# **Resumen Experimentos\_Leticia (ODIR-5K · Modelo Multiclase CNN + Metadatos)**

Arquitectura común: ResNet18 (imagen) + MLP (metadatos, meta\_dim=2), fusión por concatenación y cabeza MLP (256 + dropout=0.5).

Optimización: AdamW (con weight\_decay), CrossEntropy (+ label\_smoothing=0.05).

Desbalanceo: class\_weights en la pérdida (sin sampler).

Tracking (MLflow): train\_loss, train\_acc, train\_f1m, val\_loss, val\_acc, val\_f1m, y métricas finales en test; artefactos de classification\_report y confusion\_matrix (val/test).

## **Exp\_1 — \*Experimento\_Leticia\_ResNet18\***

Fecha: 23/08/2025

Run ID: eb3b723076f3496d895e74a4af59b742

Ejecución: 1.4 h (CPU)

Configuración:

* Transform (train): RandomHorizontalFlip(p=0.2), RandomRotation(15), normalización ImageNet.
* Parámetros: epochs=15, batch\_size=32, lr=1e-4, weight\_decay=2e-4, num\_workers=0, num\_classes=8.
* Criterio de selección: F1-macro en validación; Early Stopping paciencia=6.

Resultados:

* Train: Acc 0.851, F1 0.876
* Validación: Acc 0.579, F1 0.510
* Test: Acc 0.608, F1 0.562
* Evaluación: overfitting (train F1 0.876 vs val F1 0.510). Clases minoritarias (p.ej., clase 4) con desempeño bajo. Se requiere más regularización o data augmentation.

## Exp\_2 — \*Experimento\_Leticia\_ResNet18\*

Fecha: 24/08/2025

Run ID: df6bf684983c436680940f450870c1c7

Ejecución: 54.4 min (GPU)

Cambios respecto a Exp\_1:

* batch\_size 32→64 (tiene mejor estabilidad de gradiente en GPU).
* epochs 15→20, para convergencia
* LR 1e-4→1e-5 (controlar sobreajuste).
* weight\_decay 2e-4→5e-4 (más regularización).
* Data Augmentation: Flip(p=0.5) + Rot(10) + RandomErasing(0.1).
* num\_workers=2 (entorno Colab).
* Se añade la métrica prec\_m al logging.

Resultados:

* Train: Acc 0.899, F1 0.907
* Validación: Acc 0.582, F1 0.546
* Test: Acc 0.595, F1 0.573

Evaluación: Ligera mejora frente a Exp\_1, con algo menos de overfitting y mejor recall global gracias a la regularización y técnicas de augmentación.

## Exp\_3 — \*Experimento\_Leticia\_ResNet18 (solo con imágenes)\*

Fecha: 25/08/2025

Run ID: 06a0c25ec51c47948bc3ebe0cf0d8ae3

Ejecución: 53.44 min (GPU)

Cambios respecto a Exp\_2:

* Sin metadatos (flujo solo con imágenes).
* epochs=25; augmentation: Flip(p=0.5) + Rot(10) + RandomErasing(0.1).
* Hiperparámetros: batch\_size=32, lr=1e-4, weight\_decay=5e-4, num\_workers=2, label\_smoothing=0.05, num\_classes=8.

Resultados:

* Train: Acc 0.735, F1 0.773
* Validación: Acc 0.483, F1 0.467
* Test: Acc 0.546, F1 0.528

Evaluación: Quitar metadatos empeora el rendimiento (menor F1 val/test).

## Exp\_4 — \*Experimento\_4\_Leticia\_ResNet18\_Top5\*

Fecha: 25/08/2025

Run ID: 3189a256d9b544f5906f41cd35cfc46e

Ejecución: 1.6 h (CPU)

Cambios respecto a Exp\_3:

* Se reintroducen metadatos.
* Reducción de clases a [0,1,2,5,6] → num\_classes=5.
* epochs=20, augmentation enriquecido: RandomAffine(±7°, ±2% shift, 0.95–1.05 scale) + ColorJitter(0.10/0.10/0.10/0.02) + Flip(0.5) + RandomErasing(0.25).

Resultados:

* Train: Acc 0.922, F1 0.943
* Validación: Acc 0.710, F1 0.726
* Test: Acc 0.724, F1 0.750

Evaluación: Salto sustancial de rendimiento al centrar el problema en clases más frecuentes y mejorar la augmentación. Más robustez y mejor equilibrio entre clases. Aunque se sigue existiendo overfitting.

## Exp\_5 — \*Experimento\_4\_Leticia\_ResNet18\_Top5\*

Fecha: 26/08/2025

Run ID: 70a48d006d5d4368b50d1d92a28be053

Ejecución: 1.9 h (CPU)

Cambios respecto a Exp\_4:

* Scheduler ReduceLROnPlateau (menor LR cuando no mejora).
* AMP (Mixed Precision) habilitado cuando hay GPU.
* Logging extra: LR por época, frecuencias de clase del train, matriz de confusión normalizada, predicciones (.npz).

Resultados:

* Train: Acc 0.943, F1 0.960
* Validación: Acc 0.702, F1 0.738
* Test: Acc 0.720, F1 0.775

Evaluación: Hay una ligera mejora sobre Exp\_4. El scheduler estabiliza la convergencia; mejor F1 y recall. Se evalúa como el mejor run.

## Tabla resumen de métricas

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Train Acc** | **Train F1** | **Val Acc** | **Val F1** | **Test Acc** | **Test F1** | **Test Prec** | **Test Rec** |
| **1** | 0.851 | 0.876 | 0.579 | 0.510 | 0.608 | 0.562 | 0.570 | 0.571 |
| **2** | 0.899 | 0.907 | 0.582 | 0.546 | 0.595 | 0.573 | 0.551 | 0.609 |
| **3** | 0.735 | 0.773 | 0.483 | 0.467 | 0.546 | 0.528 | 0.484 | 0.610 |
| **4** | 0.922 | 0.943 | 0.710 | 0.726 | 0.724 | 0.750 | 0.722 | 0.784 |
| **5** | 0.943 | 0.960 | 0.702 | 0.738 | 0.720 | 0.775 | 0.772 | 0.780 |

## Hiperarámetros principales

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Batch** | **Epochs** | **LR inicial** | **Weight Decay** | **Augmentación principal** | **Dropout** | **Scheduler** | **Optimizer** |
| **1** | 32 | 15 | 1e-4 | 2e-4 | Flip(0.2)+Rot(15) | 0.5 | No | AdamW |
| **2** | 64 | 20 | 1e-5 | 5e-4 | Flip(0.5)+Rot(10)+Erasing(0.1) | 0.5 | No | AdamW |
| **3** | 32 | 25 | 1e-4 | 5e-4 | Flip(0.5)+Rot(10)+Erasing(0.1) | 0.5 | No | AdamW |
| **4** | 32 | 20 | 1e-4 | 5e-4 | Affine+ColorJitter+Flip(0.5)+Erasing(0.25) | 0.5 | No | AdamW |
| **5** | 32 | 20 | 1e-4 | 5e-4 | Affine+ColorJitter+Flip(0.5)+Erasing(0.25) | 0.5 | Sí | AdamW |

## Conclusiones

* Los Metadatos sí tienen valor en el modelo, retirarlos (Exp\_3) redujo el F1; su inclusión mejora la discriminación entre clases próximas.
* La Regularización y augmentación, como rotaciones moderadas, erasing y jitter redujeron el sobreajuste observado en Exp\_1, aunque no lo suficiente.
* Limitar el modelo al entrenamiento con las clases [0,1,2,5,6] (Exp\_4–5) incrementó notablemente F1 y estabilidad inter-clase.
* La Optimización dinámica: ReduceLROnPlateau (Exp\_5) aportó mejoras adicionales en F1 y recall.
* El Mejor resultdao se confirma que se ha obtenido en el Exp\_5, pero sin alcanzar el objetivo de eliminar el overfitting persistente.