

Automating the *Allium cepa* Bioassay Through Deep Learning

Nicolás Taurozzi¹ Joaquín Cibeira¹ Ramiro G. Rodríguez Colmeiro¹

¹Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Grupo de Inteligencia Artificial y Robótica

{ntaurozzi, jcibeira, rrodriguezcolmeiro}@frba.utn.edu.ar

Abstract This study proposes a two-stage computer vision methodology to automate the Mitotic Index (MI) in the *Allium cepa* bioassay, a key tool for the evaluation of liquid effluent toxicity and discharge auditing. The first stage utilizes YOLOv11 for cell detection, outperforming agnostic segmentation methods (e.g., SAM) by preventing fragmentation in advanced mitotic phases. The second stage uses EfficientNetB2 as a binary classifier (Interphase/Mitosis). A distinctive contribution is probabilistic calibration, allowing the MI estimation to include uncertainty quantification through confidence intervals. The research was validated against experts, confirming the results and their uncertainty bounds. The system achieves F1-Scores of 0.97 for interphase and 0.75 for mitosis, providing objective and reproducible ecotoxicological metrics, laying the technological foundations that enable the enforcement of genotoxicity parameters within current regulatory frameworks.

Keywords: computer vision, mitotic index, ecotoxicology

Automatización del bioensayo *Allium cepa* mediante aprendizaje profundo

Resumen Este estudio propone una metodología de visión artificial en dos etapas para automatizar la obtención del Índice Mitótico (IM) en el bioensayo *Allium cepa*, una herramienta clave para la evaluación de la toxicidad de efluentes líquidos y auditoría de vertidos. La primera etapa utiliza YOLOv11 para la detección celular, superando métodos de segmentación agnóstica (e.g., SAM) al evitar la fragmentación en fases mitóticas avanzadas. La segunda etapa usa EfficientNetB2 como clasificador binario (Interfase/Mitosis). Un aporte diferencial es la calibración probabilística, permitiendo que la estimación del IM incluya cuantificación de incertidumbre mediante intervalos de confianza. La investigación fue validada contra expertos, confirmando el resultado y sus cotas de incertidumbre. El sistema logra F1-Scores de 0.97 en interfase y 0.75 en mitosis, proporcionando métricas ecotoxicológicas objetivas y reproducibles, sentando las bases tecnológicas que habilitan a la exigencia de parámetros de genotoxicidad en los marcos regulatorios vigentes.

Palabras clave: visión artificial, índice mitótico, ecotoxicología

1. Introducción

El monitoreo y control de efluentes líquidos es un pilar fundamental para mitigar el impacto ambiental de los vertidos de origen industrial y doméstico. Entre los diversos contaminantes, aquellos que presentan efectos mutagénicos y genotóxicos (incluyendo agroquímicos, metales pesados y compuestos de origen industrial) resultan de particular preocupación debido a que sus alteraciones a nivel celular tienen el potencial de ser heredables, afectando a múltiples generaciones (Flores & Yamaguchi, 2008). En este contexto, el uso de bioensayos ecotoxicológicos se ha consolidado como una herramienta precisa para evaluar la calidad ambiental y la toxicidad de las muestras (Patra et al. 2004; Zhou et al. 2008).

El ensayo de *Allium cepa* (cebolla común) es un modelo biológico de planta superior ampliamente validado para la detección de este tipo de anomalías (Leme & Marin-Morales, 2009). El procedimiento consiste en la exposición de las raíces del bulbo a los efluentes o muestras de agua a evaluar, seguido de un análisis microscópico de las células meristemáticas para cuantificar alteraciones morfológicas, daños cromosómicos y disrupciones en el ciclo de la división celular (mitosis).

La importancia del ensayo *Allium cepa* para el monitoreo de efluentes radica en su alta sensibilidad y viabilidad técnica para ser estandarizado e incorporado en los marcos regulatorios vigentes. Actualmente, la legislación ambiental carece de consistencia en sus criterios de calidad; en la República Argentina, los métodos de control de vertidos varían drásticamente entre jurisdicciones (Bormioli et al. 2018). Peor aún, las normativas no contemplan la toxicidad emergente producto de las interacciones fisicoquímicas en las mezclas de contaminantes, un factor que el bioensayo de *Allium cepa* puede detectar eficazmente (Bormioli et al. 2019).

El principal obstáculo de este ensayo para su adopción en la normativa ambiental es la alta carga temporal y la variabilidad de la evaluación humana. Clasificar manualmente una inmensa cantidad de células en sus diversos estadios mitóticos (profase, metafase, anafase, telofase e interfase) ralentiza las auditorías e introduce sesgos sistemáticos (Figura 1).

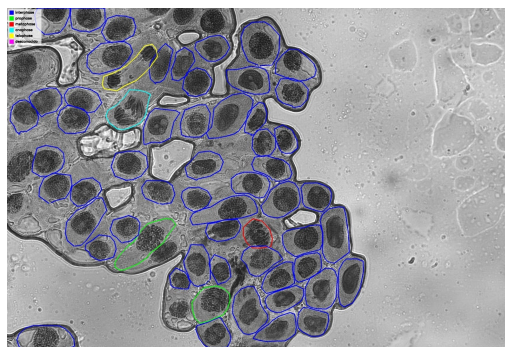


Figura 1: Fotografía con microscopio de preparado de *Allium cepa*. En color se indican las células detectadas por el operador: en verde células en profase, en azul células en interfase, en amarillo células en telofase y en rojo células en metafase.

Para que este bioensayo alcance la escalabilidad necesaria y pueda convertirse en una métrica auditable por los entes reguladores, es necesaria su estandarización procedimental. Este trabajo aborda dicha problemática proponiendo la automatización de la detección y clasificación celular mediante visión artificial. La sistematización de esta lectura no solo elimina la variabilidad humana del operador, sino que reduce los tiempos de análisis a escalas viables para el control industrial periódico, facilitando su eventual integración como requisito en la legislación ambiental.

2. Materiales y Métodos

La metodología propuesta se estructura mediante un *pipeline* de dos etapas: una fase inicial de detección de instancias celulares utilizando YOLOv11 (Khanam & Hussain, 2024) y una fase posterior de clasificación del estado mitótico basada en EfficientNetB2 (Tan & Le, 2019). Este enfoque busca optimizar la integridad de la detección, permitiendo el cálculo automático del Índice Mitótico (IM) mediante una identificación precisa de las células en división.

La transición hacia una arquitectura de detección supervisada surge como respuesta a las limitaciones observadas en investigaciones previas (Mareque et al. 2025), donde la utilización del modelo Segment Anything Model (SAM) (Kirillov et al. 2023) para la segmentación de material genético sin entrenamiento supervisado presenta errores sistemáticos. Específicamente, en fases avanzadas de la mitosis como anafase y telofase, la migración cromosómica hacia polos opuestos induce un fenómeno de sobre-segmentación. En estos casos, los métodos agnósticos tienden a interpretar los grupos cromosómicos separados como instancias celulares independientes en lugar de una unidad biológica única, error que se ilustra en la Figura 2.

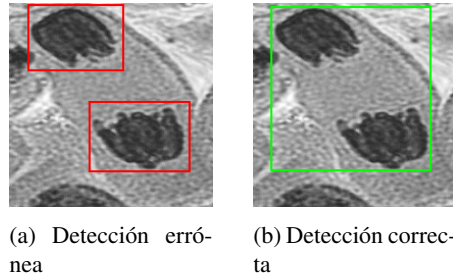


Figura 2: Sobre-segmentación del modelo SAM en telofase (a) frente a la detección correcta del modelo propuesto (b).

Para mitigar este problema de fragmentación, en el presente trabajo se sustituye la segmentación agnóstica por una arquitectura de detección. Al entrenar YOLOv11 con etiquetas reales (*ground truth*) que abarcan la morfología celular completa, se logra una identificación consistente que preserva la cohesión de la instancia incluso ante la fragmentación nuclear.

2.1. Composición y Curación del Conjunto de Datos

Ante la ausencia de bases de datos públicas estandarizadas para esta tarea, se implementa una estrategia de adquisición de datos híbrida que integra fuentes primarias de laboratorio con repositorios externos. El *dataset* final se conforma mediante la integración de tres fuentes diferentes:

- **Dataset Onion-Cell-Merged:** Recurso externo (Roboflow Universe, 2023) con 644 imágenes anotadas que mejora la generalización.
- **Dataset INA:** 392 micrografías del Instituto Nacional del Agua (INA). Tras un curado supervisado, se seleccionan 145 aptas (Figura 3), 58 con anotaciones.
- **Dataset Roboflow:** 457 imágenes de repositorios públicos, re-etiquetadas bajo criterios unificados.

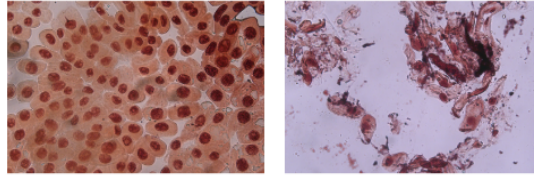


Figura 3: Comparación entre una imagen apta para el ensayo (izquierda) y una imagen no apta (derecha).

Con el objetivo de integrar estas fuentes bajo un esquema de anotación común, se implementa una estrategia de etiquetado iterativo asistida por modelos. Inicialmente, se ajusta YOLOv11 utilizando el *dataset* Onion-Cell-Merged. Posteriormente, este modelo genera anotaciones automáticas preliminares sobre *Roboflow* e INA (INA etiquetado y *Roboflow* etiquetado), cuyo flujo se resume en la Figura 4.

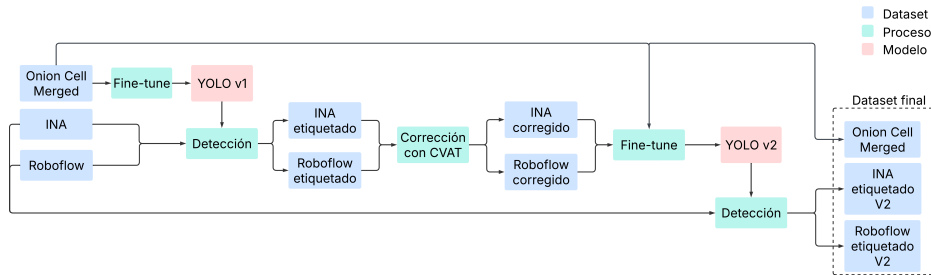


Figura 4: Flujo del proceso iterativo de etiquetado asistido por modelo. El conjunto Onion-Cell-Merged entrena un YOLOv11 inicial, cuyas predicciones sobre los conjuntos Roboflow e INA son luego validadas y corregidas manualmente en Computer Vision Annotation Tool (CVAT) (CVAT.ai Corporation, 2023) para obtener el *dataset* consolidado final.

Las predicciones preliminares se validan y corrigen manualmente en CVAT, obteniéndose los conjuntos INA corregido y *Roboflow* corregido. Con este conjunto curado

se realiza un *fine-tuning* final de YOLOv11, consolidando el *dataset* unificado detallado en la Tabla 1. A partir de estas detecciones finales, se extraen los recortes individuales para conformar el conjunto de entrenamiento del clasificador.

Tabla 1: Distribución de imágenes e instancias celulares para el entrenamiento del detector.

Subconjunto	Cantidad de Imágenes	Cantidad de Células
Entrenamiento	867	31.662
Validación	218	7.760
Evaluación	156	5.546
Total	1.241	44.968

Con el propósito de mitigar el desbalance de clases y robustecer el modelo, se implementa una estrategia de aumento de datos (*data augmentation*) que duplica el volumen de las muestras originales. Se aplican perturbaciones fotométricas aleatorias en brillo, contraste y saturación ($\pm 50\%$), junto con deformaciones elásticas (Simard et al. 2003), cuyas variaciones morfológicas se ilustran en la Figura 5. La distribución resultante de las instancias para la etapa de clasificación, integrando el aumento de datos y discriminada por clase biológica, se detalla en la Tabla 2.

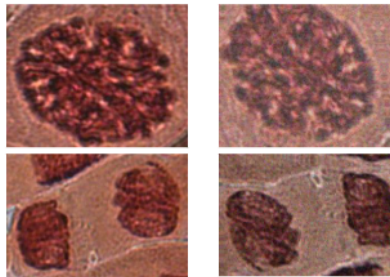


Figura 5: Comparación de imágenes de célula en reales (izquierda) y las mismas imágenes aumentadas (derecha).

Tabla 2: Distribución de recortes celulares (*crops*) para el entrenamiento del clasificador.

Clase	Entrenamiento	Validación	Evaluación	Total
Mitosis	6.162	540	282	6.984
No Mitosis	27.466	3.640	2.217	33.323
Total	33.628	4.180	2.499	40.307

2.2. Detección de Instancias Celulares (YOLOv11)

Para la detección de instancias celulares se utiliza YOLOv11 implementado con *Ultralytics v8.3.224* (Jocher et al. 2024). YOLO prioriza velocidad de inferencia y contexto global de la imagen. El modelo se entrena utilizando los cuadros delimitadores (*bounding boxes*) del *dataset* unificado de detección (véase sección 2.1), optimizándose para predecir las coordenadas x , y , w , h de cada núcleo celular. Los recortes de las regiones detectadas alimentan la etapa subsiguiente de clasificación.

2.3. Clasificación de instancias celulares

Para la determinación del estado mitótico, es implementada una arquitectura de clasificación unificada. Se compara el desempeño de cuatro modelos base (*backbones*) pre-entrenados en ImageNet: ResNet50 (He et al. 2016), VGG19 (Simonyan & Zisserman, 2015), EfficientNetB1 y EfficientNetB2 (Tan & Le, 2019). La topología final, común a todas las pruebas, sustituye el cabezal original por una estructura especializada que se muestra en la Figura ??.

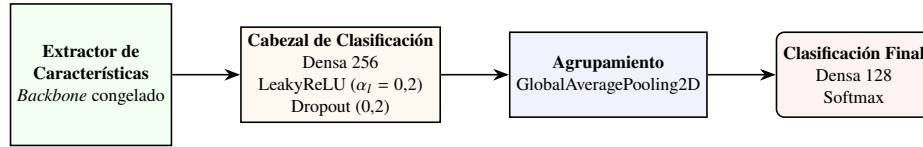


Figura 6: Arquitectura del modelo de clasificación: *backbone* congelado, cabezal denso con regularización, *GlobalAveragePooling2D* y salida *softmax* binaria (Mitosis vs. Interfase).

Se aplica una estrategia de transferencia de conocimiento (*transfer learning*) y ajuste fino (*fine-tuning*), donde el *backbone* permanece congelado con excepción de su última capa convolucional. El proceso de optimización es realizado mediante el algoritmo Adam (Kingma & Ba, 2015) con una tasa de aprendizaje de 10^{-5} y función de pérdida de entropía cruzada categórica.

El entrenamiento fue realizado sobre una estación de trabajo con GPU NVIDIA RTX 3090 (24 GB VRAM). YOLOv11 fue entrenado durante 200 épocas con *batch size* de 16 (14 min); EfficientNetB2 durante 20 épocas con *batch size* de 64 (28 min). Trabajando en CPU se procesa una imagen de 3072 x 2048px en 1,16 segundos, con un velocidad de 15 células/s. En GPU se obtiene una velocidad de 0.76 s por imagen y 197' células/s.

2.4. Calibración de modelos

Para corregir la sobreconfianza típica de las redes neuronales (Guo et al. 2017), se cuantifica la descalibración mediante el Error de Calibración Esperado (ECE). Esta métrica particiona el espacio de probabilidad en M intervalos (*bins*) y calcula el promedio de la desviación absoluta entre exactitud ($acc(B_m)$) y confianza media ($conf(B_m)$) (Ecuación 1):

$$ECE = \sum_{m=1}^M \frac{|B_m|}{n} |acc(B_m) - conf(B_m)| \quad (1)$$

Esta expresión adopta diferentes parámetros según su aplicación: en tareas de clasificación, n representa el total de muestras; en tareas de detección, la métrica es denominada Error de Calibración Esperado de Detección (D-ECE), donde n equivale al número total de detecciones y la exactitud es determinada por un criterio de superposición espacial.

2.4.1. Calibración del detector

Para la detección de clase única, la calibración se modela como clasificación binaria (Kuzucu et al. 2024). Una predicción se considera Verdadero Positivo ($Y = 1$) si su Intersección sobre Unión (IoU) con el *ground truth* supera $\sigma = 0,8$. El umbral de detección se reduce a puntajes de confanza $s = 0,1$ para modelar la incertidumbre en todo el rango de confianza. Dada la naturaleza no paramétrica de la distribución de confianza en el detector, se aplica regresión isotónica para ajustar una función no decreciente f que minimice el error cuadrático medio (Ecuación 2).

$$f = \operatorname{argmin}_f \sum_{i=1}^N (y_i - f(s_i))^2 \quad \text{sujeto a} \quad s_i \leq s_j \Rightarrow f(s_i) \leq f(s_j) \quad (2)$$

2.4.2. Calibración del clasificador

El análisis preliminar del clasificador revela un perfil de descalibración heterogéneo que depende de la clase biológica, evidenciando subconfianza en la clase Mitosis y sobreconfianza en Interfase (Figura 7). Esta asimetría invalida el uso de un escalado de temperatura global y motiva la implementación del Escalado de Temperatura Específico por Clase (CSTS) (Frenkel & Goldberger, 2021).

Mediante el método CSTS, se introduce un vector de temperaturas $T \in \mathbb{R}^K$ que ajusta los *logits* z_k de forma independiente para cada categoría. La probabilidad calibrada $\hat{q}_i$ y la optimización de T mediante la *Negative Log-Likelihood* (NLL) se definen conjuntamente en la Ecuación 3:

$$\hat{q}_i = \max_k \sigma_{SM} \left(\frac{z_k}{T_k} \right), \quad T = \operatorname{argmin}_T \left(- \sum_{i=1}^N \log(\hat{q}_i(y_i|z_i, T)) \right) \quad (3)$$

2.5. Validación estadística y estimación de incertidumbre

Para asegurar la validez del test de *Allium cepa*, se requiere analizar un mínimo de 3000 células por tratamiento (Bonciu et al. 2018). El sistema modela el IM como una variable aleatoria para cuantificar su incertidumbre mediante modelos probabilísticos.

Cada detección i se modela como una variable de Bernoulli con probabilidad p_i (confianza calibrada del detector). La población celular esperada (N_{cel}) y su varianza (σ_{cel}^2) se calculan según la Ecuación 4:

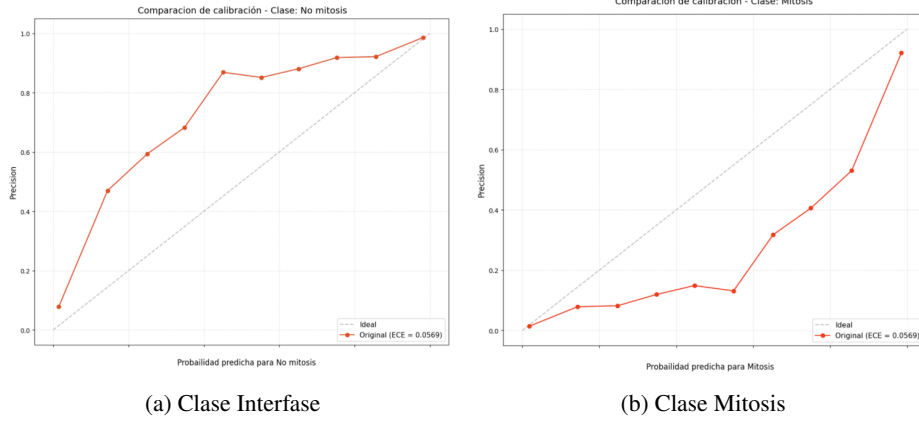


Figura 7: Diagramas de fiabilidad por clase en el subconjunto de validación, evidenciando la necesidad de una corrección diferenciada.

$$N_{cel} = \sum_i p_i, \quad \sigma_{cel}^2 = \sum_i p_i(1 - p_i) \quad (4)$$

Para la clasificación, sea $q_{j,k}$ la probabilidad calibrada de que la célula j pertenezca a la clase k , su conteo N_k y varianza σ_k^2 se estiman según la Ecuación 5.

$$N_k = \sum_j q_{j,k}; \quad \sigma_k^2 = \sum_j q_{j,k}(1 - q_{j,k}) \quad (5)$$

Dado que $N > 100$ y por el Teorema del Límite Central (Wasserman, 2004), el resultado del conteo (N_{cel}) se expresa mediante un intervalo de confianza del 95 % (Ecuación 6):

$$N_{cel} = N \pm 2\sigma \quad (6)$$

El IM se define como la proporción $\frac{N_{mit}}{N_{cel}}$. Al ser variables aleatorias donde las mitosis son un subconjunto estricto del total, presentan alta covarianza ($Cov(N_{mit}, N_{cel}) \approx \sigma_{mit}^2$). Aplicando propagación de incertidumbre en cocientes, la varianza del IM (σ_{IM}^2) se aproxima en la Ecuación 7:

$$\sigma_{IM}^2 \approx IM^2 \left[\frac{\sigma_{mit}^2}{N_{mit}^2} + \frac{\sigma_{cel}^2}{N_{cel}^2} - \frac{2\sigma_{mit}^2}{N_{mit}N_{cel}} \right] \quad (7)$$

Finalmente, dada la magnitud muestral ($N_{cel} \geq 3000$), el IM se reporta con un intervalo de confianza simétrico del 95,45 % (Ecuación 8).

$$IM_{reportado} = IM \pm 2 \cdot \sigma_{IM} \quad (8)$$

Este enfoque probabilístico permite que el sistema entregue junto con el resultado

de su evaluación un rango estadístico que absorbe las discrepancias inherentes a la variabilidad visual de las muestras microscópicas.

3. Resultados

3.1. Resultados de la evaluación del modelo de detección

El desempeño del detector se analiza sobre el conjunto de prueba mediante la curva Precisión-Recall (Figura 8) (Jocher et al. 2024).

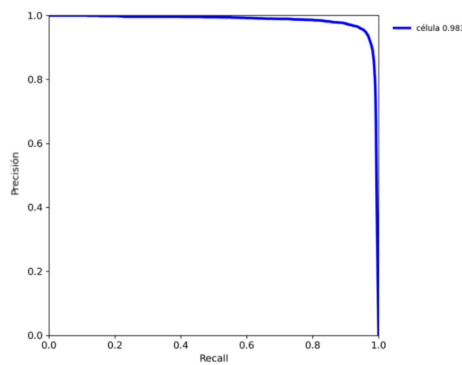


Figura 8: Curva Precisión-Recall del entrenamiento final de YOLOv11. Se observa un valor de 0,983 en el área bajo la curva (Area Under the Curve (AUC)), indicando una alta robustez en la detección.

El área bajo la curva (AUC) de 0,983 y el *recall* del 98 % evidencian una localización precisa de las instancias celulares con una tasa marginal de falsos positivos y apenas un 2 % de omisiones.

3.1.1. Resultados de calibración de YOLO

El impacto de la calibración se evalúa mediante el D-ECE. El modelo YOLOv11 original exhibe un error del 8,51 % con marcada tendencia a la subconfianza. La regresión isotónica alinea la curva de fiabilidad con la diagonal ideal (Figura 9), reduciendo el D-ECE a valores marginales y garantizando estimaciones precisas de verdaderos positivos.

3.2. Rendimiento del Clasificador y Cálculo del Índice Mitótico

La selección de la arquitectura final es fundamentada en una evaluación comparativa sobre el subconjunto de datos de prueba. Los resultados cuantitativos, detallados en la Tabla 3, permiten analizar el compromiso entre precisión y sensibilidad para la identificación de la fase mitótica.

EfficientNetB2 es consolidada como la arquitectura óptima al alcanzar la mayor exactitud global (95 %) y un equilibrio superior en la clase minoritaria (Mitosis), con

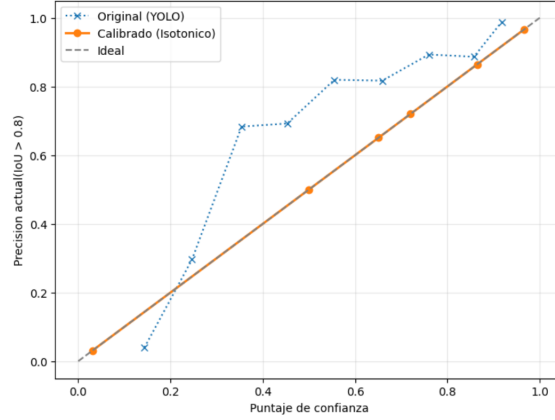


Figura 9: Diagrama de fiabilidad para el modelo YOLOv11. La línea naranja ilustra el alineamiento de la curva post-calibración con la diagonal ideal tras la aplicación de la regresión isotónica.

Tabla 3: Resultados comparativos de los modelos evaluados sobre el conjunto de prueba.

Modelo	Exactitud (%)	Recall (%)		Precisión (%)	
		Mitosis	Interfase	Mitosis	Interfase
EfficientNetB2	95	77	97	77	97
EfficientNetB1	94	81	96	70	98
VGG19	90	75	92	56	97
ResNet50	93	70	96	67	96

una precisión del 77 %. El clasificador final demuestra un elevado desempeño con un F1-Score de 0,97 para la clase Interfase y de 0,75 para Mitosis. Estos valores aseguran una discriminación de la clase mayoritaria y capacidad para identificar células en división, sentando las bases para una estimación fiable del IM.

3.2.1. Resultado de calibración del modelo clasificador

La aplicación de CSTS reduce el ECE del clasificador de 4,58 % a un despreciable 2,28 %. Como muestra la Figura 10, el modelo calibrado corrige la subconfianza en la clase Mitosis y la sobreconfianza en Interfase, legitimando el uso de sus probabilidades para el cálculo de incertidumbre.

3.3. Presentación del resultado final

El procedimiento finaliza mediante la integración de la detección calibrada de YOLOv11 y la clasificación de EfficientNetB2 para la generación del reporte biológico. Se realiza un conteo de las estructuras celulares y se estima la población total (N_{total})

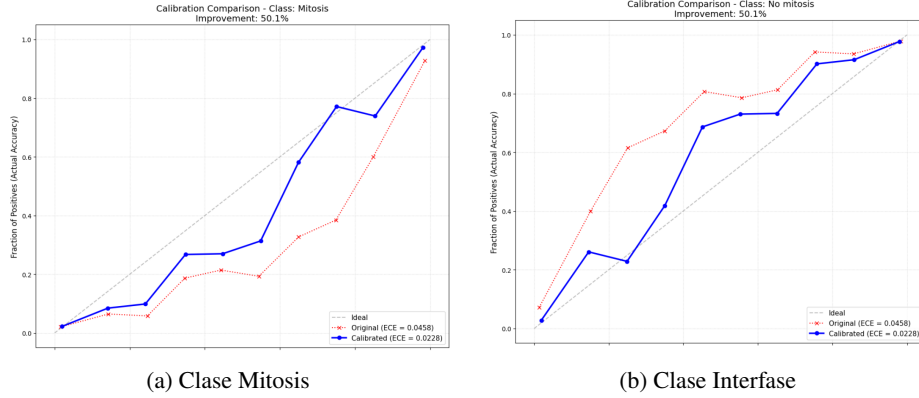


Figura 10: Diagramas de fiabilidad por clase previos y posteriores a la calibración. Se ilustra la rectificación de las curvas de confianza tras la aplicación de CSTS.

con su respectivo intervalo de confianza al 95 % ($IC_{95\%}$), considerando el error de calibración del detector.

La validez estadística del estudio es validada siguiendo los criterios estandarizados para el bioensayo de *Allium cepa*, los cuales exigen que el límite inferior del intervalo de confianza cumpla con la condición $L_{inf} \geq 3000$. En caso de no alcanzarse dicho umbral, el IM es calculado e informado junto con una advertencia de falta de significancia estadística.

Posteriormente, el clasificador diferencia las células detectadas en mitosis (N_{mit}) e interfase (N_{nmit}), reportándose cada conteo con su incertidumbre asociada. Los resultados obtenidos sobre el conjunto de evaluación final son contrastados con las anotaciones de expertos en la Tabla 4. Es observado que el valor de referencia obtenido por los expertos se encuentra contenido dentro del intervalo de confianza generado por el modelo.

Tabla 4: Comparación de resultados entre la metodología automática propuesta y el conteo experto manual sobre el conjunto de evaluación.

Métrica Biológica	Pipeline Automático (IC 95 %)	Conteo Experto
Total Células Detectadas	432 ± 12	428
Células en Mitosis (N_{mit})	69 ± 12	71
Células en Interfase (N_{nmit})	363 ± 12	357
Índice Mitótico (IM)	$18,65\% \pm 2,7\%$	16,59 %

4. Discusión y Conclusiones

A pesar de ser un método ampliamente utilizado, no se encontró trabajo publicado que aborde la segmentación automática y clasificación de células de *Allium cepa* en cuanto su fase mitótica. Las revisiones vigentes del modelo describen explícitamente que el análisis de las preparaciones citogenéticas se realiza mediante observación microscópica y conteo manual, identificando visualmente las fases mitóticas y las aberraciones cromosómicas sin recurrir a ningún método de segmentación automática (Leme & Marin-Morales, 2009; Nicuță et al. 2025). Si bien existen algoritmos automatizados de detección y clasificación de mitosis basados en aprendizaje profundo, estos se aplican exclusivamente a células humanas (en particular, en conjuntos de datos de histopatología oncológica (Nateghi & Tizhoosh, 2024; Saha et al. 2018)) y no son directamente transferibles al dominio de *Allium cepa*. En este contexto, el presente trabajo se propone como el primero en automatizar dicho método, aportando además el conjunto de datos generado como referencia para futuras investigaciones sobre mitosis celular en plantas superiores.

La incorporación de protocolos de calibración probabilística representa un avance significativo sobre los sistemas de visión artificial convencionales. La regresión isotónica en el detector y el escalado CSTS en el clasificador corrigen las distorsiones de confianza permitiendo transitar de una clasificación determinista a una estimación basada en intervalos de confianza, proporcionando un nivel de transparencia estadística esencial para la validación de ensayos ecotoxicológicos.

El desbalance estructural de la clase Mitosis (~13 % del total de instancias) constituye la principal limitación del sistema. La problemática se concentra en fases limítrofes (profase temprana vs. interfase tardía) donde la diferenciación morfológica es ambigua incluso para expertos entrenados. Si bien las métricas reportadas cubren las tres etapas principales del problema (detección, clasificación y calibración) con el objetivo de estimar de forma robusta el índice mitótico, evaluaciones futuras podrían complementarse con Intersection over Union (IoU) en detección, y *balanced accuracy* o AUC-PR en clasificación, según el aspecto del desempeño que se desee enfatizar.

Se concluye que el pipeline desarrollado exhibe una elevada concordancia con los resultados del análisis experto manual, reduciendo los tiempos de procesamiento y eliminando la subjetividad inter-observador. La robustez del sistema, validada mediante el cumplimiento de 3000 células, posiciona a esta herramienta como una solución escalable y reproducible para el biomonitoreo de la calidad del agua. Al superar las barreras operativas de este bioensayo, el presente desarrollo hace viable la exigencia de parámetros de genotoxicidad y toxicidad emergente en los marcos regulatorios, sentando las bases para una legislación ambiental más abarcativa y estricta. Investigaciones futuras buscarán integrar modelos generativos que permitirán el aumento de datos orientado a clasificar las células según su fase con el objetivo de calcular el índice por fase y mejorar los puntajes de clasificación del modelo.

Aclaraciones

Durante la preparación de este trabajo, el/los autor(es) utilizaron Gemini y Claude con el fin de mejorar el lenguaje y la legibilidad del manuscrito. Tras utilizar esta herramienta/servicio, el/los autor(es) revisaron y editaron el contenido según fuera necesario y asumen plena responsabilidad por el contenido de la publicación. Además, los autores no tienen ningún conflicto de intereses que declarar que sea relevante para el contenido de este artículo.

Reconocimientos

Este trabajo fue desarrollado en el marco de la colaboración entre el Grupo de Inteligencia Artificial y Robótica de la Universidad Tecnológica Nacional, Facultad Regional Buenos Aires y el Laboratorio Experimental de Tecnologías Sustentables del Instituto Nacional del Agua.

Referencias

- Bonciu, E., et al. (2018). An evaluation for the standardization of the Allium cepa test as cytotoxicity and genotoxicity assay. *Caryologia*, 71(3), 191-209.
- Bormioli, N., Hanela, S., Reale, M., El Kassis, Y., Davico, M., Duek, A., Comellas, E., Tagliavini, D., & Gómez, C. (2019). Análisis de la normativa aplicable al control de vertidos de establecimientos industriales en ocho jurisdicciones de Argentina. *IV Congreso Nacional de Ciencia y Tecnología Ambiental*.
- Bormioli, N., Reale, M., Hanela, S., de los Ángeles Davico, M., El Kassis, Y., Tagliavini, D., Duek, A., Comellas, E., & Gómez, C. (2018). Estudio comparativo de límites de vertido para efluentes industriales en cinco regiones de la República Argentina.
- CVAT.ai Corporation. (2023). *Computer Vision Annotation Tool (CVAT)*. <https://github.com/cvat-ai/cvat>
- Flores, M., & Yamaguchi, M. (2008). Teste do micronúcleo: uma triagem para avaliação genotóxica. *Saúde e Pesquisa*, 1(3), 337-340.
- Frenkel, L., & Goldberger, J. (2021). Network Calibration by Class-based Temperature Scaling. *arXiv preprint arXiv:2110.01022*.
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 70, 1321-1330.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- Jocher, G., Chaurasia, A., & Qiu, J. (2024). *Ultralytics YOLOv11* (Ver. 8.3.224). <https://github.com/ultralytics/ultralytics>
- Khanam, R., & Hussain, M. (2024). YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv preprint arXiv:2410.17725*.
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*.

- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4015-4026.
- Kuzucu, C., et al. (2024). A Comparative Study on Calibration of Object Detectors. *arXiv preprint arXiv:2405.20459*.
- Leme, D. M., & Marin-Morales, M. A. (2009). Allium cepa test in environmental monitoring: A review on its application. *Mutation Research/Reviews in Mutation Research*, 682(1), 71-81. <https://doi.org/10.1016/j.mrrev.2009.06.002>
- Mareque, L., Taurozzi, N., Cibeira, J., Santo Domingo, A., & Rodríguez Colmeiro, R. G. (2025). Segmentación automática de células de Allium cepa por métodos no supervisados de visión artificial. *Proyecciones*, 23(2).
- Nateghi, R., & Tizhoosh, H. R. (2024). Toward Interpretable and Generalized Mitosis Detection in Digital Pathology Using Deep Learning. *Scientific Reports*, 14, 11560. <https://doi.org/10.1038/s41598-024-62360-y>
- Nicuță, D., Grosu, L., Patriciu, O.-I., Voicu, R.-E., & Alexa, I.-C. (2025). The *Allium cepa* Model: A Review of Its Application as a Cytogenetic Tool for Evaluating the Biosafety Potential of Plant Extracts. *Methods and Protocols*, 8(4), 88. <https://doi.org/10.3390/mps8040088>
- Patra, M., Bhowmik, N., Bandopadhyay, B., & Sharma, A. (2004). Comparison of mercury, lead and arsenic with respect to genotoxic effects on plant systems and the development of genetic tolerance. *Environmental and Experimental Botany*, 52(3), 199-223.
- Roboflow Universe. (2023). Onion Cell Merged Dataset.
- Saha, M., Chakraborty, C., & Racoceanu, D. (2018). Efficient Deep Learning Model for Mitosis Detection Using Breast Histopathology Images. *Knowledge-Based Systems*, 136, 119-128. <https://doi.org/10.1016/j.knosys.2017.09.005>
- Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. *Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR)*, 958-963.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations (ICLR)*.
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 6105-6114.
- Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer.
- Zhou, S., Duan, C., Wang, X., Wang, M., Yu, Z., & Fu, H. (2008). Assessing cypermethrin-contaminated soil with three different earthworm test methods. *Journal of Environmental Sciences*, 20(11), 1381-1385.