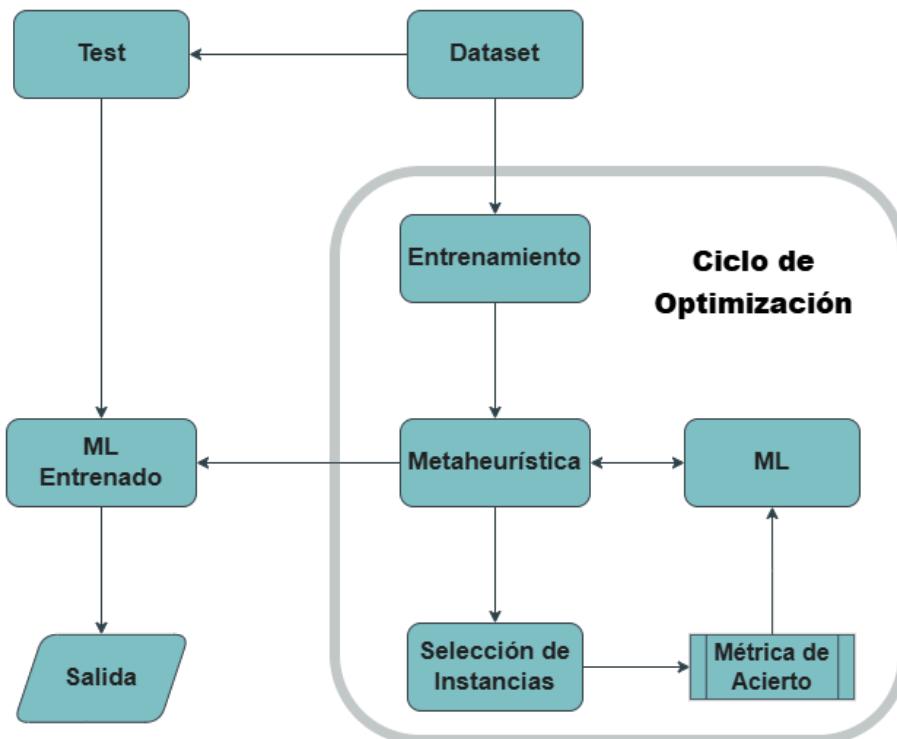


Figura 6.1: Esquema general del flujo de ejecución.

2. **Configuración inicial:** Se definen el algoritmo metaheurístico, el modelo de red neuronal y el conjunto de datos, configurados mediante los parámetros del script `generator.py`.
3. **Inicio del ciclo de optimización:** El script `main.py` organiza el flujo de ejecución e inicia el ciclo de optimización, en el que la metaheurística genera subconjuntos de entrenamiento candidatos.
4. **Evaluación de subconjuntos:** Cada subconjunto generado es evaluado mediante un modelo de *deep learning*, que se entrena brevemente para obtener métricas de rendimiento (como la *accuracy*), utilizadas como función de fitness por la metaheurística.
5. **Entrenamiento del modelo final:** Una vez seleccionado el mejor subconjunto, se entrena un modelo final usando dicho subconjunto, empleando *transfer learning* con pesos preentrenados.
6. **Evaluación final y almacenamiento de resultados:** El modelo final se evalúa con el conjunto de test, y se almacenan las métricas obtenidas, así como gráficos y tablas que permiten analizar el rendimiento del sistema.

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.

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 [28]. 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 [29].

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.

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

En el caso multiclase, se generaliza como:

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

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

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

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

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

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

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

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.

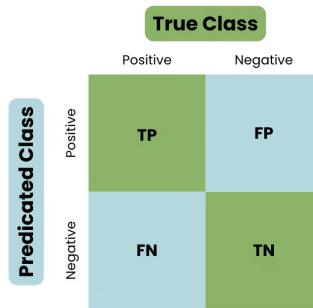


Figura 6.2: Representación gráfica de una matriz de confusión

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 , 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 [30] 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 preprocessamiento 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”.

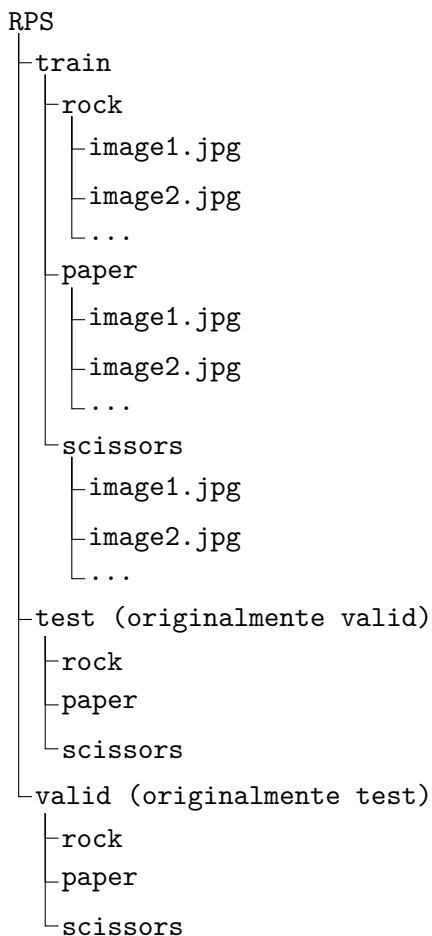


Figura 7.2: Estructura de carpetas del dataset Rock, Paper, Scissors

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.

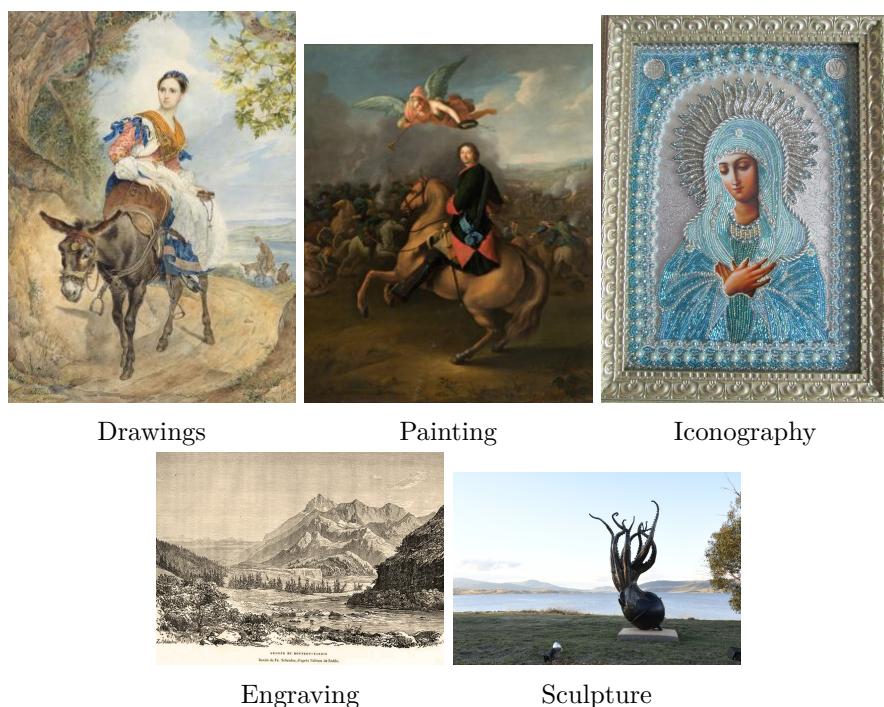


Figura 7.3: Ejemplos de clases en el dataset PAINTING

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.

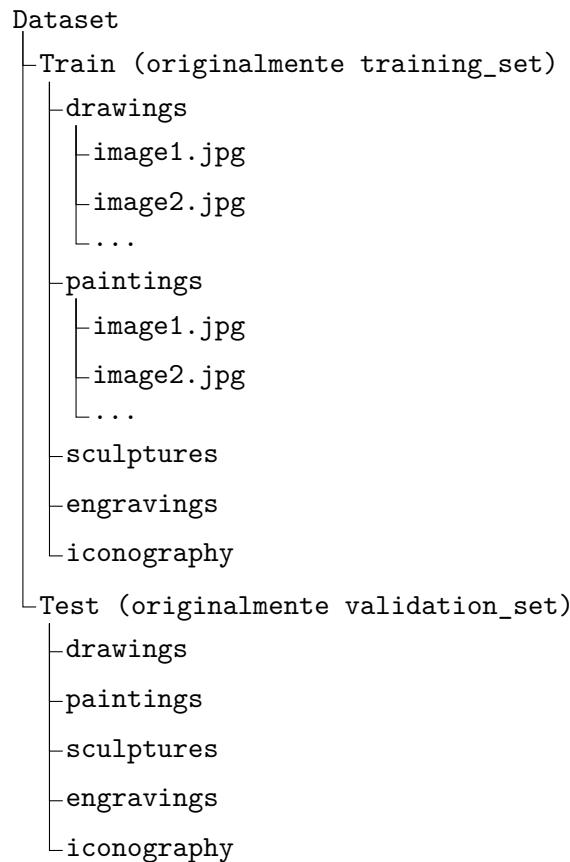


Figura 7.4: Estructura de carpetas del dataset PAINTING

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 [31]

Sin embargo, debido a la presencia de archivos innecesarios y algunas imágenes corruptas, se opta por una versión limpia disponible en Kaggle [32].

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. CIFAR-10

El **CIFAR-10** es un conjunto de datos clásico de visión por computador propuesto por la Universidad de Toronto y el Canadian Institute for Advanced Research (CIFAR) [33]. Consta de 60 000 imágenes en color, divididas en 10 clases equilibradas (*airplane, automobile, bird, cat, deer, dog, frog, horse, ship* y *truck*).

En la Figura 7.5 se han mostrado diez imágenes de cada una de las clases del dataset CIFAR-10, para que se pueda observar la variabilidad de las imágenes.

Estructura del Dataset

CIFAR-10 se distribuye en dos particiones predeterminadas: **50.000** imágenes de entrenamiento (cinco lotes de 10.000) y **10.000** imágenes de prueba (un lote). Al usar `torchvision.datasets.CIFAR10` [34], la descarga y la partición quedan gestionadas automáticamente.

Como se puede observar en la Figura 7.6, las imágenes se distribuyen en dos particiones predeterminadas: **50.000** imágenes de entrenamiento (cinco

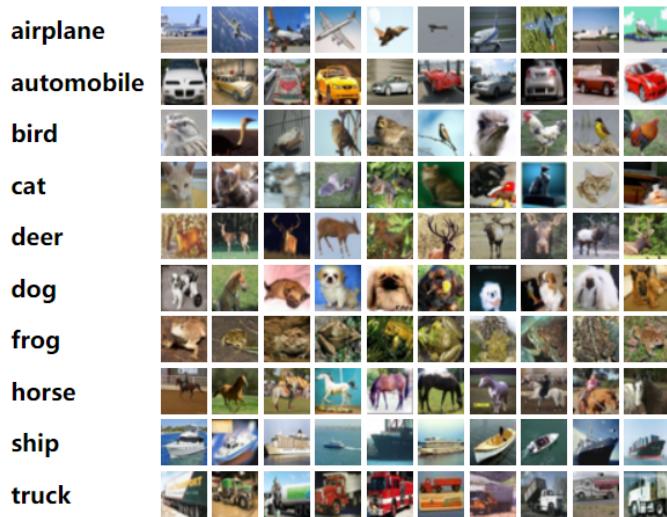


Figura 7.5: Ejemplos de cada clase en el dataset CIFAR-10

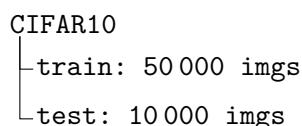


Figura 7.6: Estructura lógica del dataset CIFAR-10

lotes de 10.000) y **10.000** imágenes de prueba (un lote). Al usar `torchvision.datasets.CIFAR10`, la descarga y la partición quedan gestionadas automáticamente.

Formato de los Datos

Las imágenes se almacenan comprimidas en ficheros `.bin`. Cada imagen es RGB de 32x32 píxeles. Al integrar el dataset con `torchvision`, se aplicaron transformaciones estándar consistentes en:

- Redimensionado a 224x244 píxeles (para adaptarse a arquitecturas como MobileNet o ResNet).
- Conversión a tensor con `ToTensor()`.
- Normalización con media y desviación estándar para CIFAR-10: (0.5, 0.5, 0.5).

No se utilizaron pesos preentrenados ni transformaciones específicas dependientes del modelo, dado que CIFAR-10 no dispone de *weights* predefinidos en `torchvision`.

Uso del Dataset

CIFAR-10 se ha empleado para comprobar el comportamiento de los algoritmos en un problema de clasificación multiclase de dificultad media-alta: imágenes pequeñas, gran variabilidad intraclasa y 10 categorías. Esto permite evaluar la capacidad de generalización de los algoritmos meméticos en un escenario muy usado en la literatura.

Licencia y Acceso

El dataset CIFAR-10 está disponible públicamente y se distribuye bajo la licencia *MIT*. Se descarga automáticamente al ejecutar el código, sin necesidad de registro adicional.

7.1.4. 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
CIFAR-10	60.000	10	BIN	32×32 px

Tabla 7.1: Resumen comparativo de los datasets utilizados

La selección de estos tres datasets responde a la necesidad de evaluar los algoritmos en distintos niveles de complejidad.

El dataset **Rock, Paper, Scissors** se utiliza en las primeras fases del proyecto como punto de partida, ya que ofrece 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 coste computacional.

Posteriormente, se emplea el dataset **PAINTING** para validar el comportamiento de los algoritmos en un entorno más exigente. Al incluir cinco categorías de arte con distintos estilos visuales, este conjunto introduce una mayor variabilidad tanto semántica como estructural, lo que permite evaluar la robustez y capacidad de generalización de las soluciones desarrolladas.

Finalmente, se incorpora el dataset **CIFAR-10**, ampliamente utilizado en la literatura científica, para analizar el rendimiento de los algoritmos en un entorno estandarizado y con mayor dificultad intrínseca. Con diez

clases y una alta variabilidad visual, CIFAR-10 supone un desafío adicional tanto en términos de precisión como de capacidad de generalización, lo que permite obtener una evaluación más completa de la eficacia de los algoritmos propuestos.

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 [Aparatado 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 7.7 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.

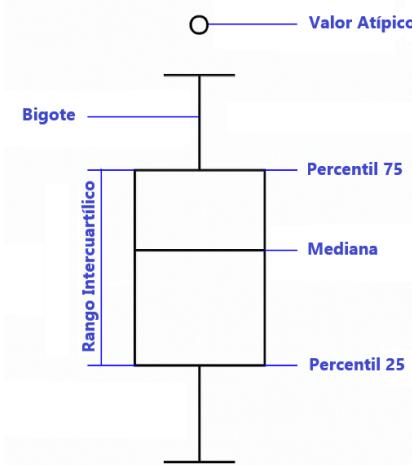


Figura 7.7: Ejemplo de boxplot explicado.

Los barplots (gráficos de barras), por su parte, se emplean para re-

presentar 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 legi-

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)	GA-AM
Genético con Mutación Adaptativa Libre	GA-AM-F
Genético con Reinicio Poblacional (Population Restart)	GA-PR
Algoritmo Memético	MA
Algoritmo Memético Libre	MA-F

Tabla 7.2: Nomenclatura simplificada de los algoritmos.

bilidad de los resultados y facilitando la interpretación de las comparativas entre métodos.

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 multiclas.

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

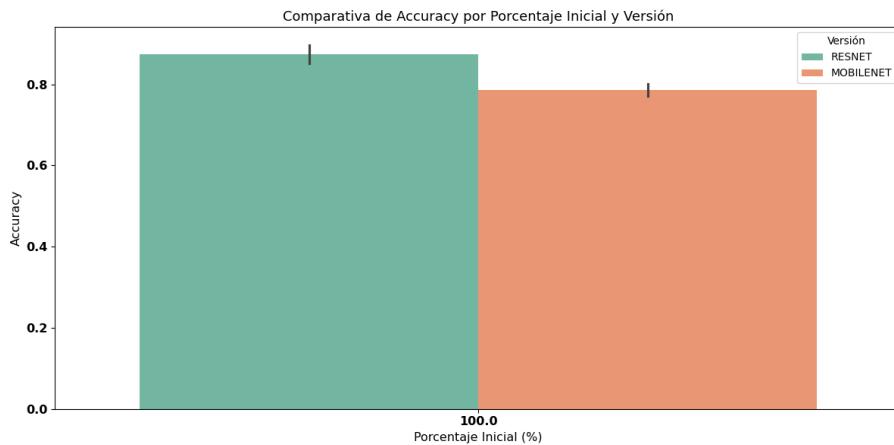


Figura 8.1: Comparación de *accuracy* de los modelos ResNet50 y MobileNet usando el 100 % del conjunto de datos.

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

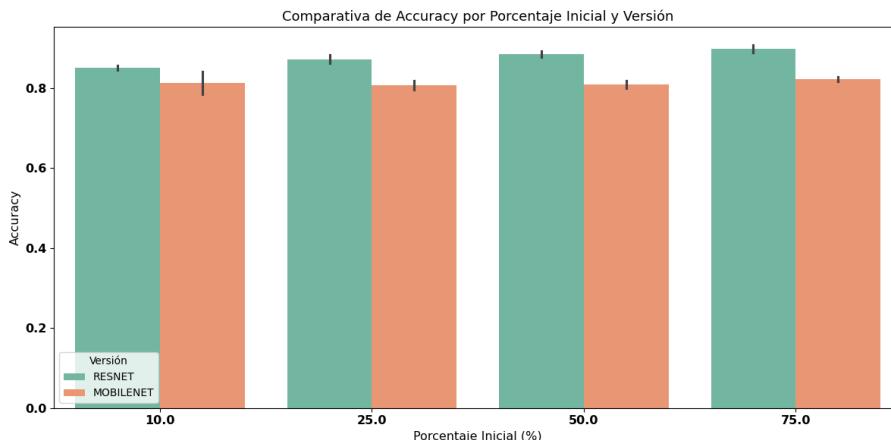


Figura 8.2: Diagrama de barras para comparar los modelos usando el *accuracy* alcanzado por cada porcentaje inicial.

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 Fi-

gura 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.

Aunque **ResNet50** alcanza la mayor precisión con el 100 % de los datos, su huella computacional (aproximadamente 25M parámetros) encarece la ejecución de los cientos de entrenamientos requeridos en los experimentos de reducción. **MobileNetV2**, con solo 3,5 M parámetros, reduce entre un 45 % y un 60 % el tiempo total de entrenamiento y permite reproducir el estudio en hardware modesto sin sacrificar más de 5-6 puntos porcentuales de precisión en los subconjuntos evaluados. Por tanto, se adopta **MobileNetV2** como modelo principal, priorizando la eficiencia sobre la ligera ventaja de **ResNet50**. El RS queda como línea base para cuantificar las mejoras aportadas por los algoritmos de reducción propuestos

Aunque **ResNet50** obtiene los mejores resultados cuando se utiliza 100 % de los datos, su elevado coste computacional hace que no sea práctico para este trabajo. Dado que los experimentos requieren entrenar el modelo cientos de veces con diferentes subconjuntos, usar una red tan pesada como **ResNet50** resultaría demasiado costoso en tiempo y recursos. En cambio, **MobileNetV2** ofrece una alternativa mucho más ligera, reduciendo notablemente los tiempos de entrenamiento y permitiendo realizar todos los experimentos incluso en equipos con menos capacidad. Además, la pérdida de precisión respecto a **ResNet50** es moderada y aceptable para los objetivos del estudio. Por ello, se elige **MobileNetV2** como modelo principal, priorizando la eficiencia. Así, el RS se utiliza como punto de referencia para evaluar el impacto de las estrategias de reducción de datos.

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.

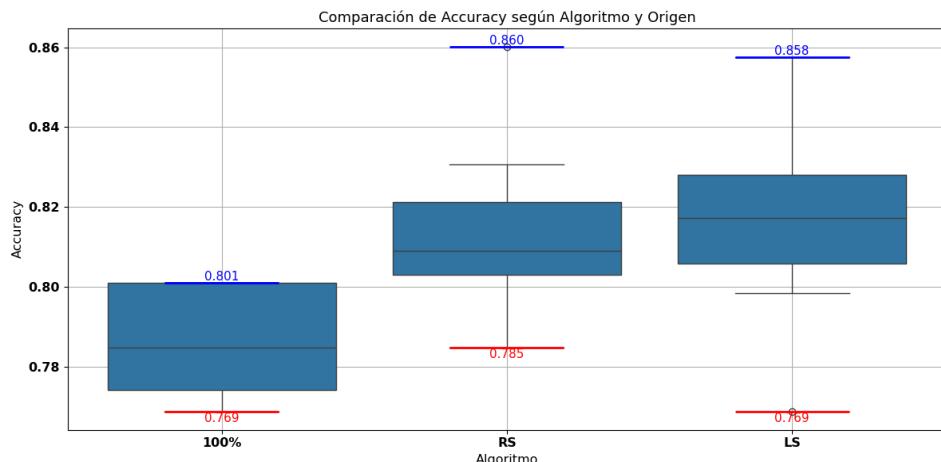


Figura 8.3: Boxplot comparando el RS con la LS usando *accuracy*.

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 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.

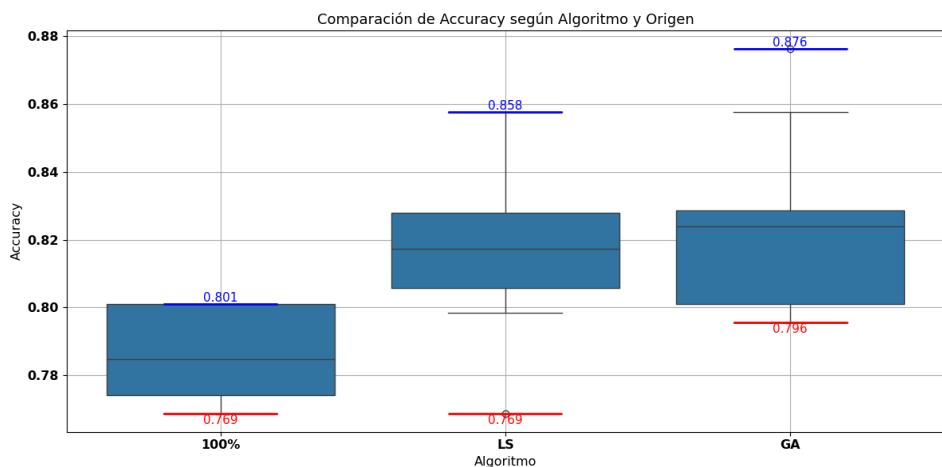


Figura 8.4: Boxplot comparando LS con GA usando *accuracy*.

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 resultados 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.

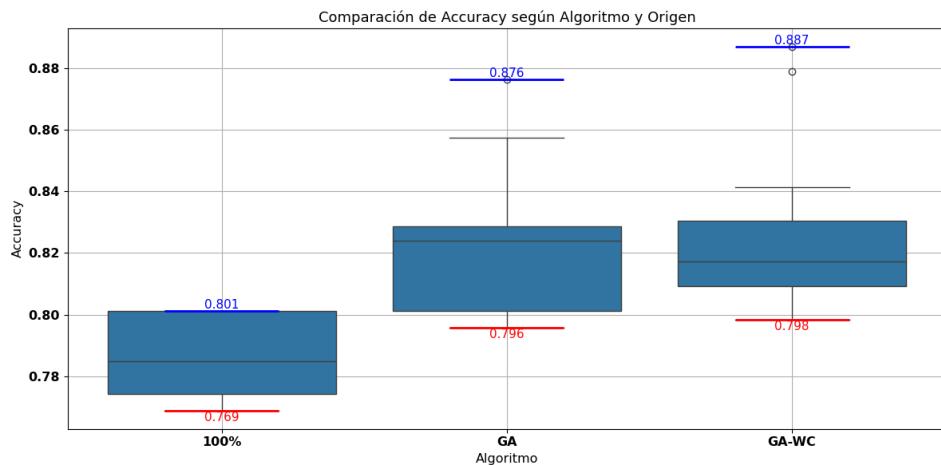


Figura 8.5: Boxplot de *accuracy* comparando el GA y el GA-WC.

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 evalua 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.

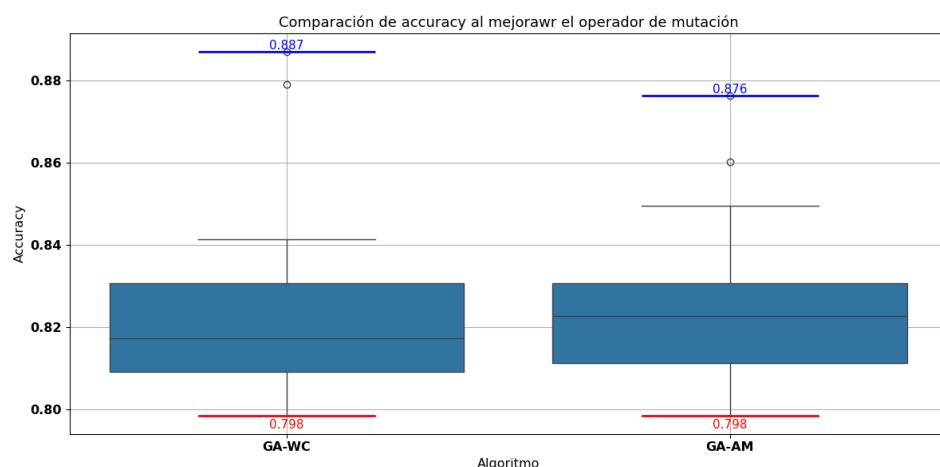


Figura 8.6: Comparación de *accuracy* entre el GA-WC y el GA-AM.

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 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

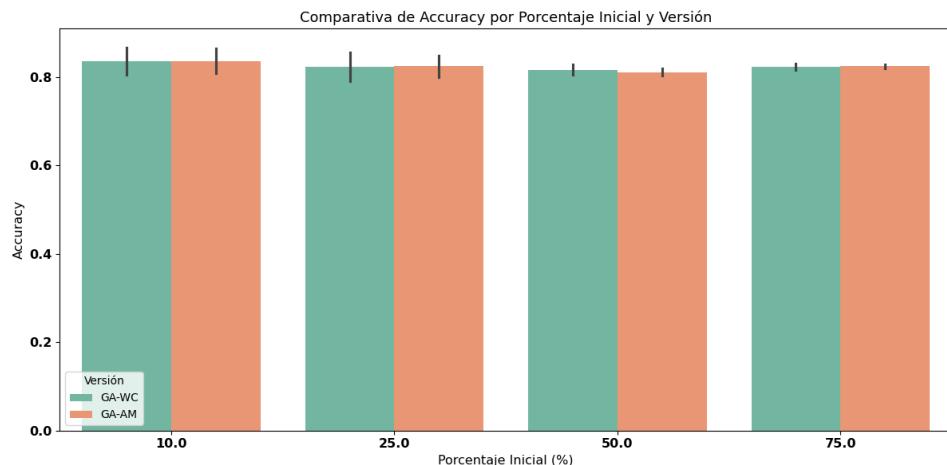


Figura 8.7: Diagrama de comparación usando el *accuracy* entre el GA-WC Y el GA-AM, separados por porcentaje inicial.

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 explotación 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.

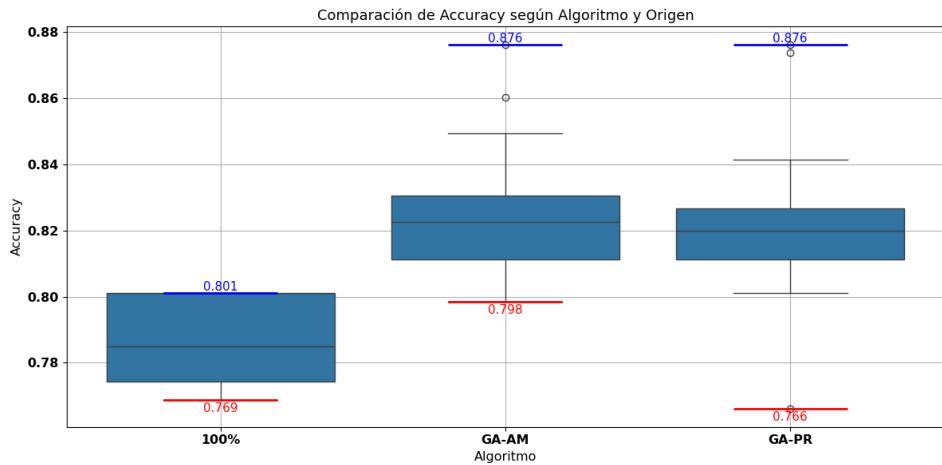


Figura 8.8: Comparación de *accuracy* entre el GA-AM y con GA-PR.

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

Tabla 8.3: Resultados de los algoritmos GA-AM y GA-PR por porcentaje inicial.

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 GA-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 GA-PR (**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 GA-PR no logra una reducción significativa en la dispersión de los valores: aunque los cuartiles y medianas de GA-AM y GA-PR son similares, se observa un leve desplazamiento hacia valores ligeramente más bajos en GA-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 GA-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 rendimiento 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.

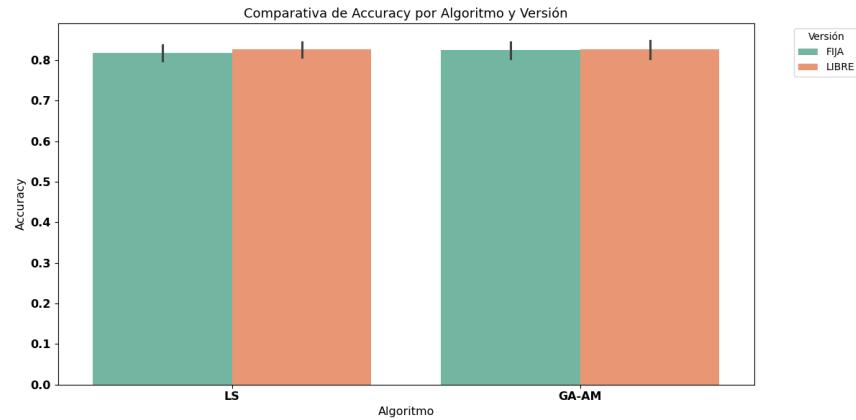


Figura 8.9: Comparación de *accuracy* entre el LS, GA-AM y sus versiones Libres (formato BARPLOT).

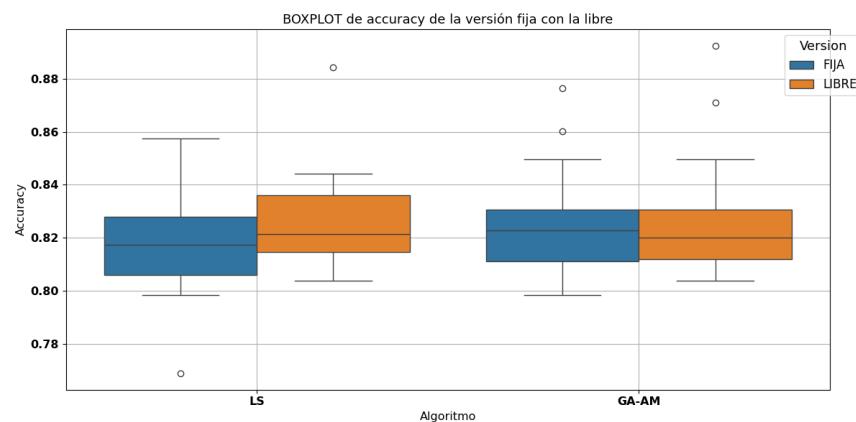


Figura 8.10: Comparación de *accuracy* entre el LS, GA-AM y sus versiones Libres (formato BOXPLOT).

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.

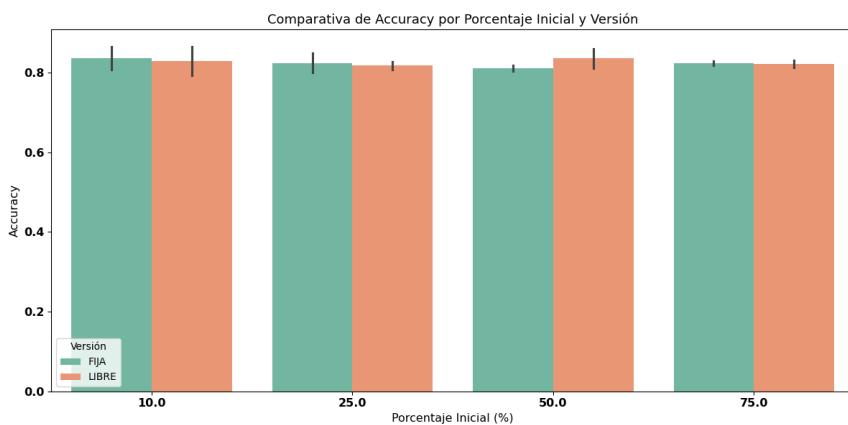


Figura 8.11: Comparación de *accuracy* en función del porcentaje inicial para el algoritmo GA-AM.

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, junto con las Tablas 8.4 y 8.5, permiten visualizar y comparar con mayor profundidad la relación entre el

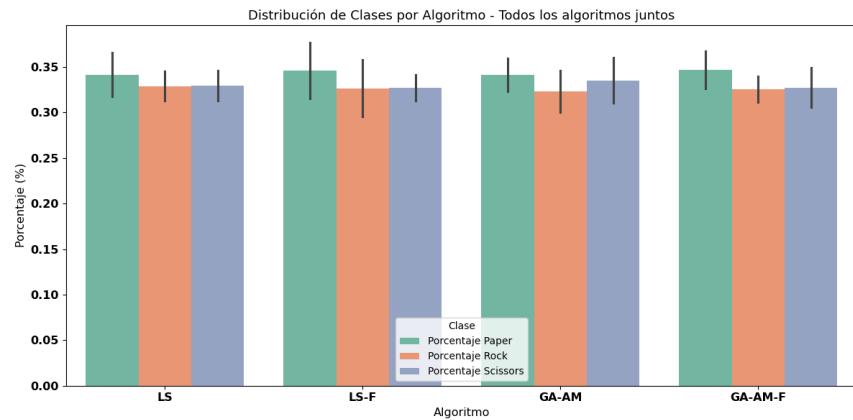


Figura 8.12: Distribución de clases en los subconjuntos generados por los algoritmos estándar y libres.

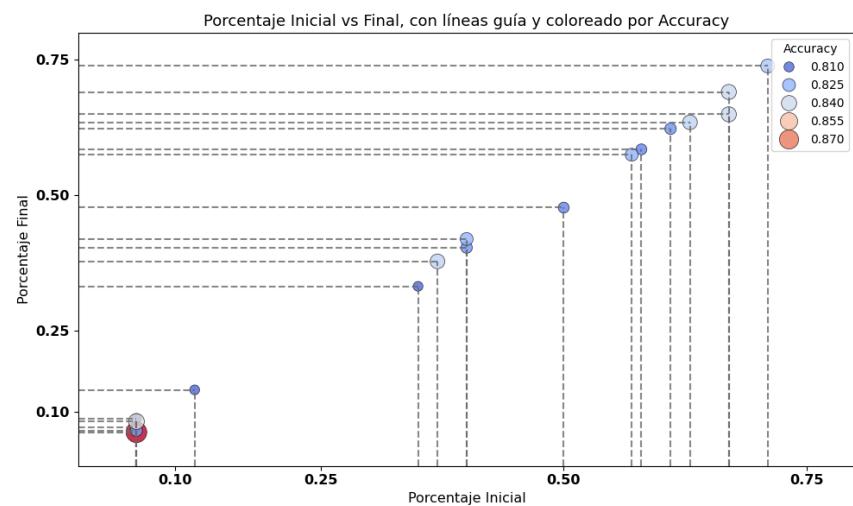


Figura 8.13: Relación entre *accuracy* y porcentaje final de datos seleccionados por el algoritmo LS-F.

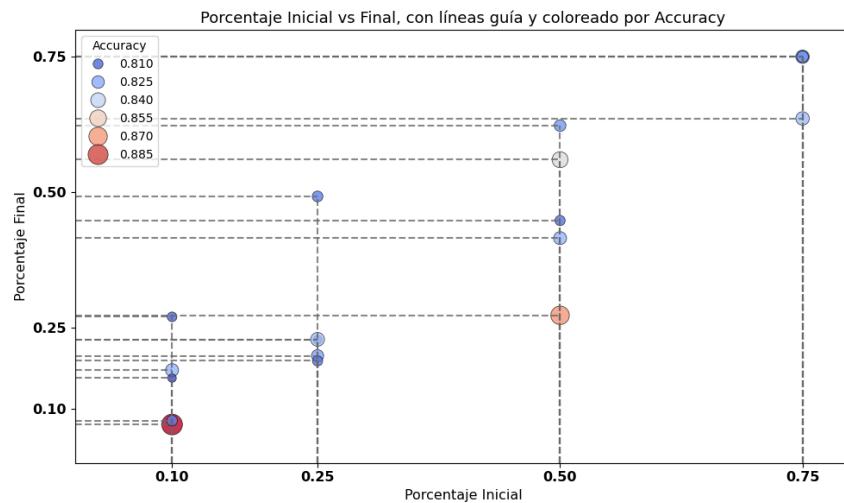


Figura 8.14: Relación entre *accuracy* y porcentaje final de datos seleccionados por el algoritmo GA-AM-F.

Algoritmo	Duración Total	Duración por Eval.	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)
10 %						
GA-AM	00:27:39	00:00:16	83.66 %	84.24 %	83.66 %	83.09 %
GA-AM-F	00:34:32	00:00:20	82.96 %	84.07 %	82.96 %	82.12 %
25 %						
GA-AM	00:57:28	00:00:34	82.42 %	83.68 %	82.42 %	81.78 %
GA-AM-F	01:07:13	00:00:40	81.77 %	82.66 %	81.77 %	81.18 %
50 %						
GA-AM	01:47:31	00:01:04	81.13 %	82.62 %	81.13 %	80.41 %
GA-AM-F	01:51:24	00:01:06	83.60 %	85.22 %	83.60 %	83.03 %
75 %						
GA-AM	02:36:40	00:01:34	82.42 %	83.94 %	82.42 %	81.85 %
GA-AM-F	02:33:25	00:01:32	82.20 %	83.94 %	82.20 %	81.47 %
100 %						
	100 %	–	78.60 %	81.55 %	78.6 %	77.68 %

Tabla 8.4: Resultados de los algoritmos GA-AM y GA-AM-F por porcentaje inicial.

Algoritmo	Porc. Inicial	Duración Total	Duración por Eval.	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)
Porcentaje inicial $\leq 10\%$							
LS	10 %	00:33	00:00	81.24 %	81.81 %	81.24 %	80.61 %
LS-F	6 %	00:38	00:00	83.23 %	83.7 %	83.23 %	82.71 %
Porcentaje inicial $\leq 25\%$							
LS	25 %	01:07	00:00	82.47 %	83.95 %	82.47 %	81.76 %
LS-F	12 %	01:00	00:00	80.91 %	82.73 %	80.91 %	79.82 %
Porcentaje inicial $\leq 50\%$							
LS	50 %	02:08	00:01	80.76 %	82.77 %	80.76 %	80.09 %
LS-F	35 %	01:36	00:00	80.91 %	83.12 %	80.91 %	79.76 %
LS-F	37 %	01:57	00:01	83.6 %	83.96 %	83.6 %	83.19 %
LS-F	40 %	02:20	00:01	82.12 %	83.3 %	82.12 %	81.34 %
LS-F	50 %	03:13	00:01	81.45 %	82.6 %	81.45 %	80.61 %
Porcentaje inicial $\leq 75\%$							
LS	75 %	03:06	00:01	82.37 %	83.97 %	82.37 %	81.67 %
LS-F	57 %	03:54	00:02	82.53 %	83.72 %	82.53 %	81.77 %
LS-F	58 %	04:02	00:02	81.45 %	82.37 %	81.45 %	80.76 %
LS-F	61 %	04:12	00:02	81.72 %	83.3 %	81.72 %	81.29 %
LS-F	63 %	02:36	00:01	83.6 %	84.43 %	83.6 %	82.77 %
LS-F	67 %	03:02	00:01	83.87 %	84.59 %	83.87 %	83.4 %
LS-F	71 %	03:18	00:01	83.06 %	84.93 %	83.06 %	82.42 %
Porcentaje inicial $\leq 100\%$							
LS-F	85 %	03:49	00:02	81.99 %	83.71 %	81.99 %	81.56 %
LS-F	87 %	04:32	00:02	82.26 %	84.22 %	82.26 %	81.76 %
100 %	100 %	–	00:03	78.6 %	81.55 %	78.6 %	77.68 %

Tabla 8.5: Resultados de los algoritmos LS y LS-F agrupados por franjas de porcentaje inicial.

accuracy, el porcentaje final de datos seleccionados y otros aspectos clave como la duración y estabilidad en las versiones libres.

En el caso de la LS-F, la Figura 8.13 muestra una tendencia a incrementar ligeramente el tamaño del subconjunto final en las ejecuciones con mayor precisión, lo que sugiere una adaptación flexible en función del rendimiento alcanzado. Esta observación se refuerza al revisar los datos de la Tabla 8.5, donde se evidencia que LS-F mantiene una precisión elevada incluso con porcentajes iniciales relativamente bajos, y logra hacerlo con duraciones contenidas por evaluación, especialmente en las franjas hasta el 50 %.

Por su parte, el GA-AM-F muestra en la Figura 8.14 una mayor dispersión en los tamaños finales seleccionados, lo que refleja su capacidad adaptativa para ajustarse a diferentes condiciones de búsqueda. Esta flexibilidad también se ve reflejada en la Tabla 8.4, donde se observa que GA-AM-F alcanza una precisión superior a la versión fija GA-AM, especialmente en porcentajes iniciales del 50 %, manteniendo una duración por evaluación razonable.

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 general, 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. Además, las tablas permiten apreciar esta versatilidad en detalle, al mostrar cómo se comportan métricas clave como *Precision*, *Recall* y *F1-score* a lo largo de las distintas franjas de porcentaje inicial.

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.

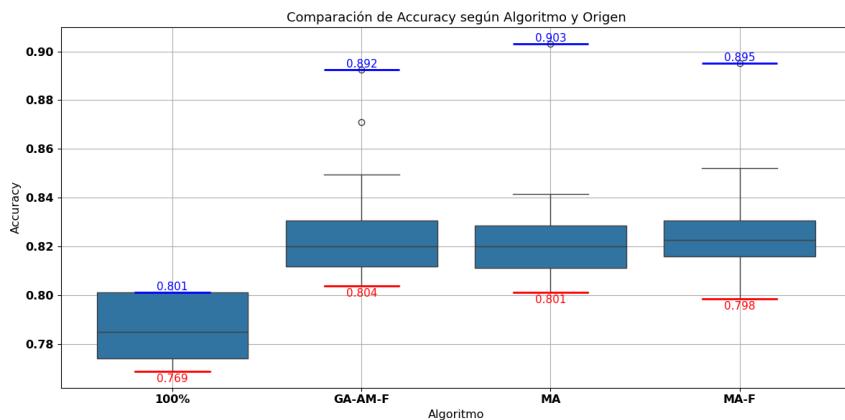


Figura 8.15: Comparación de *accuracy* entre el algoritmo memético (MA) y su versión libre.

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 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.

El análisis detallado de la Tabla 8.6 confirma estas observaciones y permite extraer conclusiones más matizadas sobre el comportamiento de los

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

Tabla 8.6: Resultados de los MA y del GA-AM-F por porcentaje inicial.

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 sobrerepresentación 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

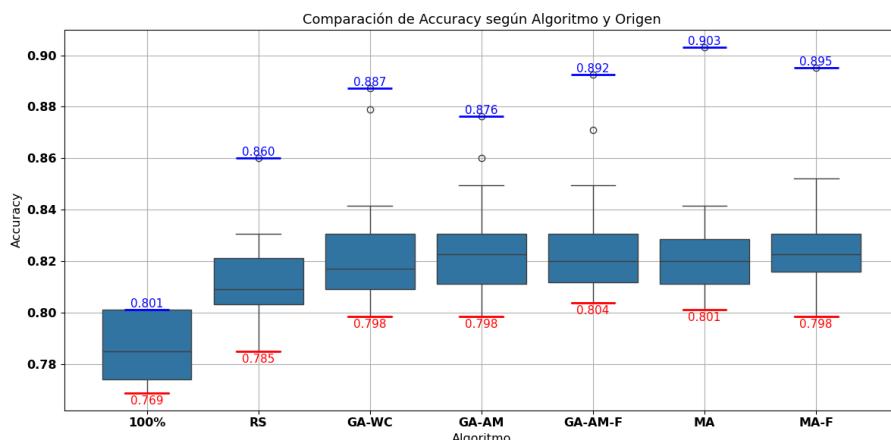


Figura 8.16: Boxplot de *accuracy* para los algoritmos RS, GA-WC, GA-AM, GA-AM-F, MA y MA-F.

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 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

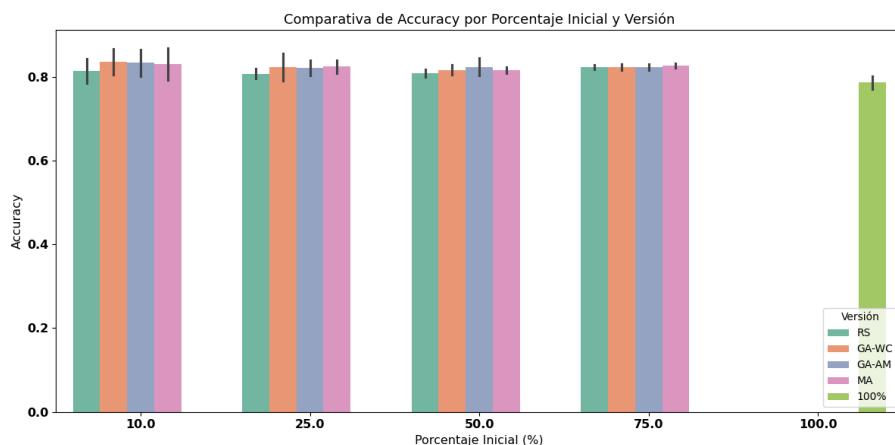


Figura 8.17: Porcentaje final de datos seleccionados por cada algoritmo.

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

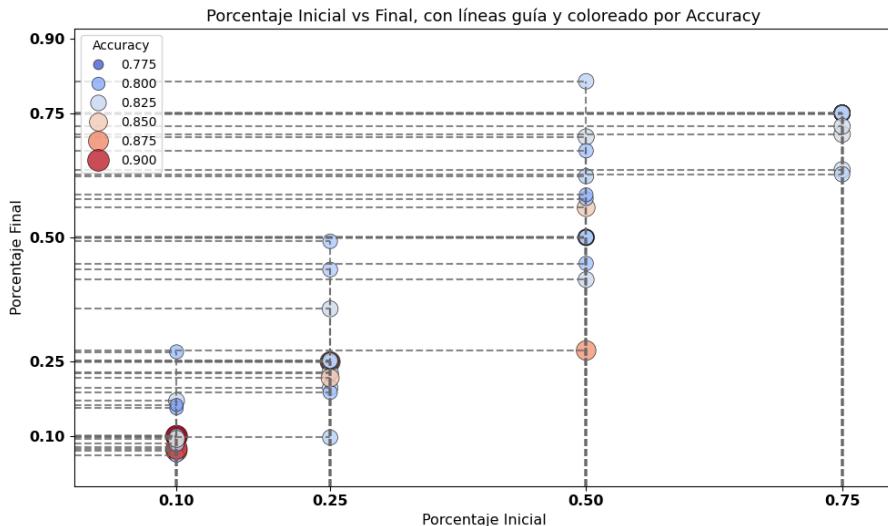


Figura 8.18: Relación entre *accuracy* y porcentaje final de datos seleccionados por cada algoritmo.

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

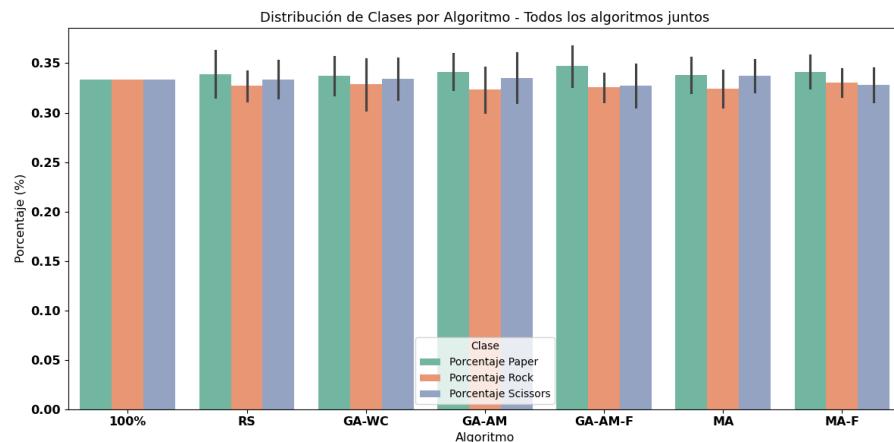


Figura 8.19: Distribución de clases seleccionadas por cada algoritmo en el conjunto reducido final.

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

meter la calidad del aprendizaje.

8.10.4. Síntesis final

Algoritmo	Duración Total	Duración por Eval.	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)	Porc. Final
10 %							
RS	00:38:17	00:00:22	81,34 %	82,46 %	81,34 %	80,52 %	10,00 %
GA-WC	00:34:03	00:00:21	83,60 %	84,38 %	83,60 %	82,82 %	10,00 %
GA-AM	00:27:39	00:00:17	83,66 %	84,24 %	83,66 %	83,09 %	10,00 %
GA-AM-F	00:34:32	00:00:21	82,96 %	84,07 %	82,96 %	82,12 %	14,94 %
MA	00:27:43	00:00:17	82,85 %	84,44 %	82,85 %	81,91 %	10,00 %
MA-F	00:32:39	00:00:20	83,28 %	84,70 %	83,28 %	82,46 %	9,62 %
25 %							
RS	01:18:22	00:00:47	80,70 %	81,90 %	80,70 %	79,90 %	25,00 %
GA-WC	01:10:21	00:00:43	82,37 %	83,86 %	82,37 %	81,57 %	25,00 %
GA-AM	00:57:29	00:00:34	82,42 %	83,68 %	82,42 %	81,78 %	25,00 %
GA-AM-F	01:07:13	00:00:40	81,77 %	82,66 %	81,77 %	81,18 %	26,68 %
MA	00:57:12	00:00:34	81,99 %	83,11 %	81,99 %	81,21 %	25,00 %
MA-F	01:08:54	00:00:41	82,80 %	83,77 %	82,80 %	82,17 %	27,18 %
50 %							
RS	02:26:40	00:01:28	80,86 %	82,65 %	80,86 %	80,22 %	50,00 %
GA-WC	02:15:43	00:01:21	81,61 %	83,78 %	81,61 %	81,06 %	50,00 %
GA-AM	01:47:31	00:01:05	81,13 %	82,62 %	81,13 %	80,41 %	50,00 %
GA-AM-F	01:51:24	00:01:07	83,60 %	85,22 %	83,60 %	83,03 %	46,36 %
MA	01:47:32	00:01:05	81,51 %	83,36 %	81,51 %	80,89 %	50,00 %
MA-F	02:16:51	00:01:22	81,61 %	83,09 %	81,61 %	80,96 %	67,07 %
75 %							
RS	03:05:49	00:01:51	82,26 %	84,16 %	82,26 %	81,67 %	75,00 %
GA-WC	03:33:06	00:02:08	82,31 %	83,91 %	82,31 %	81,59 %	75,00 %
GA-AM	02:36:41	00:01:34	82,42 %	83,94 %	82,42 %	81,85 %	75,00 %
GA-AM-F	02:33:25	00:01:32	82,20 %	83,94 %	82,20 %	81,47 %	72,72 %
MA	02:37:43	00:01:35	82,74 %	84,33 %	82,74 %	82,14 %	75,00 %
MA-F	03:08:21	00:01:53	82,69 %	84,18 %	82,69 %	82,10 %	71,16 %
100 %							
100 %	–	00:03:15	78,60 %	81,55 %	78,60 %	77,68 %	100,00 %

Tabla 8.7: Resultados detallados por porcentaje inicial.

Los resultados globales obtenidos para el conjunto de datos Rock, Paper, Scissors (véase Tabla 8.7) 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 media en múltiples porcentajes iniciales, sino que también logra mantener un porcentaje reducido de datos seleccionados, optimizando la eficiencia sin comprometer el 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 algoritmo **GA-AM-F**, incorporado como mejora del genético adaptativo, muestra una evolución destacable. Aunque en algunos escenarios no supera completamente al enfoque memético, sí logra mejores precisiones que otros algoritmos genéticos, manteniendo además un porcentaje de datos intermedio, lo cual refuerza su potencial como alternativa más estable y eficaz dentro del grupo genético.

El enfoque **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 estabilidad algo inferior en ciertos escenarios.

En cuanto a **GA-AM**, aunque mejora claramente respecto a **RS** y **GA-WC**, no alcanza los niveles de rendimiento ni estabilidad de los enfoques meméticos. El algoritmo **GA-WC**, por su parte, obtiene ligeras mejoras frente a la selección aleatoria, pero requiere un mayor porcentaje de datos para lograr precisión competitiva, y sufre mayor dispersión entre ejecuciones.

Finalmente, la referencia aleatoria **RS** queda superada por todos los enfoques evolutivos y meméticos. Sus resultados muestran una alta variabilidad, menor precisión global y una propensión a generar subconjuntos con distribuciones de clases desbalanceadas, evidenciando la importancia de emplear estrategias de selección informadas.

Además de los análisis cuantitativos, las Figuras 8.20, 8.21, 8.22 y 8.23 refuerzan visualmente estas conclusiones mediante barplots comparativos para cada porcentaje inicial. Estos gráficos permiten observar cómo varía el rendimiento de cada algoritmo manteniendo fijo el punto de partida, destacando especialmente la superioridad de los enfoques meméticos (**MA**, **MA-F**) y la mejora progresiva introducida por **GA-AM-F**.

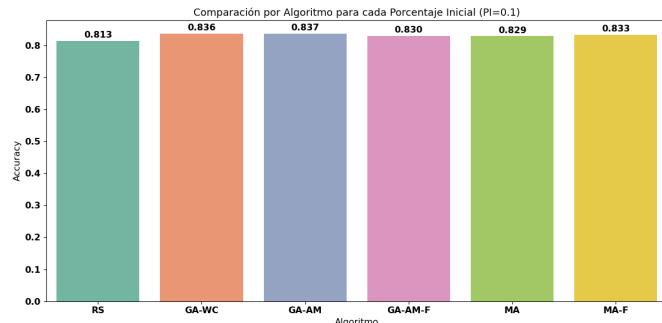


Figura 8.20: Comparativa de *accuracy* para cada algoritmo con porcentaje inicial 10 %.

En conjunto, este análisis confirma que los algoritmos meméticos, especialmente en su versión libre (**MA-F**), representan la mejor alternativa para

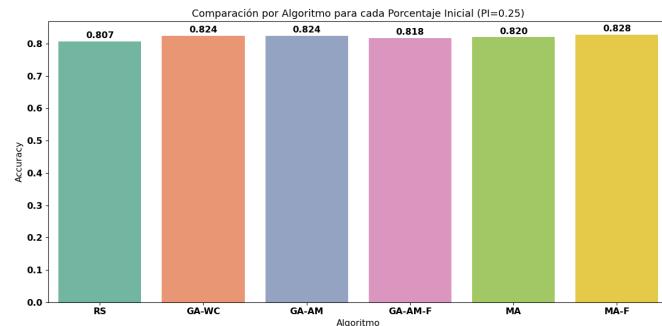


Figura 8.21: Comparativa de *accuracy* para cada algoritmo con porcentaje inicial 25 %.

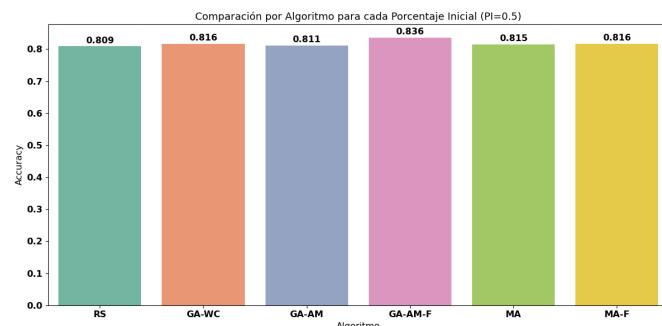


Figura 8.22: Comparativa de *accuracy* para cada algoritmo con porcentaje inicial 50 %.

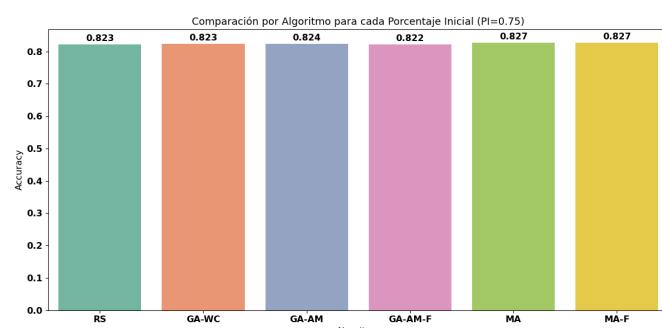


Figura 8.23: Comparativa de *accuracy* para cada algoritmo con porcentaje inicial 75 %.

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 y a los otros enfoques genéticos. Los resultados medios obtenidos a lo largo de todos los porcentajes iniciales refuerzan esta conclusión, situando a MA-F como el algoritmo más robusto en términos globales.

No obstante, al analizar los tiempos de entrenamiento, se observa que algunos algoritmos genéticos mejorados también ofrecen ventajas competitivas en contextos específicos. Aunque no existen diferencias drásticas entre algoritmos dentro de cada porcentaje inicial, la selección de instancias puede reducir significativamente el tiempo total y el de inferencia por evaluación. Esta eficiencia es especialmente relevante en escenarios donde el tiempo de respuesta o los recursos computacionales son críticos. En particular, resulta destacable que con porcentajes iniciales tan reducidos como el 10 % o 25 %, algunos algoritmos alcanzan rendimientos comparables o superiores a los obtenidos con el 100 % de los datos, y lo hacen en una fracción del tiempo. En este sentido, GA-AM-F se posiciona como la alternativa más eficiente con un 50 % de datos: logra la mayor precisión dentro de ese grupo, mantiene tiempos de ejecución moderados y presenta una menor inestabilidad respecto a otros porcentajes iniciales. Esto lo convierte en una opción especialmente atractiva cuando se prioriza la relación entre rendimiento, estabilidad y coste computacional.

8.11. Validación con el dataset PAINTING

Para validar la robustez de los algoritmos propuestos, se emplea el dataset PAINTING, caracterizado por una elevada complejidad visual y un número superior de clases respecto al conjunto RPS. Los experimentos se realizan con los porcentajes iniciales del 25 %, 50 % y 75 %, utilizando los algoritmos MA y MA-F, identificados previamente como los más efectivos. Como referencia comparativa, se incluyen también resultados con el 100 % del dataset y la RS.

8.11.1. Comparación entre algoritmos

En la Figura 8.24 se muestran los valores de *accuracy* obtenidos por cada algoritmo. El algoritmo MA-F sigue destacando, alcanzando una mediana de **0.933** cuando se parte del 50 % de los datos, superando al modelo entrenado con el 100 % del conjunto (mediana **0.924**). El algoritmo MA obtiene también resultados sólidos, especialmente con el 75 %.

La RS presenta de nuevo mayor variabilidad, y aunque alcanza un valor

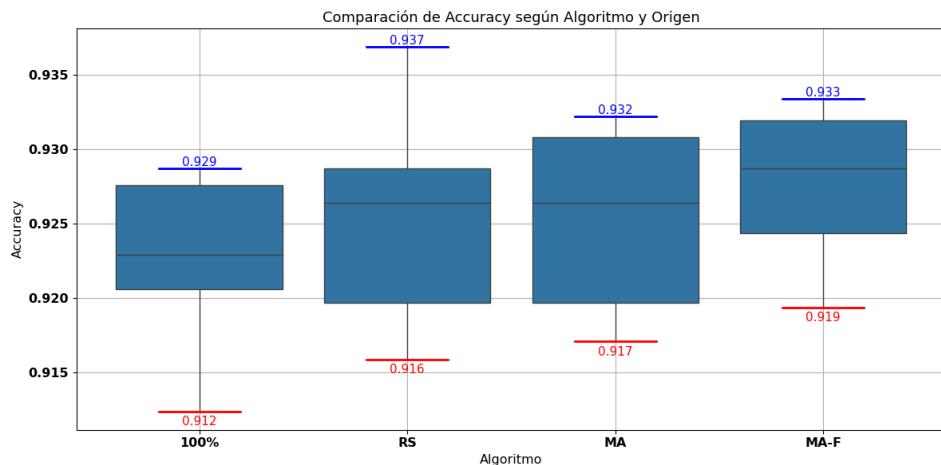


Figura 8.24: Boxplot comparando resultados con el dataset PAINTING usando *accuracy*.

máximo competitivo, su rendimiento es menos estable. Estos resultados reforzarán que la representatividad del subconjunto es un factor más relevante que su tamaño absoluto.

8.11.2. Evolución del tamaño del subconjunto

La Figura 8.25 muestra cómo evoluciona dinámicamente el tamaño final de los subconjuntos generados por MA-F. El algoritmo incrementa el número de instancias cuando detecta beneficios significativos en la precisión, siendo más agresivo con puntos de partida bajos. Por ejemplo, al iniciar con el 25 % y 50 %, el tamaño final aumenta hasta un 41,35 % y 68,69 % respectivamente. En cambio, con un 75 % inicial, el incremento es más moderado, alcanzando un 78,29 %.

8.11.3. Impacto del porcentaje inicial

Los resultados de las Figuras 8.26 y 8.25, junto con la Tabla 8.8, reforzarán la conclusión de que no es necesario utilizar el 100 % del conjunto para lograr rendimientos óptimos. Los porcentajes iniciales del 50 % y 75 % muestran una excelente capacidad para generalizar, superando al conjunto completo en estabilidad y máxima precisión alcanzada (**0.933**). El 25 % inicial, si bien logra resultados competitivos, evidencia una mayor sensibilidad a la variabilidad del subconjunto, una tendencia que también se manifestó en los resultados del dataset RPS cuando se empleaban subconjuntos muy reducidos.

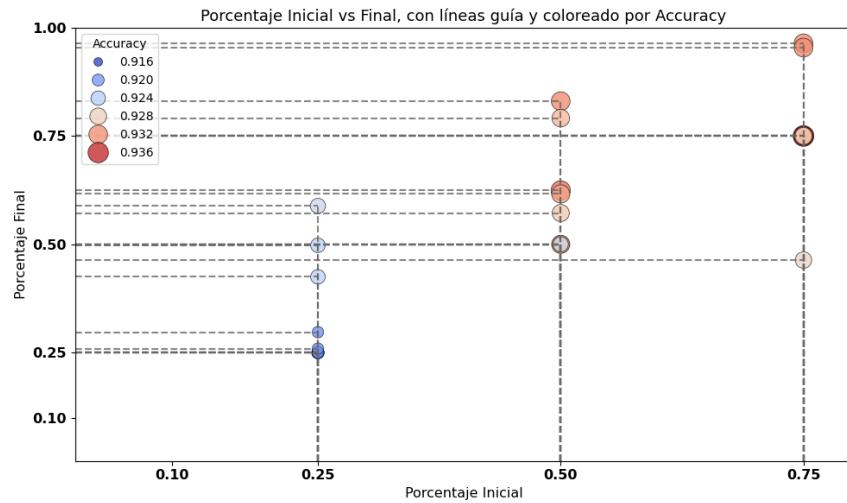


Figura 8.25: Comparacion del tamaño del subconjunto seleccionado por los MA y RS según el porcentaje inicial y mostrando accuracy.

Algoritmo	Accuracy (Avg)	Precision (Avg)	Recall (Avg)	F1-score (Avg)	Porc. Final
25 %					
RS	91,80 %	91,49 %	91,80 %	91,57 %	25,00 %
MA	91,89 %	91,63 %	91,89 %	91,67 %	25,00 %
MA-F	92,24 %	92,06 %	92,24 %	92,09 %	41,35 %
50 %					
RS	92,69 %	92,52 %	92,69 %	92,55 %	50,00 %
MA	92,73 %	92,61 %	92,73 %	92,63 %	50,00 %
MA-F	93,11 %	93,00 %	93,11 %	93,01 %	68,69 %
75 %					
RS	93,14 %	92,97 %	93,14 %	93,01 %	75,00 %
MA	93,11 %	92,93 %	93,11 %	92,99 %	75,29 %
MA-F	93,02 %	92,88 %	93,02 %	92,92 %	78,29 %
100 %					
100 %	92,24 %	92,05 %	92,24 %	92,06 %	100,00 %

Tabla 8.8: Resultados de los algoritmos meméticos, aleatorio y uso del 100 % en el dataset PAINTING.

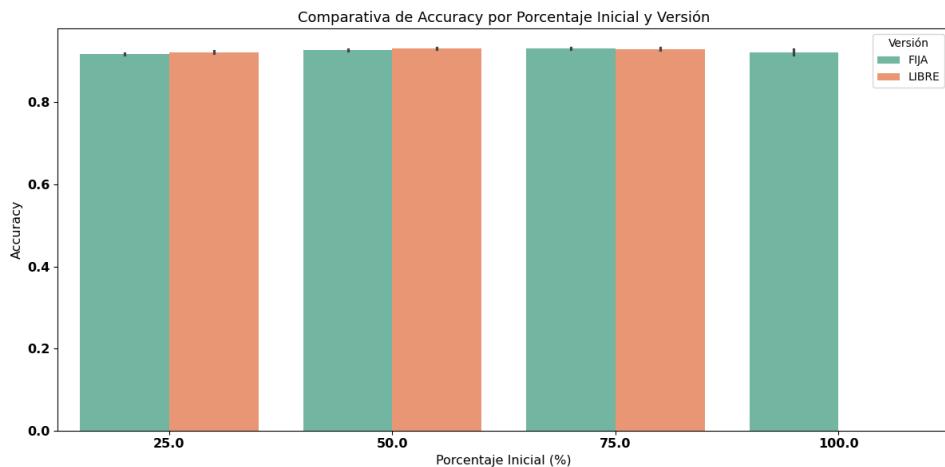


Figura 8.26: Diagrama de barras de *accuracy* según el porcentaje inicial de datos del MA, MA-F y el 100 %.

8.11.4. Distribución y equilibrio de clases

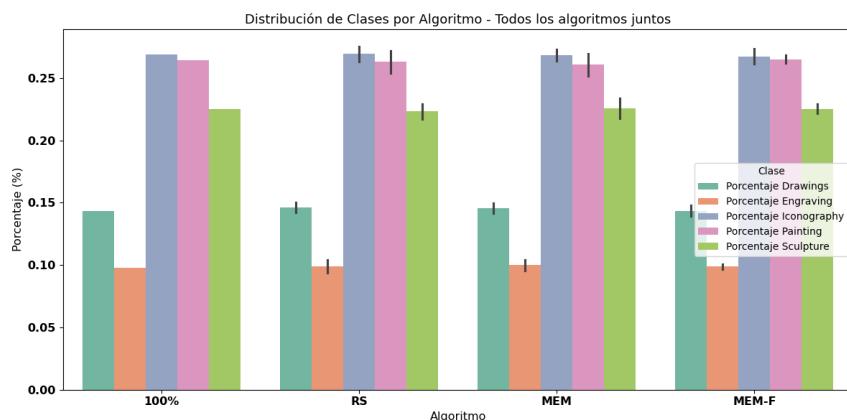


Figura 8.27: Distribución de clases seleccionadas por cada algoritmo.

La Figura 8.27 muestra cómo se preserva la proporción entre clases, un aspecto que también se respetó en experimentos anteriores. La capacidad de mantener esta diversidad es clave para evitar sesgos de entrenamiento, como se discutió al comparar los métodos evolutivos frente a selección aleatoria.

8.11.5. Síntesis final

La validación con el dataset PAINTING confirma la efectividad y generalización de los algoritmos meméticos, especialmente la variante libre. Los

resultados subrayan que seleccionar datos representativos es más beneficioso que usar el conjunto completo, ofreciendo precisión elevada con menor coste computacional.

Este análisis complementa y confirma las conclusiones obtenidas con RPS, incluso en un entorno más complejo y multiclasé, demuestra que el enfoque evolutivo no solo es eficaz, sino también robusto y transferible, reforzando el valor práctico de los enfoques evolutivos propuestos.

8.12. Validación con el dataset CIFAR10

Esta sección tiene como objetivo comprobar si las conclusiones obtenidas en los apartados anteriores, especialmente sobre el comportamiento de los algoritmos meméticos y las versiones libres, se mantienen en un entorno más desafiante como CIFAR10. A diferencia de los datasets anteriores, este conjunto presenta una mayor complejidad visual, baja resolución y una fuerte variabilidad intraclase, lo que lo convierte en un entorno ideal para validar la generalización y robustez de las soluciones propuestas. Asimismo, se busca identificar posibles contradicciones o limitaciones que pudieran no haber sido evidentes en entornos más simples.

Para ello, se evalúan los algoritmos que han mostrado mejor rendimiento hasta ahora (GA-AM, GA-AM-F, MA, MA-F) y se incluye como referencia el modelo entrenado con el 100 % del conjunto de datos.

En este caso, se han realizado únicamente 4 ejecuciones por configuración en lugar de las 5 habituales en el resto del proyecto, debido a una nueva limitación impuesta en el servidor utilizado, su elevada saturación y la falta de tiempo disponible para completar todas las evaluaciones con un dataset de este tamaño.

8.12.1. Evaluación general del rendimiento

La Figura 8.28 presenta la distribución de precisión obtenida por cada algoritmo. Las cuatro variantes superan la mediana del modelo de referencia (**0.778**), confirmando que una selección informada puede ser más eficaz que el uso completo e indiscriminado de datos. Destacan GA-AM-F y MA-F, que alcanzan máximos de **0.805** y **0.808**, respectivamente.

Más allá de la precisión máxima, se observa que las versiones libres reducen la dispersión entre ejecuciones, reforzando su estabilidad en condiciones más adversas. Este comportamiento respalda lo ya observado en los datasets RPS y PAINTING.

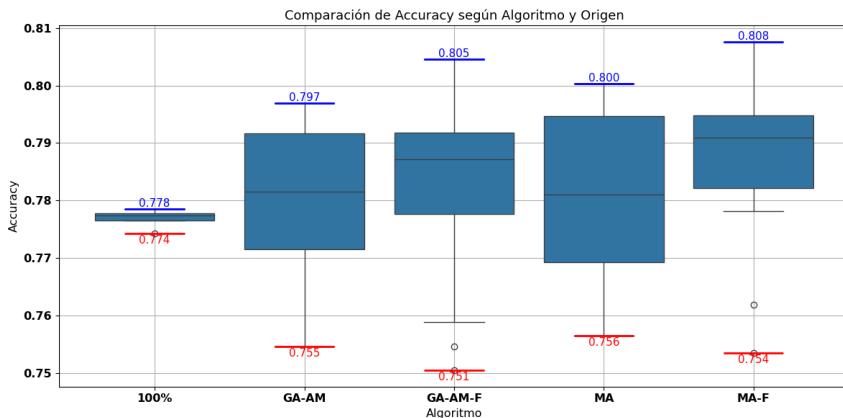


Figura 8.28: Distribución de accuracy obtenida para cada algoritmo sobre CIFAR10.

8.12.2. Comparativa con las versiones libres

La Figura 8.29 compara directamente las versiones FIJA y LIBRE para los algoritmos GA-AM y MA. Al igual que en datasets anteriores, las versiones libres muestran una ligera mejora en precisión media y una reducción de la varianza.

No obstante, en este entorno más exigente, la diferencia entre versiones fijas y libres se reduce, lo que sugiere que la ventaja del ajuste dinámico del tamaño del subconjunto puede verse atenuada por la complejidad del conjunto de datos. Aun así, las versiones libres mantienen una ligera superioridad, lo que valida su utilidad incluso en escenarios de mayor dificultad.

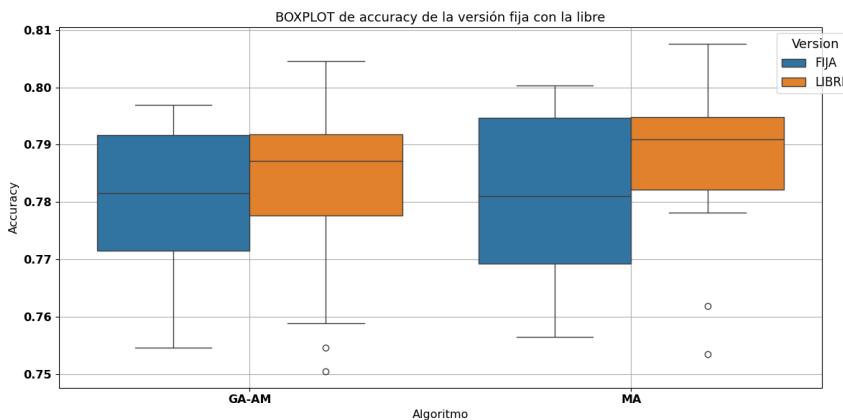


Figura 8.29: Comparativa de precisión entre versiones fijas y libres para los algoritmos evolutivos.

8.12.3. Consistencia frente al porcentaje inicial

La Figura 8.30 muestra cómo evoluciona la precisión media en función del porcentaje de datos iniciales. Se observa que los valores iniciales del 50 % y 75 % destacan de forma consistente, logrando precisiones iguales o superiores al entrenamiento completo. Este comportamiento reafirma la validez de los resultados obtenidos en PAINTING, donde el 50 % superaba al 100 %, y sugiere que estos porcentajes representan puntos de partida especialmente eficientes para la selección de subconjuntos informativos.

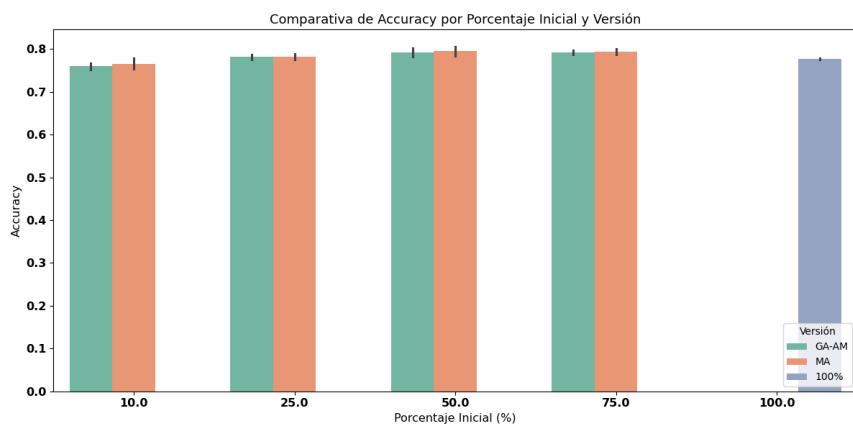


Figura 8.30: Precisión media alcanzada según el porcentaje inicial de datos.

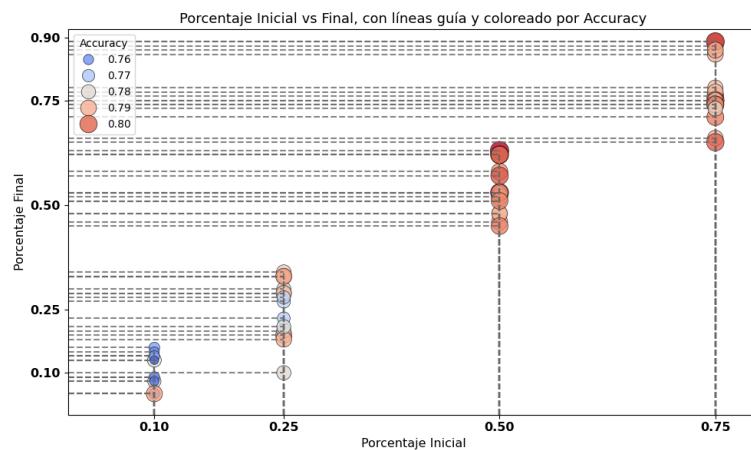


Figura 8.31: Porcentaje inicial vs porcentaje final de datos seleccionados.

Para comprender mejor el comportamiento de los algoritmos, se muestra la Figura 8.31. A partir de esta visualización, se pueden extraer varios patrones relevantes:

1. **Trayectorias eficientes desde valores intermedios.** Los puntos de partida del 50 % y el 75 % no solo conducen a altas precisiones, sino que lo hacen manteniendo tamaños de subconjuntos finales relativamente estables, lo que sugiere que permiten al algoritmo trabajar con confianza desde etapas tempranas.
2. **Zona de alta precisión estabilizada.** Existe una franja de subconjuntos finales, en torno al 60–80 %, donde se concentra la mayor parte de las ejecuciones con mejor precisión. Esto sugiere que este rango podría representar un punto de equilibrio entre información útil y redundancia.
3. **Tendencia a expandir el subconjunto de datos.** A diferencia de experimentos anteriores, donde algunos algoritmos reducían el número de ejemplos seleccionados respecto al conjunto inicial, en este caso se observa una inclinación general a ampliar el tamaño del subconjunto final. Esta conducta sugiere una estrategia orientada a reforzar la representatividad del conjunto, priorizando una cobertura más amplia cuando la información inicial es limitada.
4. **Baja rentabilidad en extremos.** Las ejecuciones que parten de 10 % o 100 % (teniendo el contexto de las otras figuras) rara vez conducen a los mejores resultados: en el primer caso, la falta de diversidad inicial limita la exploración efectiva; en el segundo, la inclusión de datos redundantes parece perjudicar la optimización.

En conjunto, los resultados demuestran que los algoritmos son capaces de ajustar dinámicamente el tamaño de los subconjuntos, maximizando la precisión con una fracción moderada de datos. Especialmente, los arranques desde el 50 % y 75 % ofrecen un equilibrio favorable entre diversidad y eficiencia, lo que los convierte en opciones recomendables en escenarios de selección activa de datos.

8.12.4. Distribución de clases

La Figura 8.32 presenta la distribución de clases en los subconjuntos generados por los algoritmos. Todos mantienen una proporción equilibrada, lo que garantiza que las mejoras de rendimiento no provienen de sesgos estructurales, sino de un proceso evolutivo eficaz.

Este equilibrio ya fue observado en RPS y PAINTING, lo que refuerza la consistencia de los algoritmos propuestos en distintos contextos.

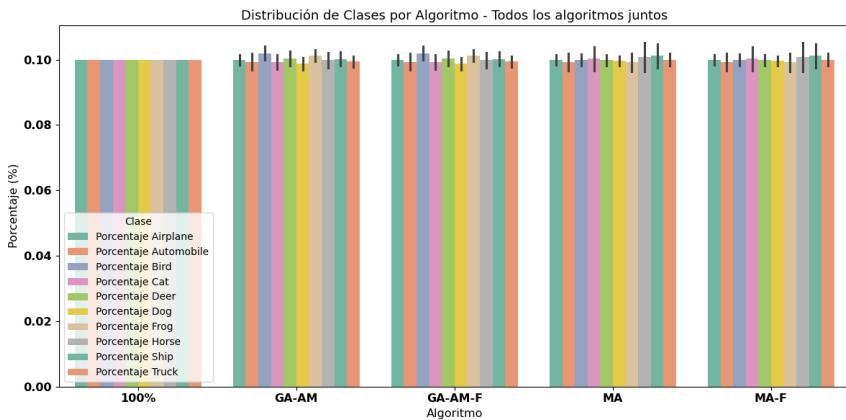


Figura 8.32: Porcentaje de instancias por clase tras aplicar cada algoritmo. La distribución permanece prácticamente uniforme.

8.12.5. Síntesis final de la validación con CIFAR10

Los resultados obtenidos con CIFAR10 permiten validar las principales conclusiones extraídas en las fases anteriores del estudio:

- Todos los algoritmos evaluados superan el rendimiento del modelo entrenado con el 100 % del dataset, confirmando que una selección óptima puede ser más eficaz que el uso de datos sin filtrar.
- Las variantes libres (-F) mantienen su superioridad frente a las versiones fijas, aunque en este caso la ventaja es más moderada.
- A partir del 50 % de datos, los algoritmos alcanzan rendimientos muy competitivos, llegando incluso a superar bastante al entrenamiento completo.
- La distribución de clases permanece estable en todos los casos, descartando que las mejoras observadas se deban a sesgos de clase.
- En contraste con datasets anteriores, en CIFAR10 los algoritmos tienden a expandir los subconjuntos seleccionados, lo que sugiere una adaptación activa frente a la mayor complejidad visual del entorno.

En resumen, MA-F vuelve a consolidarse como la opción más prometedora, combinando precisión elevada, baja varianza y adaptabilidad ante diferentes porcentajes iniciales. Además, esta validación refuerza que las conclusiones alcanzadas en entornos más simples son extrapolables a escenarios de mayor dificultad, demostrando la solidez general del enfoque propuesto.

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 tres conjuntos de datos con distintos niveles de complejidad: ‘Rock, Paper, Scissors’, ‘PAINTING’ y ‘CIFAR-10’.

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 porcentaje 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, y con un tercer conjunto ('CIFAR-10'), considerablemente más exigente por su variabilidad intraclasa y baja resolución, 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. Aunque la selección de instancias conlleva un coste computacional adicional durante su ejecución, este esfuerzo inicial se justifica al permitir una notable reducción en el tiempo de entrenamiento e inferencia de los modelos explorados. Además, facilita la auditoría y trazabilidad de los datos, aspectos fundamentales en entornos donde estas exigencias normativas son prioritarias. A medio plazo, el uso de subconjuntos optimizados puede reducir el coste global cuando se entrena múltiples modelos o configuraciones diferentes.

Como líneas futuras, se plantea:

- Aplicar las técnicas propuestas en tareas más complejas, como la segmentación semántica o la clasificación multietiqueta, donde la reducción del conjunto de datos puede tener un impacto aún más relevante.
- Incorporar múltiples objetivos en el proceso de selección, convirtiéndolo en un problema de optimización multiobjetivo. Por ejemplo, se podría considerar aspectos como el tiempo de entrenamiento, el coste computacional total o la eficiencia en la fase de inferencia, además de las métricas clásicas de rendimiento.

- Explorar la integración de los algoritmos meméticos con otras estrategias de reducción de datos, como la destilación de conocimiento o la poda de redes neuronales, con el objetivo de combinar distintas aproximaciones complementarias.

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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.48550/arXiv.1704.04861). arXiv: [1704.04861 \[cs\]](https://arxiv.org/abs/1704.04861).
- [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](https://doi.org/10.48550/arXiv.1801.04381). arXiv: [1801.04381 \[cs\]](https://arxiv.org/abs/1801.04381).
- [16] 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.
- [17] 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](https://doi.org/10.1007/978-1-4842-5364-9_2).
- [18] *torch.cuda — PyTorch 2.4 documentation*, <https://pytorch.org/docs/stable/cuda.html>. visitado 8 de oct. de 2024.
- [19] 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](https://doi.org/10.1007/978-3-319-33383-0_5).
- [20] *NumPy v2.0 Manual*, <https://numpy.org/doc/2.0/index.html>. visitado 24 de feb. de 2025.
- [21] *Polars — Python API reference*, <https://docs.pola.rs/api/python/stable/reference/>. visitado 24 de feb. de 2025.

- [22] *Pandas 2.2.3 documentation*, <https://pandas.pydata.org/pandas-docs/version/2.2/index.html>. visitado 8 de jun. de 2025.
- [23] *Matplotlib 3.9.3 documentation*, <https://matplotlib.org/3.9.3/index.html>. visitado 24 de feb. de 2025.
- [24] *Seaborn 0.13.2 documentation*, <https://seaborn.pydata.org/tutorial.html>. visitado 3 de mayo de 2025.
- [25] *Openpyxl 3.1.3 documentation*, <https://openpyxl.readthedocs.io/en/stable/>. visitado 3 de mayo de 2025.
- [26] *Creation of virtual environments*, <https://docs.python.org/3/library/venv.html>. visitado 24 de feb. de 2025.
- [27] *Conda Documentation*, <https://docs.conda.io/en/latest/>. visitado 24 de feb. de 2025.
- [28] *cuBLAS Deterministic Algorithms*, https://docs.nvidia.com/cuda/cublas/index.html#cublasApi_reproducibility. visitado 24 de feb. de 2025.
- [29] *Early Stopping Discussion*, <https://www.geeksforgeeks.org/early-stopping-on-validation-loss-or-on-accuracy/>, feb. de 2024. visitado 4 de mayo de 2025.
- [30] *Rock Paper Scissors Dataset*, <https://public.roboflow.com/classification/rock-paper-scissors>. visitado 24 de feb. de 2025.
- [31] *(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.
- [32] *(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.
- [33] *CIFAR-10 dataset*, <https://www.cs.toronto.edu/~kriz/cifar.html>. visitado 12 de jun. de 2025.
- [34] *Cifar10-PyTorch*, https://docs.pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html. visitado 12 de jun. de 2025.

