

AUTOMATIZACIÓN DE ENSAYOS ECOTOXICOLÓGICOS CON MÉTODOS DE APRENDIZAJE AUTOMÁTICO

Joaquín Nicolás Cibeira, Nicolás Taurozzi
 Director: Dr. Ing. Ramiro Germán Rodríguez Colmeiro

Introducción

El ensayo *Allium cepa* [1] es un bioensayo ecotoxicológico fundamental para evaluar la toxicidad celular mediante el cultivo de bulbos de cebolla en contacto con agentes químicos. En este estudio, el experto debe inspeccionar manualmente miles de células de las raíces para diferenciar la interfase de las fases mitóticas (profase, metafase, anafase y telofase), tarea central para el cálculo del Índice Mitótico (*IM*). No obstante, este proceso resulta intrínsecamente subjetivo y propenso a la fatiga visual, degradando la precisión del estudio. Esta investigación propone automatizar dicha tarea mediante aprendizaje automático, logrando un cálculo del *IM* objetivo y de alto rendimiento.

Metodología

La solución propuesta se articula en cuatro etapas clave: refinamiento del *Ground Truth* para mitigar el ruido en las etiquetas, detección mediante YOLOv11 [2], clasificación de fases con EfficientNet-B2 y aumento generativo para compensar el desbalance biológico. El flujo de trabajo integral, desde la captura hasta el reporte, se ilustra en la Fig. 1.

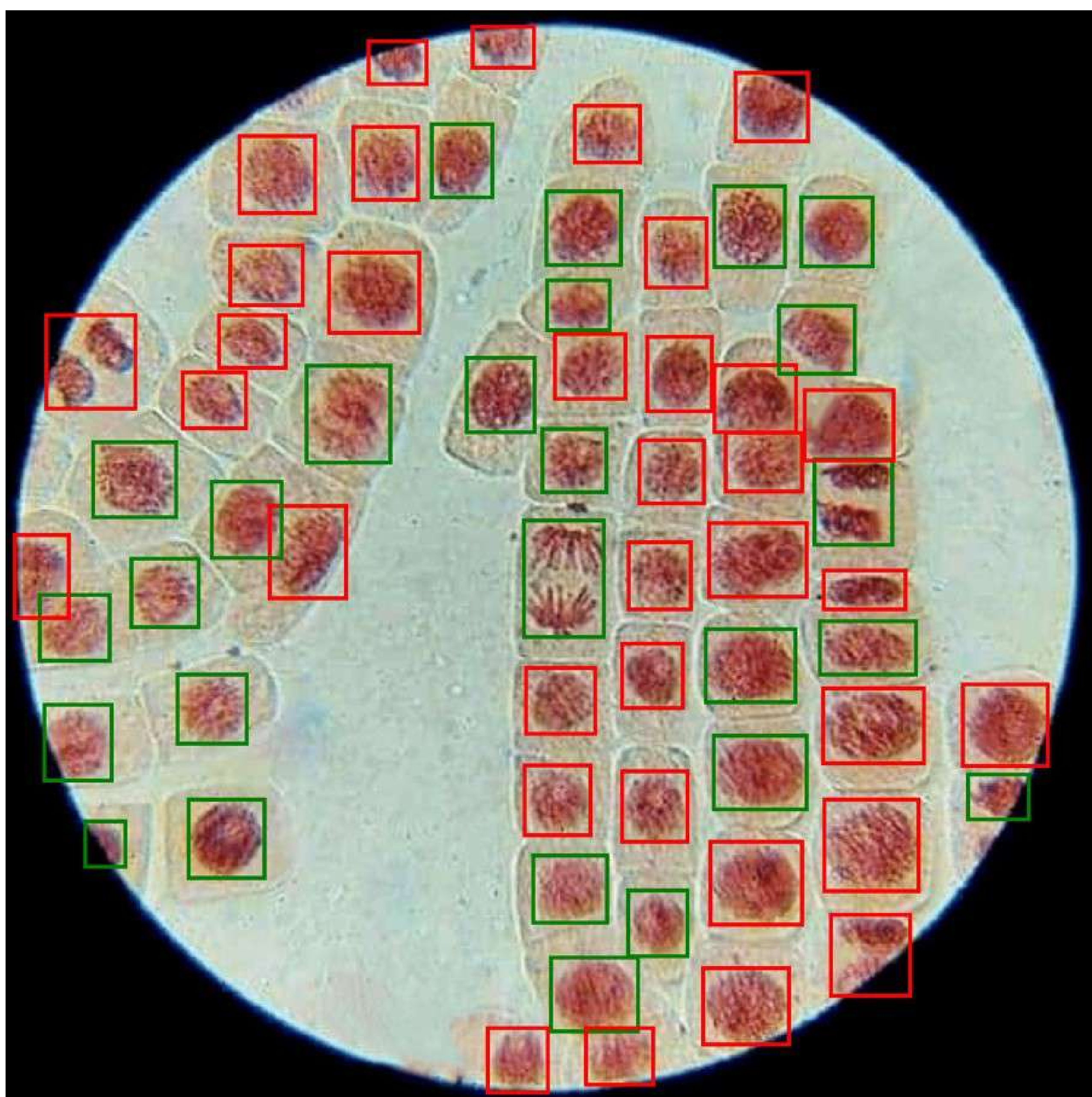


Fig. 1: Resultado de la detección y clasificación de células. Recuadradas en rojo se observan las células en interfase y en verde las que se encuentran en mitosis.

Calibración de modelos e incertidumbre

Debido a la tendencia de las redes neuronales hacia la sobreconfianza, se implementó una calibración probabilística [3] para que las salidas reflejen la precisión real. En el detector YOLOv11, la Regresión Isotónica redujo el error de calibración (*D-ECE*) de un 8.51 % a valores cercanos al 0 %. En el clasificador EfficientNet-B2, el escalado de temperatura (*CSTS*) redujo el error (*ECE*) en un 50 %, permitiendo fundamentar el cálculo estadístico del *IM*. Los diagramas de fiabilidad resultantes, que comparan el desempeño del detector y el clasificador, se detallan en la Fig. 2.

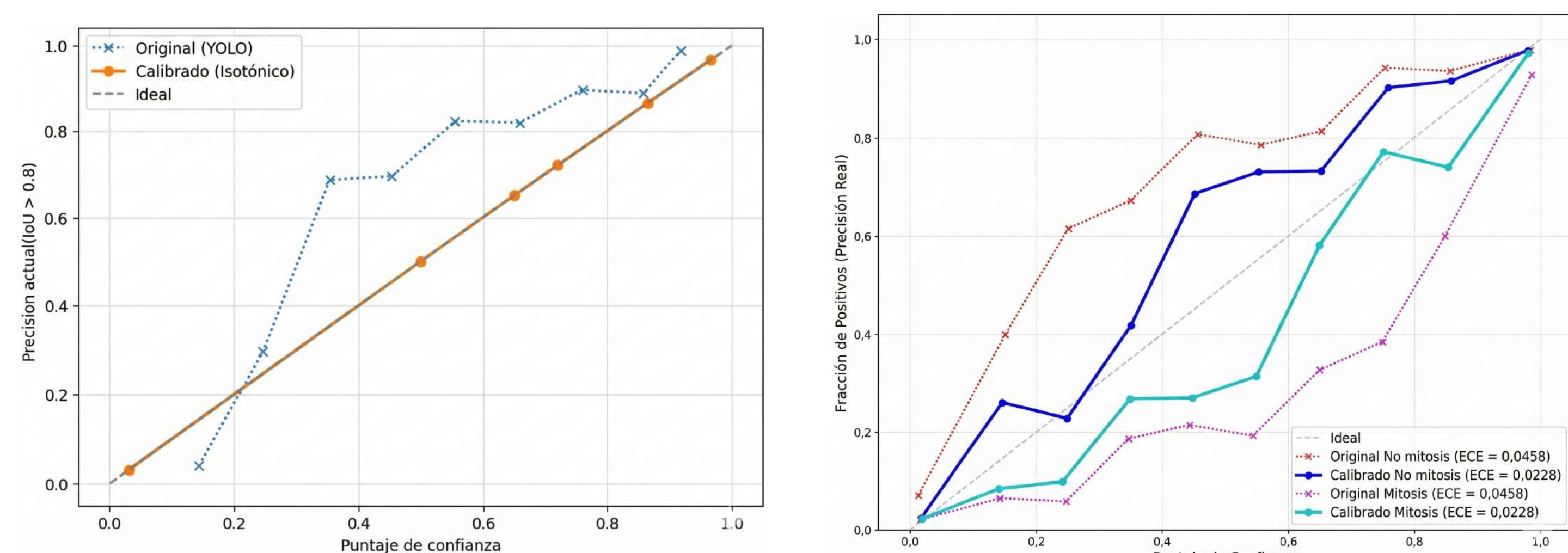


Fig. 2: Diagramas de fiabilidad: calibración del detector mediante regresión isotónica (izq.) y del clasificador mediante CSTS (der.).

Resultados del conteo mitótico

La validación del sistema se realizó comparando el desempeño automático frente al experto manual. Como se observa en el Cuadro 1, el sistema proporciona una estimación de la incertidumbre asociada, permitiendo al toxicólogo evaluar la confianza del ensayo en tiempo real.

Parámetro	Método manual	Resultado automático
Instancias totales	428	432 ± 12
Índice Mitótico (<i>IM</i>)	16.59 %	18,65 % ± 2,7 %

Cuadro 1: Comparación de resultados del Índice Mitótico.

Generación de datos sintéticos

Para abordar la escasez intrínseca de datos en las fases menos frecuentes del ciclo celular, se propone la generación de muestras sintéticas. Mediante un β -VAE [4], se logra una representación compacta del espacio latente que permite tratar la mitosis como una transición biológica continua. Este enfoque otorga un control preciso sobre las células generadas, cuya evolución suave se presenta en la caminata latente de la Fig. 3.

A partir de estas imágenes estructurales, se aplican modelos de difusión para dotar a las imágenes de texturas realistas y alta calidad fotográfica, tal como ilustra la Fig. 4. Como trabajo futuro, se espera integrar estos datos generados al conjunto de entrenamiento para mitigar el desequilibrio de clases, logrando así aumentar la precisión de los clasificadores.

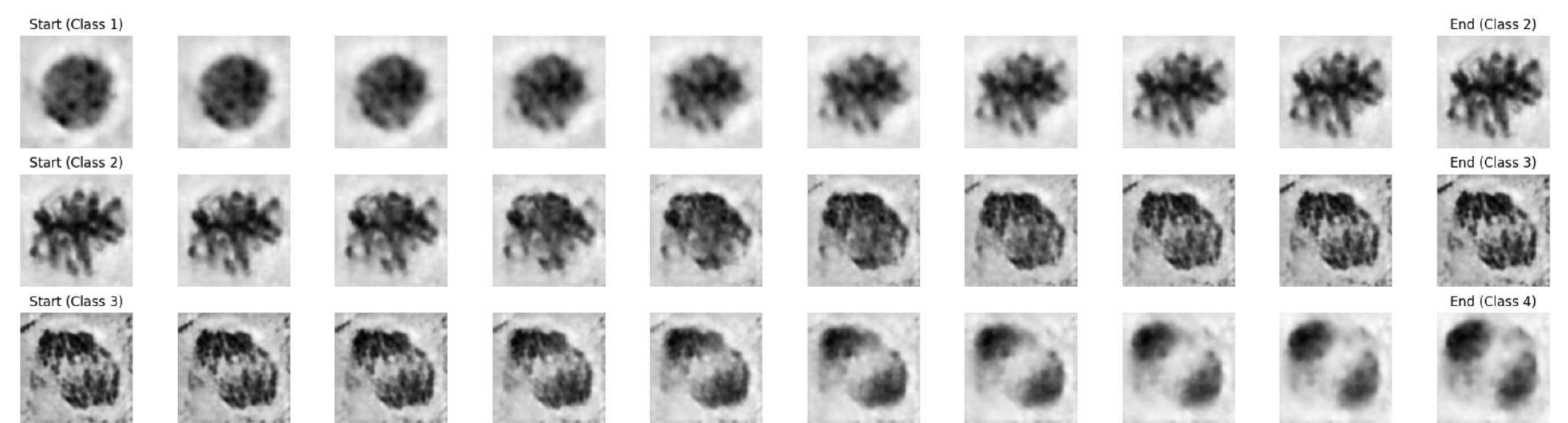


Fig. 3: Interpolación latente (Fig. 27a) que muestra la evolución de la cromatina entre estados.

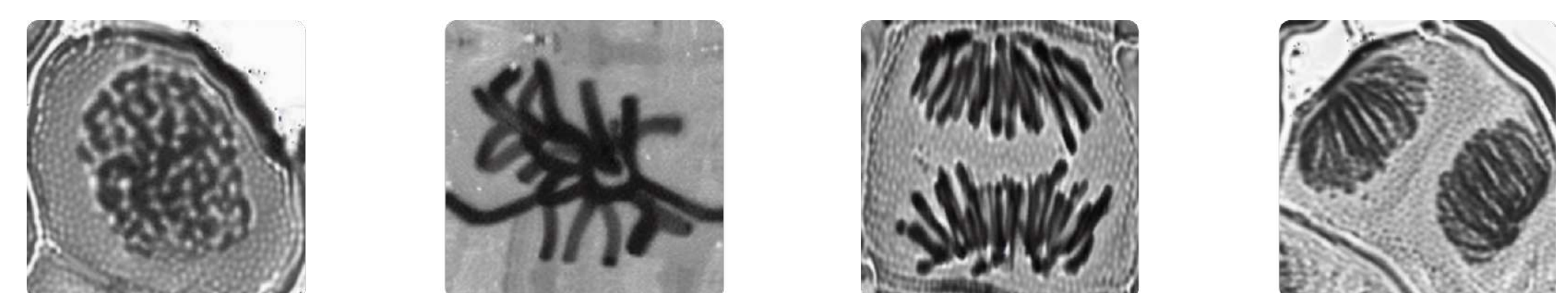


Fig. 4: Muestras generadas por difusión de las fases mitóticas, cuya textura biológica está sujeta a perfeccionamiento futuro.

Conclusiones y trabajo futuro

La implementación del sistema permite alcanzar niveles de precisión comparables a los de un experto humano, logrando simultáneamente una reducción drástica en los tiempos de procesamiento. Asimismo, la robustez del *dataset* consolidado garantiza la invariabilidad del modelo frente a las diversas condiciones de las muestras biológicas, estableciendo de esta manera una herramienta objetiva para el monitoreo ambiental continuo y reproducible que elimina la subjetividad del análisis tradicional.

Como línea de investigación futura dentro del GIAR, se continuará con el desarrollo de *datasets* sintéticos. La integración de estas muestras generadas permitirá robustecer el entrenamiento de los clasificadores frente al desequilibrio de clases, consolidando metodologías avanzadas de aumento de datos para habilitar el cálculo de índices mitóticos específicos por cada fase.

Referencias

- [1] D. M. Leme and M. A. Marin-Morales, "Allium cepa test in environmental monitoring: A review on its application," *Mutation Research/Reviews in Mutation Research*, vol. 682, no. 1, pp. 71-81, 2009.
- [2] G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics yolov11," 2024. Arquitectura utilizada para la detección de instancias celulares.
- [3] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International Conference on Machine Learning (ICML)*, pp. 1321-1330, PMLR, 2017. Referencia técnica para las métricas ECE y los métodos de calibración.
- [4] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," in *International Conference on Learning Representations (ICLR)*, 2017. Base teórica para el modelado del espacio latente y el Latent Walk.