



UNIVERSIDAD NACIONAL AUTONOMA DE MÉXICO

CENTRO DE NANOCIENCIAS Y NANOTECNOLOGÍA

*DETECCIÓN AUTOMÁTICA DE ABERRACIONES GENÉTICAS
DERIVADAS DEL ENSAYO DE MICRONÚCLEOS EN LINFOCITOS DE
SANGRE PERIFÉRICA HUMANA PARA TINCIIONES NO FLUORESCENTES*

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
LICENCIADO EN NANOTECNOLOGÍA

P R E S E N T A:

José Antonio Rodríguez Arenas

TUTOR DE TESIS

Dr. Juan Carlos García Ramos

Dra. Yanis Toledano Magaña

Ensenada Baja California a diciembre 2020





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
CENTRO DE NANOCIENCIAS Y NANOTECNOLOGÍA



LICENCIATURA EN NANOTECNOLOGÍA
NANOCATÁLISIS

DETECCIÓN AUTOMÁTICA DE ABERRACIONES GENÉTICAS DERIVADAS DEL ENSAYO DE
MICRONÚCLEOS EN LINFOCITOS DE SANGRE PERIFÉRICA HUMANA PARA TINCIONES NO
FLUORESCENTES

TESIS
QUE PARA OPTAR POR EL TÍTULO DE:
LICENCIADO EN NANOTECNOLOGÍA

PRESENTA:
José Antonio Rodríguez Arenas

DIRECTOR DE TESIS
Juan Carlos García Ramos
Yanis Toledano Magaña

ENSENADA, BAJA CALIFORNIA DICIEMBRE 2020

Hago constar que el trabajo que presento es de mi autoría y que todas las ideas, citas textuales, datos, ilustraciones, gráficas, etc. Sacados de cualquier obra o debidas al trabajo de terceros, han sido debidamente identificados y citados en el cuerpo del texto y en la bibliografía y acepto que en caso de no respetar lo anterior puedo ser sujeto de sanciones universitarias.

Afirmo que el material presentado no se encuentra protegido por derechos de autor y me hago responsable de cualquier reclamo relacionado con la violación de derechos de autor.

José Antonio Rodríguez Arenas

ÍNDICE

	Página
Resumen	6
Introducción	7
Antecedentes	10
Justificación	14
Objetivos	15
General	
Particular	
Marco teórico	16
Ensayo de micronúcleos en linfocitos de sangre periférica humana	
Procesamiento de imágenes	16
Algoritmos usados en procesamiento de imágenes.	17
Clasificación de eventos	18
Sistemas basados en decisiones	18
Sistemas supervisados	19
Sistemas no supervisados	19
Análisis de imágenes de células	20
Metodología	22
Ambiente de programación	22
Características particulares de banco de imágenes.	22
Sistema de preprocesamiento	22
Sistema de segmentación y extracción de características	24
Generación de grupos principales de eventos por clasificación no supervisada.	25
Reclasificación de eventos identificados para corroboración.	25
Determinación de probabilidad de aberraciones genéticas de MN, puentes y yemas.	26
Validación	27
Resultados	27
Clasificación de cuerpos detectados.	27

	Efectividad de detección directa en imágenes de entrada.	30
	Sistema	31
	Clasificación por probabilidad de presentación de cuerpos de interés.	32
Discusión	Criterio de elección de método de clasificación	34
	Clasificación de grupos de importancia y detección de cuerpos de interés	36
	Cualidades de cuerpos clasificados	38
	Comparativa con trabajos similares.	40
	Robustez	43
	Simplicidad	44
	Actualización respecto a acercamientos reportados	45
	Relevancia Actual	48
Conclusiones		50
Perspectivas		51
Referencias		52

A mi familia

Por su apoyo incondicional y esfuerzo.

Los amo

RESUMEN

El ensayo de micronúcleos es una técnica frecuente en laboratorios que realizan pruebas de genotoxicidad. El ensayo puede realizarse con tinciones fluorescentes y no fluorescentes, y puede ser asistido por contadores automáticos especializados. Sin embargo, en ensayos bajo tinciones no fluorescentes, estos sistemas no son considerados óptimos debido a la posibilidad de falsos positivos derivados de conteos de cuerpos no celulares teñidos, ya que los sistemas automáticos se basan en mediciones morfológicas básicas como área y forma. El presente trabajo da un acercamiento a un nuevo sistema de conteo celular para tinción no fluorescente, capaz de discriminar eventos celulares mono y binucleados de entre artefactos con alta similitud. Así como de identificar cuerpos micro nucleares correspondientes a micronúcleos (MN), puentes nucleares (PN) y yemas, en linfocitos aislados de sangre periférica humana, comúnmente usados en ensayo de micronúcleos.

INTRODUCCIÓN

La identificación y clasificación de morfologías celulares son métodos aplicables en técnicas de diagnóstico (Acevedo et al., 2019; Andrade et al., 2019; Gómez Flores et al., 2014; Q. Wang et al., 2016) y análisis (Andreini et al., 2020; Araújo et al., 2019; Delannoy et al., 2020; Kleczek et al., 2020; Sert et al., 2019; W. Wang et al., 2019) celular. Algunas de estas técnicas son aplicadas en la detección de enfermedades a partir de tejidos, o cuerpos celulares de interés. Un caso particular de evaluación basada en métodos visuales es la detección de cuerpos intracelulares, utilizado para determinar daños genotóxicos inducidos por algún agente externo (Kirsch-volders et al., 2018; Murphy et al., 2018).

El ensayo de micronúcleos en linfocitos aislados de sangre periférica humana (LSPH) es un ensayo de determinación de micronúcleos por bloqueo de la citocinesis (*cytokinesis-block micronucleus, CBMN*) y permite cuantificar el daño genotóxico provocado por un agente externo. Se realiza mediante la identificación y cuantificación de la frecuencia de cuerpos intracelulares inducidos por errores durante el proceso de división celular, que pueden generar cuerpos tales como, micronúcleos, puentes celulares y yemas. Estos pueden asociarse directamente a la interacción con el agente externo. Lo que vuelve a este ensayo importante en trabajos relacionados con la determinación de daño a nivel genético.

La metodología tradicional del ensayo de micronúcleos se basa en una tinción de LSPH con tinciones no específicas como Giemsa, Papanicolau, o acetato-orceína (Bolognesi et al., 2020), para poder realizar una cuantificación visual por microscopía óptica de campo claro, y detectar si presentan o no, cuerpos micro nucleares de interés (Sram et al., 2016). Se cuentan entre 1,000 y hasta más de 10,000 células individuales por cada condición de evaluación, lo que implica un conteo de más de 25,000 células, adicionalmente se deben contar los micronúcleos, yemas y puentes. Por esta razón, la evaluación manual por un usuario requiere de un consumo de tiempo elevado y puede tener asociada una baja reproducibilidad, lo que dependerá de la experiencia del evaluador (Roemer et al., 2015; Taghi et al., 2017). Sin embargo, en 2019 se reportó que en laboratorios con más amplia

experiencia en el conteo se realiza de forma manual, con poca variabilidad entre los conteos y sugiere que la variación se debe principalmente al donante (Lovell et al., 2019).

La aplicación para la detección de micronúcleos en cuerpos celulares por medio de técnicas automáticas ha sido poco estudiada de forma específica en los últimos 5 años, a pesar de ser un método que puede reducir alrededor de 70% del tiempo requerido (Roemer et al., 2015). Estudios recientes se han enfocado en la mejora de las condiciones metodológicas (Swee et al., 2021) específicamente, en cuanto al uso de sistemas automáticos se ha optado por sistemas comerciales que reportan una precisión experimental de entre 50 y 80% (Fenech et al., 2013) (Frieauff et al., 1998). La mayoría de los trabajos recientes de sistemas especializados de identificación automática tanto comerciales como personalizados, se basan principalmente en la cuantificación de los eventos a partir de su tinción con fluorescencia (Baeyens et al., 2011; Bolognesi et al., 2011; Roemer et al., 2015) y solo dos han sido reportado en los últimos 5 años para la identificación de micronúcleos en cuerpos celulares de LSPH para tinciones no fluorescentes con una precisión >70% (Hülber et al., 2020; Taghi et al., 2017). La elección de sistemas de técnicas por fluorescencia se debe a que en tinciones no fluorescentes se dificulta generar un contraste óptimo entre el citoplasma y el fondo del campo óptico, lo que puede provocar falsos positivos (Bolognesi et al., 2011) aparentemente con mayor frecuencia que en conteos a partir de tinciones fluorescentes (Bolognesi et al., 2020).

Por otra parte, la aplicación de técnicas automáticas requiere de un pre procesamiento manual para separar cuerpos de interés o de la corroboración manual por la baja precisión que presentan; esto requiere tanto software como hardware especializado y debido a que la mayoría se basa en aspectos morfológicos como área, forma y relación, la efectividad de detección de los sistemas frente a eventos celulares que no hayan sido previamente identificados no es clara. Además, los sistemas que existen actualmente necesitan condiciones muy

específicas de trabajo y en ningún caso se aborda la detección de cuerpos como puentes nucleares o yemas. Por lo que una exclusión manual individual previa y una validación manual posterior sigue siendo requerida.

El presente trabajo tiene por finalidad resolver un problema de identificación de cuerpos celulares en LSPH teñidos con colorante no fluorescente (Giemsa), para detectar con precisión la potencial presencia de micronúcleos de forma semiautomática, a partir de la extracción y aislamiento de cuerpos presentes en un banco de imágenes no procesado, de fotografías del campo de visión de un microscopio óptico convencional. El sistema presentado elimina la necesidad de una preselección individual de cuerpos celulares e identifica eventos distinguibles como cuerpos celulares mononucleados o binucleados respecto a artefactos no celulares con una alta similitud presentes en la imagen. Así mismo, se presenta un acercamiento con la intención de identificar aquellos casos donde se cuente con una alta probabilidad de presencia de micronúcleos, puentes nucleares o yemas, con la finalidad de limitar el conteo manual de eventos a sólo aquellos cuerpos celulares que requieran atención detallada.

ANTECEDENTES

En la literatura se identificaron varios sistemas enfocados en la identificación automática de micronúcleos en diferentes tipos de células (**Tabla 1**).

Tabla 1. Detección automática de micronúcleos en cuerpos celulares. Trabajos encontrados en la literatura que abarcan especialmente la identificación de micronúcleos en cuerpos celulares, principalmente de LSPH. Se muestran en orden de publicación siendo el más reciente de 2020.

Año	Aplicación	Descripción de interés
2020 (Hülber et al., 2020)	Generación de sistema de detección y conteo de MN en tinción no fluorescentes*.	Generación de software + hardware con incertidumbre de 20-38% en el sistema automático.
2017 (Taghi et al., 2017)	Detección automática de MN en células de sangre periférica mediante procesamiento de imágenes microscópicas*.	Detección de MN mediante filtrado con plantilla para separar núcleos de MN en imágenes RGB de muestras de microscopio en campo claro. Precisión de 96% para conteo de MN en células binucleadas en cuerpos manualmente seleccionados. No menciona detección de puentes nucleares o de yemas.
2010 (Bolognesi et al., 2011)	Prueba de MN para biodosimetría*.	Validación de conteo automático de sistema comercial basado en detección por fluorescencia. Correlación de 0.705 entre conteo manual y conteo automático de MN presentes en células binucleadas. No menciona detección de puentes nucleares o de yemas.

1998 (Frieauff et al., 1998)	Análisis automático de ensayo <i>in vitro</i> de MN en células V79.	Detección de MN en células V79 por medio del procesamiento de canal G de una cámara convencional. Donde no se hace discernimiento de presencia de artefactos.
---------------------------------------	---	---

*Trabajos en LSPH aislados.

A pesar de que el ensayo de micronúcleos por bloqueo de la citocinesis es el recomendado por la OCDE para la evaluación de genotoxicidad, es poco utilizado probablemente por el tiempo requerido para la observación y análisis. Lo que puede ser la razón de que existan pocos sistemas automáticos para este ensayo. Una revisión realizada a finales de 2015 reporta que de un total de 540 trabajos donde se emplea la técnica de micronúcleos por bloqueo de la citocinesis, solo en 34 se utilizaron linfocitos humanos (Sram et al., 2016). Situación que se repite en revisiones de estudios de genotoxicidad en pacientes con cáncer de cabeza y cuello, donde sólo 5 de 19 son estudiados en linfocitos (Bolognesi et al., 2020).

Un factor importante para el conteo automático de MN por análisis de imagen, son los criterios de inclusión o exclusión. En la literatura se describen algunos de los criterios más comúnmente utilizados para detectar cuerpos binucleados y MN (Tabla 2).

Tabla 2. Criterios comunes de identificación morfológica

Criterio	Descripción	Usado en
Relación de aspecto	$\frac{DíametroMáximo}{DíametroMínimo}$	(Bolognesi et al., 2011; Taghi et al., 2017)
Limitante de asimetría entre diferentes áreas	Determina la relación entre diámetro, área y perímetro.	(Bolognesi et al., 2011)
Área máxima	Número de píxeles extraídos por cada evento	(Bolognesi et al., 2011)

Identificación por plantilla	Identifica eventos definidos como MN y en un segundo proceso los definidos como células binucleadas utilizando una plantilla predefinida	(Bolognesi et al., 2011)
Solidez	$\frac{Area}{Convex Area}$	(Taghi et al., 2017)
Extensión	$\frac{Area total}{\acute{a}rea de caja contenedora}$	(Taghi et al., 2017)
Límites mínimos y máximos de inclusión	Definen el intervalo de pixeles a utilizar	(Bolognesi et al., 2011)
Uso de plantillas predefinidas	Determinación de semejanza de cuerpos respecto a un cuerpo predefinido	(Taghi et al., 2017)
Excentricidad	$\frac{Foci}{longitud de eje mayor}$	(Taghi et al., 2017)
Convexidad	$\frac{P\acute{e}rimetro Convexo}{Per\acute{ı}metro}$	(Taghi et al., 2017)
Redondez	$\frac{4 * Area}{\pi * Di\acute{a}metroMax^2}$	(Taghi et al., 2017)
Compacidad	$\frac{\sqrt{\left(\frac{4}{\pi}\right) Area}}{Di\acute{a}metro m\acute{a}ximo}$	(Taghi et al., 2017)
Factor de forma	Factor personalizado	(Taghi et al., 2017)

Los criterios para considerar cuerpos positivos pueden no ser los óptimos. En el caso de C. Bolognesi et al. (Bolognesi et al., 2011), se basan en 5 únicos aspectos para detectar células binucleadas, mientras que en el caso de M. Taghi et al. (Taghi et al., 2017), utilizan 9 y resuelven la identificación de MN a partir de la extracción de los núcleos principales mediante una plantilla.

Un criterio de exclusión manual fue necesario en los casos de C. Bolognesi et al. (Bolognesi et al., 2011), para rechazar células inadecuadas y mejorar el conteo final. Y en el caso de M. Taghi et al. no queda claro el estudio de casos donde cuerpos de similares características son evaluados o excluidos (Taghi et al., 2017).

JUSTIFICACIÓN

El ensayo de determinación de micronúcleos por bloqueo de la citocinesis en LSPH permite obtener información sobre el daño genotóxico que un agente externo puede causar a las células y por ende a un organismo más complejo. Sin embargo, el tiempo necesario para la identificación de MN, puentes y yemas; el alto costo de sistemas automatizados de conteo y del equipamiento necesario para la obtención de imágenes que puedan ser analizadas, dificulta su aplicación y uso en laboratorios de investigación universitarios. Además, la herramienta de asistencia manual para el análisis de ensayos celulares reporta posibles inconsistencias en la identificación de micronúcleos.

En este sentido, el crecimiento en las áreas involucradas en el procesamiento de imágenes dentro del área médico-biológicas generadas en los últimos dos años muestran que, es posible generar sistemas automáticos para la determinación de MN en LSPH. En la literatura se identificaron 2 trabajos directamente relacionados con la asistencia en el procesamiento automático de imágenes de linfocitos de sangre periférica humana para el ensayo de micronúcleos, pero sin detección de puentes nucleares o yemas.

Todo lo anterior pone en evidencia la necesidad de generar un sistema especializado de fácil uso para que asista al evaluador en la determinación de MN en LSPH. Esto permitiría reducir el tiempo necesario para determinar el efecto genotóxico de agentes externos utilizando el modelo recomendado por la OCDE.

El presente trabajo pretende contribuir a la generación de un sistema de análisis de imágenes de ensayos de determinación de micronúcleos por bloqueo de la citocinesis en LSPH, que sea capaz de determinar y cuantificar la presencia o ausencia de células binucleadas, micronúcleos, puentes y yemas. Adicionalmente, se espera que el sistema pueda realizar la identificación en imágenes de diferentes resoluciones y que cuente con la precisión y reproducibilidad necesarias.

OBJETIVOS

Objetivo General:

Generar un sistema de evaluación automática de imágenes del ensayo de determinación de micronúcleos por bloqueo de la citocinesis con colorantes no fluorescentes, obtenidas de campos de visión de microscopios de campo claro.

Objetivos específicos:

- Desarrollar un algoritmo base capaz de clasificar cuerpos de forma no supervisada, extraídos de imágenes principales.
- Generar un banco de imágenes etiquetadas para implementar un sistema de clasificación supervisada basada en imágenes.
- Determinar un algoritmo óptimo para evaluar la presencia de micronúcleos, puentes celulares y yemas, en cuerpos celulares.
- Generar una red neuronal capaz de identificar cuerpos celulares mononucleados y binucleados bajo condiciones mínimas de captura y alta presencia de artefactos no celulares.
- Validar la predicción de la red neuronal implementada en el proceso de clasificación de imágenes.

MARCO TEÓRICO

Ensayo de micronúcleos en linfocitos de sangre periférica humana

Los linfocitos son una especie celular presente en la sangre, son relativamente fáciles de cultivar en suspensión y se utilizan ampliamente como biomarcadores para determinar el potencial daño genotóxico provocado por agentes externos (Lovell et al., 2019). En particular, el estudio de micronúcleos en LSPH se realiza mediante el bloqueo o no de citocinesis utilizando un inhibidor de polimerización de actina como la citocalasina B, la cual se puede añadir antes del proceso mitótico con la finalidad de evaluar células binucleadas. En este caso, las células se cultivan durante 24h en presencia del agente externo cuyo potencial genotóxico se desea evaluar y se procede al conteo de micronúcleos presentes en las células. (OCDE, 2016)

El conteo puede ser manual o automático. El conteo automático de los cuerpos micro nucleares ha sido de interés comercial en diversas épocas y ha generado algunos sistemas de conteo automático. Para un conteo automático por medio de sistemas comerciales o por sistemas especializados se evalúa directamente sobre la imagen del campo utilizando un sistema acoplado al microscopio, o una cámara directamente acoplada al microscopio, que excluye cuerpos que no cumplan con los criterios de decisión establecidos por el programador. Este tipo de sistemas, por lo tanto, presentan problemas de detección en casos donde los criterios de exclusión no son lo suficientemente robustos para lograr discernir entre cuerpos celulares y artefactos de características similares.

Procesamiento de imágenes

El procesamiento de imágenes se basa en la extracción de características de información, mientras que la separación de cuerpos utiliza valores umbrales (en inglés, thresholding), o de separación de cuerpos de interés por detección de ejes, texturas o contornos. En casos donde las características de la imagen presenten demasiadas irregularidades, derivadas de contrastes poco definidos, presencia de ruido en la imagen, baja resolución de los cuerpos estudiados, entre otros, la

extracción de información puede ser mejorada directamente en los canales de información específicos o parcialmente por medios de filtros de imagen.

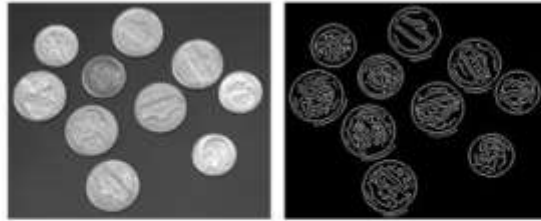
Los filtros de imagen consisten en la homologación de información presente en una vecindad de píxeles, que es capturada por la ventana del filtro, procesada por ella y el resultado reasignado a la posición del píxel. Algunos de los filtros más ampliamente utilizados en diferentes trabajos de segmentación de imágenes son, los gaussianos, de correcciones de brillo y contraste como filtros de transformación "top-bottom hat", o de desvanecimiento ("Smoothing") de los píxeles.

La segmentación de los cuerpos es extraída en capas de información, y es representada en "imágenes binarias", de forma tal que los cuerpos de interés son representados por secciones de píxeles con un valor de intensidad de 1, mientras que toda la información que no se considera de utilidad es representada por píxeles de valor 0. Una técnica de segmentación automática es la técnica por separación basada en vecinos cercanos, "*kmeans*". Esta segmentación separa la información de píxeles de composición similar y puede ser utilizada en arreglos multidimensionales. Los que lo vuelve una buena técnica de separación de información en arreglos tridimensionales de información, y permite separar cuerpos relevantes a partir de la determinación de su color, en un total de $k= 1, 2, 3, 4, \dots, n$ secciones.

Algoritmos usados en procesamiento de imágenes.

Los conteos automáticos de software para el procesamiento de imágenes se basan en la información extraída por una combinación de filtros, separación por umbrales mínimos de información, o por sistemas de separación de cuerpos. Con la finalidad de extraer características como el número total de píxeles, la forma y textura de los cuerpos de interés en función de las imágenes, y/o la intensidad de los bordes y ejes detectados, ejemplo de ellos se muestran en la **Figura 1**.

a) Detección de bordes.



b) Separación por valores umbrales



c) Separación multinivel de valores umbrales

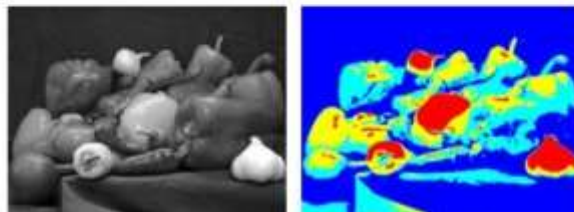


Figura 1. Diferentes sistemas de detección de propiedades presentes en imágenes. (imágenes tomadas de la página: <https://www.mathworks.com/help/images/ref/imsegkmeans.html>)

Clasificación de eventos

La clasificación de eventos de interés puede realizarse a través de criterios establecidos, sistemas no supervisados de clasificación automática, o sistemas supervisados de clasificación automática.

Sistemas basados en decisiones

Los sistemas basados en criterios establecidos se basan en la clasificación a partir de parámetros específicos. De forma tal que, cualquier evento que cumpla con las características evaluadas, será clasificado dentro de un mismo grupo, y en caso de no cumplirlas, dependiendo de la cantidad de decisiones y de la robustez

de ellas, será clasificado en un grupo independiente, o será descartado. Este tipo de arquitecturas se basan principalmente en la programación de las decisiones de forma directa por el programador. Lo que provoca una labor tediosa en casos donde se requiera evaluar múltiples eventos, donde no se tengan definidos los criterios de inclusión y exclusión, o donde la variabilidad de los cuerpos es tal, que puedan ser agrupados con facilidad en grupos erróneos.

Sistemas supervisados

Los sistemas supervisados dependen de una interacción directa con el usuario para generar y validar su sistema de clasificación. De forma general, los pasos de la clasificación de imágenes se encargan de realizar: Correcciones de iluminación, brillo, contraste y limpieza de ruido de la imagen, durante el preprocesamiento; segmentación de objetos y extracción de características de interés, durante el procesamiento y; clasificación por algoritmos discriminatorios o mediante el entrenamiento, validación y testeo de modelos matemáticos generados por técnicas de aprendizaje automático.

Sistemas no supervisados

Los sistemas de clasificación no supervisada pretenden identificar un número específico de grupos de interés, que compartan similares características, y que los haga distinguibles del resto de eventos que conforman el espacio muestral.

Los sistemas de clasificación no supervisada son algoritmos como: sistemas basados en super vectores (SVM); Ensamblados de árboles de decisiones; relaciones de Naïver-Bayes, o agrupación por vecinos cercanos “*Kmeans*”.

Uno de los sistemas de clasificación no supervisada de mayor relevancia para el área médico-biológica, es el uso de redes neuronales, las cuales se han aplicado para la resolución de clasificación de imágenes tanto en diagnóstico como en evaluación. Su uso es ampliamente utilizado y la mayoría de trabajos reportados se basan en la detección de cuerpos a partir de imágenes.

Las redes neuronales pueden ser desarrolladas en su totalidad, o se puede optar por la técnica de transferencia de conocimiento en redes preexistentes. Esta

técnica se usa principalmente en casos donde los cuerpos compartan similares dimensiones o características a los cuerpos analizados por la red neuronal originalmente diseñada.

Análisis de imágenes de células

En lo que análisis de imágenes de células respecta, diferentes autores han dado una gran importancia al primer paso de preprocesamiento, siendo necesario para eliminar algunos artefactos no deseados (Braiki et al., 2020). Uno de los preprocesamientos más utilizados es la conversión de la imagen original a un espacio de trabajo con el cual resulte más sencillo realizar la separación de información de interés (Braiki et al., 2020).

La segmentación es un proceso indispensable para el análisis de imágenes (Adão et al., 2020). Se han realizado mejoras en el proceso de segmentación para mejorar la precisión (Adão et al., 2020; Biswas & Hazra, 2020; Kaur & Sahambi, 2016), la combinación entre filtros, preprocesamiento y una serie de funciones operacionales permite mejorar sustancialmente el desempeño de los trabajos relacionados.

Distintos enfoques pueden ser tomados para lidiar con el buen análisis de las imágenes dependiendo de las características relacionadas con el sistema de captura (calidad de imagen, sistema de captura) y con la naturaleza de las muestras biológicas a analizar (tamaño, forma, variación) (Braiki et al., 2020; Gómez Flores et al., 2014). Existen múltiples herramientas manuales de identificación morfológica o conteo celular a partir del análisis de imágenes. Sin embargo, se reporta en la literatura que a pesar de su aplicabilidad, la inversión en tiempo es alta para obtener resultados deseables que no siempre pueden ser óptimos debido a la ineficiencia de estas herramientas para trabajar sobre cuerpos muy similares o muy complejos (Smith et al., 2018a). Variaciones leves en las características de la imagen afectan sustancialmente la calidad del análisis en general (W. Wang et al., 2019), es por esto que existe una tendencia a optar por métodos específicos o generar sistemas especializados para resolver casos particulares (Braiki et al., 2020; W. Wang et al., 2019).

En general, la clasificación por morfología es óptima en casos con gran cantidad de información como en imágenes de super alta resolución (Shen et al., 2019; X. Zhang et al., 2020). Sin embargo, una aproximación basada únicamente en identificación morfológica se ha identificado como de baja precisión en trabajos con células de morfología muy similar, muy compleja, o una combinación de ambas (Gómez Flores et al., 2014). En casos donde el acercamiento por segmentación no resulta óptimo en primera instancia, muchos sistemas añaden nuevos canales de información tales como información hiper espectral (Q. Wang et al., 2016) o basan su análisis en una evaluación por super píxeles (Tareef et al., 2017).

El uso de un análisis supervisado es un buen principio cuando se tienen trabajos donde se encuentren datos identificados con buenas variables obtenidas por medio de un segmentado y extracción de características adecuadas. La combinación de todas estas técnicas podría permitir analizar un elevado número de imágenes con mayor rapidez y seguridad que al realizar dicho análisis de forma manual.

METODOLOGÍA

Ambiente de programación

Toda la programación fue realizada en el ambiente Matlab® R2020b bajo licencia de uso académico proporcionada por la UNAM. Se utilizaron las herramientas especializadas (TOOLBOX) de *MACHINE LEARNING*, *PARALLEL COMPUTING*, y *IMAGE PROCESSING*. La compilación y los datos de rendimiento fueron generados en una computadora Lenovo ThinkPad SO W10, 8Gb-RAM, SSD 240Gb, 2 Núcleos. Intel® Core™ i7-4600 CPU 2.10GHz 2.69GHz.

Características particulares del banco de imágenes.

Se analizaron un total de 49 imágenes de ensayos de conteo de micronúcleos amablemente donadas por un miembro del grupo de trabajo del *Laboratorio de Nanomedicina* del CNYN-UNAM: *Victoria Vera Pineda*. Cada imagen fue tomada por la misma cámara de teléfono celular a distancias aleatorias con respecto al ocular del microscopio. Las imágenes que conforman la galería principal tienen variación en las condiciones de captura respecto a, luz, intensidad, magnitud, tamaño, saturación y número de células por campo. Tienen una resolución de entre 540x766 a 3120x4160 pixeles, y fueron capturadas con un teléfono celular Samsung S7. Las características del ensayo de micronúcleos son LSPH aislados bajo protocolo estándar, sometidos a citocalasina B, meta- o para- carbamatos de dietileno, teñidos con colorante Giemsa. La muestra fue obtenida de un donante hombre aparentemente sano, previo consentimiento informado.

Sistema de preprocesamiento.

El sistema de preprocesamiento consiste de dos partes, la primera para generar información en diferentes espacios de color de los cuerpos, de forma directa sobre la imagen de entrada. Con la finalidad de identificar la Región Principal de Interés (ROI_p), y las regiones de la periferia de los cuerpos principales identificados.

Para cada imagen se extrajo la ROI_p y todas fueron redimensionadas al tamaño de la ROI_p más grande del nuevo set generado. Se almacenan las ROI de los campos de visión que fueron reconocidos con circularidad ≥ 0.2 y Área $>70\%$ del

Área Total. Estas ROI fueron pre procesadas para su conversión a distintos espacios de color con la finalidad de extraer más fácilmente información de interés. Los espacios de colores fueron los espacios HSV, Cie L*a*b* y RGB invariante de iluminación.

A partir de la información de los canales de a* y G, se identificaron los cuerpos principales de interés, que corresponden a aquellos cuerpos que presentan una coloración marcada respecto al fondo, en dirección a bajos niveles de intensidad del canal G y a*. De esta forma se identifican tanto los cuerpos nucleares de posibles eventos celulares, así como de cualquier otro artefacto de características similares. Como criterios de limitación se consideraron un mínimo tamaño de área y relación de aspecto, los cuales son criterios básicos de filtrado de eventos. Estos criterios no son suficientemente robustos para separar una totalidad de cuerpos celulares respecto a los artefactos, sin embargo, son suficientes para reducir el número de eventos derivado de picos de intensidad característicos de imágenes de bajas resoluciones y rendimiento ineficiente del sistema de captura.

En pasos intermedios de la segmentación se aplicaron distintos filtros a las imágenes obtenidas de cada evento principal identificado. Estos filtros tuvieron el objetivo de mejorar de forma local el contraste entre el fondo y el cuerpo del citoplasma, permitiendo un mejor desempeño del sistema de segmentación. Se utilizó un ajuste de contraste y un mejoramiento de la sección central de las imágenes extraídas. Para ello se utilizaron el sistema de ajuste básico para la generación de iluminación por luz blanca. Finalmente, para una relación de aspecto >0.8 se almacenó la región circundante del cuerpo de interés principal correspondiente a 1.5 veces el diámetro del área del cuerpo principal. En caso el caso donde la relación de aspecto es menor a 0.6, el área circundante almacenada corresponde a 0.75 veces dicho diámetro. Se generó una identificación de la presencia de cuerpos principales cercanos dentro de esos radios de cercanía y se agruparon aquellos eventos aislados que se encontraran dentro, así mismo, se identificó al cuerpo de mayor área como el cuerpo principal del grupo, y se procedió

a una segunda expansión de la ventana con el fin de poder procesar la detección del citoplasma durante el proceso de segmentación

Sistema de segmentación y extracción de características.

La segmentación de los cuerpos fue realizada por medio de la detección de los cuerpos de mayor contraste del canal G aumentado, así como de sus respectivas regiones de campo de los canales S del espacio HSV, a^* del Cie $L^*a^*b^*$ y sus colores del espacio RGB invariante.

A partir de los datos extraídos de la segmentación, así como de las características de cuerpos principales y secundarios. Se determinó el cálculo de diferentes relaciones de aspecto de carácter morfológico. Se extrajeron las características morfológicas enlistadas en la sección de antecedentes, en la **Tabla 2**, para ambos cuerpos, y adicionalmente, las características extraídas mostradas en la **Tabla 3**.

Tabla 3. Características de interés de medición de relación entre cuerpos principal y secundario.

Característica extraída	Ecuación
Diferencia de áreas secundaria y primaria	$A_s - A_p$
Diferencia de áreas secundaria convexa y primaria	$A_{sc} - A_p$
Relación de áreas secundarias respecto a principal	$\frac{A_p}{A_s}$
Lejanía de puntos secundarios respecto a los bordes del cuerpo principal	$\frac{\sum \min(A_{s(x,y)} A_{p(x,y)})}{N * (A_{s(x,y)} A_{p(x,y)})}$
Perímetro del cuerpo secundario	# de pixeles que conforman el contorno de A_s
Área convexa del cuerpo secundario	Cantidad de pixeles que conforman el cuerpo poligonal más pequeño que englobe al cuerpo secundario
Intensidades promedio, media, moda y desviación estándar de pixeles del cuerpo principal	$mean(V_p); median(V_p); std(V_p);$
Intensidades promedio, media, moda y desviación estándar de pixeles del secundario	$mean(V_s); median(V_s); std(V_s);$
Circularidad de los cuerpos secundarios	$\frac{4 * A_s}{\pi * DiámetroMax^2}$

Donde A_s , A_p y A_{sc} representan las correspondientes áreas secundarias y principales de cada evento y al área convexa del cuerpo secundario. $A_s(x,y)$ y $A_p(x,y)$ representan los vectores de posición de los elementos de los cuerpos A_s y A_p , que sirven para determinar la distancia entre elementos del vector $A_s(x,y)A_p(x,y)$

La cantidad de características extraídas tuvo la finalidad de identificar propiedades de relación entre las diferentes secciones de cuerpos y sus respectivos diámetros, que generaran una óptima clasificación de los distintos tipos de eventos detectados. Para así generar una clasificación inicial de cuerpos a partir de estas características.

Generación de grupos principales de eventos por clasificación no supervisada.

El sistema de clasificación de cuerpos se obtiene a partir de un análisis de componentes principales realizado a los datos extraídos. Los datos no fraccionarios unitarios fueron previamente normalizados en función del z-score. Esto con la finalidad de mantener una variación en la cercanía de los datos de forma uniforme entre las variables, lo que evita que durante la determinación de los grupos óptimos de clasificación (o clústeres) se generen grupos a partir de la variación de los datos no normalizados, lo que culminaría en una clasificación no óptima.

La clasificación fue realizada con los componentes principales necesarios para explicar un mínimo del 95% de variación de los datos. Realizada por medio de una agrupación mediante cercanía promedio entre valores vecinos (*kmeans*). Se identificó la importancia de los predictores y se trabajó con aquellos de mayor importancia. La clasificación generada fue almacenada en un vector independiente y se generó un banco de imágenes con un número de grupos diferentes, igual al número de grupos óptimos determinados por la evaluación del sistema *kmeans*. Se generaron un total de 3926 eventos, almacenados en un total final de 7 grupos diferentes. Este banco de imágenes sirvió para generar el primer paso del sistema de clasificación basado en imágenes.

Reclasificación de eventos identificados para corroboración.

La reclasificación basada en imágenes tuvo por objetivo disminuir la probabilidad de mal agrupar cuerpos no similares que fueron almacenados en

grupos incorrectos, debido a que las características extraídas no fueron lo suficientemente adecuadas para clasificar dichos cuerpos.

El sistema de clasificación basado en imágenes consistió en el entrenamiento por medio de transferencia de conocimiento de una red neuronal preexistente, AlexNet. Esta red neuronal fue desarrollada para clasificar objetos cotidianos como coches, bicicletas, casas, o autos. La transferencia de conocimiento consistió en reemplazar las capas 4 y 23 de la red neuronal por capas con la cantidad de grupos identificados en el paso anterior. Para posteriormente, comenzar con el sistema de reentrenamiento de la red neuronal, alimentándola con un set de entrenamiento extraído del banco de imágenes de eventos individuales generado en el paso anterior.

Este sistema de entrenamiento fue probado con el set de prueba generado a partir del banco de imágenes inicial para esto, se extrajeron el 10% de imágenes de forma aleatoria, pertenecientes a cada clase. En los casos donde dichas imágenes no contaran con un puntaje significativo superior, o predicción igual a su clase, determinada por el sistema no supervisado, dicho evento fue re etiquetado. Este proceso se repitió para todo el set de prueba. Posteriormente, un segundo reentrenamiento fue realizado con la nueva configuración de imágenes. Este proceso se repitió en 2 ocasiones, generando un total de 4 sistemas de clasificación de cuerpos.

La agrupación final de los eventos identificados fue etiquetada manualmente con los nombres: “cuerpos celulares mononucleares”, “cuerpos celulares binucleares”, y tipo 1, 2 y 3 de cuerpos negativos a células. Los cuerpos celulares mononucleares y binucleares fueron elegidos por medio de la selección del usuario previsualizando los tipos de eventos agrupados en cada una de los grupos. Dichos grupos fueron seleccionados para su posterior análisis para la detección de posibles aberraciones genéticas.

Determinación de probabilidad de aberraciones genéticas de MN, puentes y yemas.

La determinación de potencial existencia de aberraciones genéticas de interés en los cuerpos celulares identificados como “cuerpos celulares binucleares”, fueron seleccionados para ser post procesados. Se clasifica la presencia de yemas y micronúcleos, así como la presencia de puentes nucleares. La identificación de yemas se realiza mediante la determinación de irregularidades en zonas contrarias a la dirección del núcleo, es decir, en zonas exteriores al centro de la célula.

La detección de micronúcleos se realiza mediante la segmentación basada en bordes de cuerpos nucleares presentes en la imagen analizada. Con la intención de cuantificar la presencia de bordes en cuerpos alejados del eje de los cuerpos celulares. La detección de puentes celulares, procesada únicamente para cuerpos binucleados, se realiza mediante la determinación de regiones de unión entre cada uno de los dos núcleos celulares.

Validación.

La validación de la presencia de los cuerpos celulares se realiza mediante el conteo manual del catálogo de imágenes detectadas como cuerpos celulares, así como de la existencia de cuerpos celulares mal agrupados que forman parte de los etiquetados, como los diferentes cuerpos negativos a células

RESULTADOS

Clasificación de cuerpos detectados.

A partir de un total de 3926 eventos, el sistema generado clasifica un total de 3926 eventos aislados. Los cuales son clasificados en 7 grupos principales finales. Ejemplares de esto grupos se muestran en la **Figura 2**.



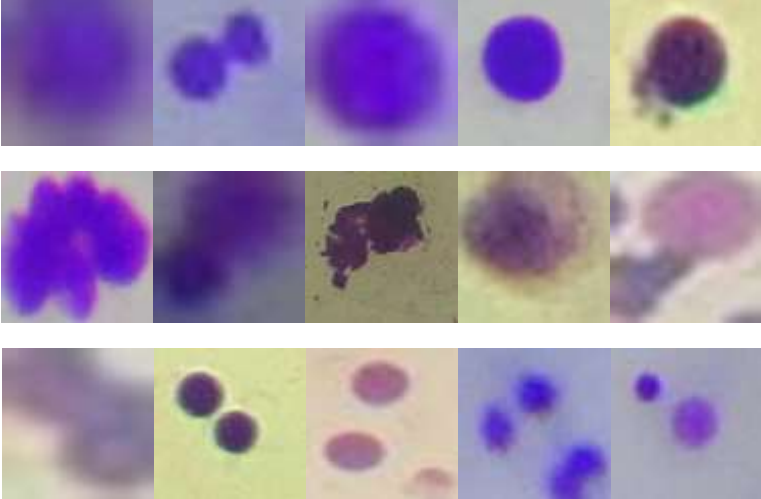
Tipo de cuerpo	Ejemplo de Clasificado
Eventos+ para células binucleados	
Eventos+ para células mononucleadas	
Eventos negativos a cuerpos celulares	

Figura 2. Ejemplos de cuerpos que conforman los grupos celulares.

La **Figura 3** muestra el conteo manual y automático total enfocado a células mononucleares y células binucleares.

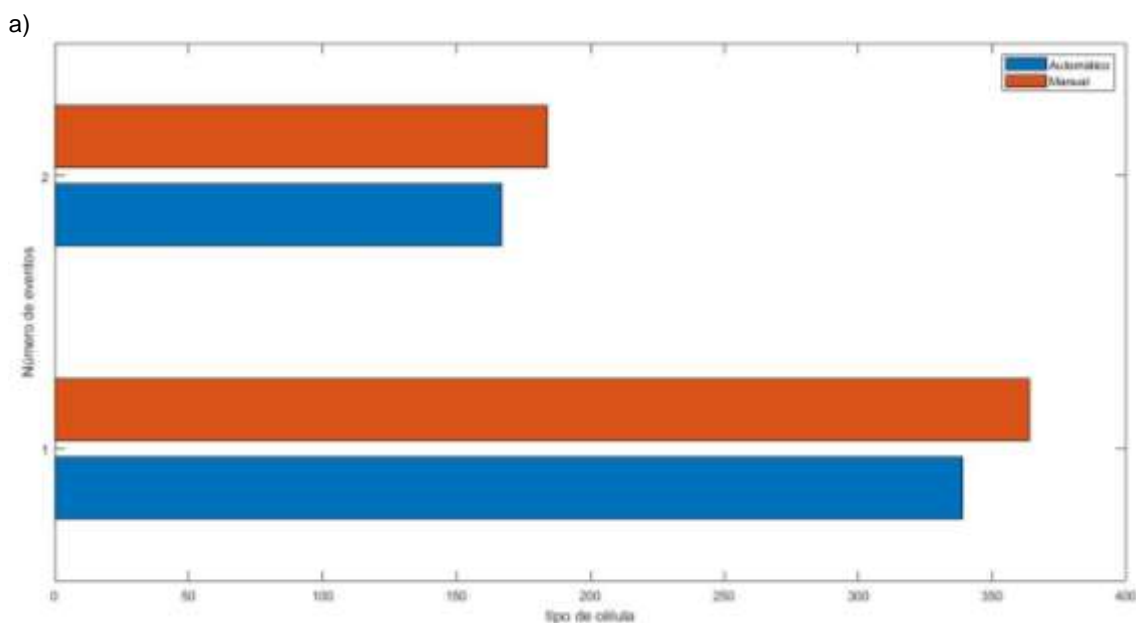


Figura 3. Detección de micronúcleos, puentes celulares y yemas.

La cantidad total de eventos positivos para cuerpos celulares detectados por el sistema son 167 de cuerpos celulares mononucleares y 339 para cuerpos binucleares. Bajo conteo manual, se distinguen e identifica un total de 184 células mononucleadas, 364 binucleadas y 61 multinucleadas. La **Tabla 4** muestra la relación entre los cuerpos identificados por el sistema validando manualmente la agrupación de cuerpos mono y binucleares.

Tabla 4. Relación de clasificación de cuerpos mononucleados y binucleados positivos generada por el modelo

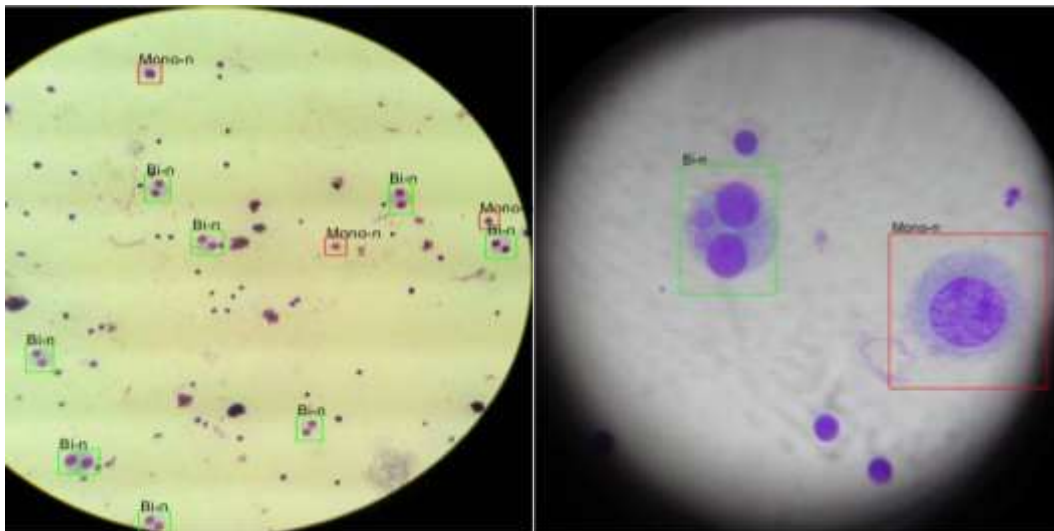
Tipo de cuerpos	Numero de cuerpos agrupados en las distintas secciones				
	Eventos	Verdaderos positivos	Falsos positivos	Precisión (%)	No detectado
Mononucleadas	167	128	59	76.65	84
Binucleadas	339	280	39	82.6	56

De los eventos identificados la precisión final es de 82.6% en la clasificación final para los eventos de interés para el postprocesamiento (eventos celulares binucleados), y presenta una pérdida de detección del número total de eventos identificados manualmente del 25.54%.

Efectividad de detección directa en imágenes de entrada.

Las características particulares del banco de imágenes presentan un elevado número de eventos potencialmente identificables como eventos falsos positivos a cuerpos celulares por su similitud morfológica en tamaño, forma y valor de intensidad de los píxeles que conforman a dichos cuerpos.

La **Figura 4** muestra ejemplares de las regiones de interés de imágenes escogidas por su representatividad del banco de imágenes usado. Los eventos de interés identificados como cuerpos celulares se encuentran marcados en rojo, aquellos detectados como mononucleares, y en verde, aquellas identificadas como binucleares.



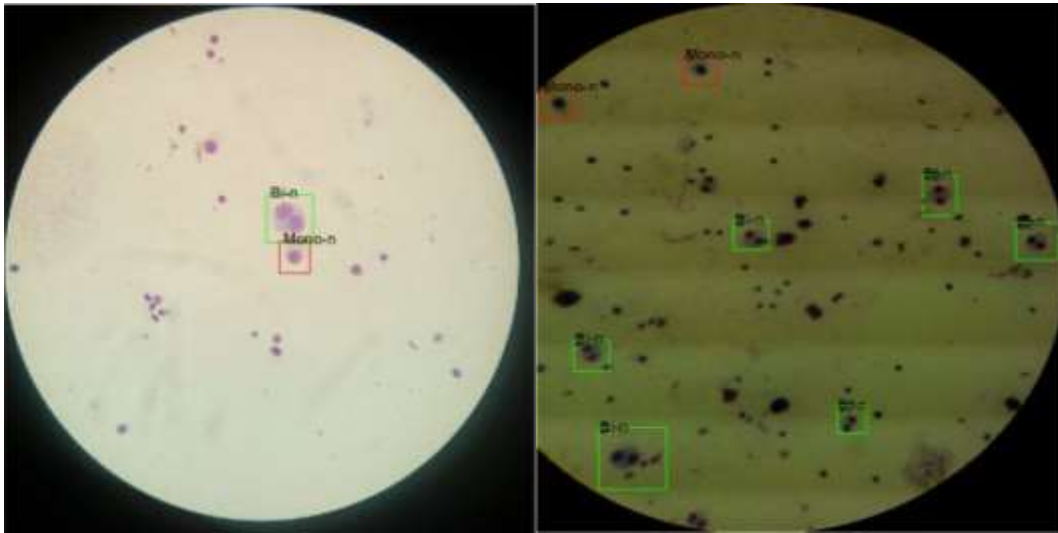


Figura 4. Muestras aleatorias de las regiones de interés de las imágenes de entrada.

Se observa una buena efectividad en la identificación de células en los casos donde los objetos se encuentran mejor definidos y la magnificación utilizada genera cuerpos con dimensiones de más de 100 píxeles de diámetro.

Sistema

La red neuronal (NN) reentrenada para identificar los cuerpos mononucleares y binucleares de entre elementos muy similares. Las características técnicas de la NN se muestran en la **Tabla 5**.

Tabla 5. Características técnicas de la red neuronal entrenada generada

Medición	Ecuación	Valor (%)
Precisión	$= \frac{TP + TN}{TP + FP + TN + FN}$	82.6
Sensibilidad	$= \frac{TP}{TP + FN}$	0.93
Especificidad	$= \frac{TN}{TN + TP}$	0.28
Perdida del modelo en entrenamiento	$Kfold = \frac{\text{predicción de imágenes de entrenamiento}}{\text{categoría imágenes reales entrenamiento}}$	0.12

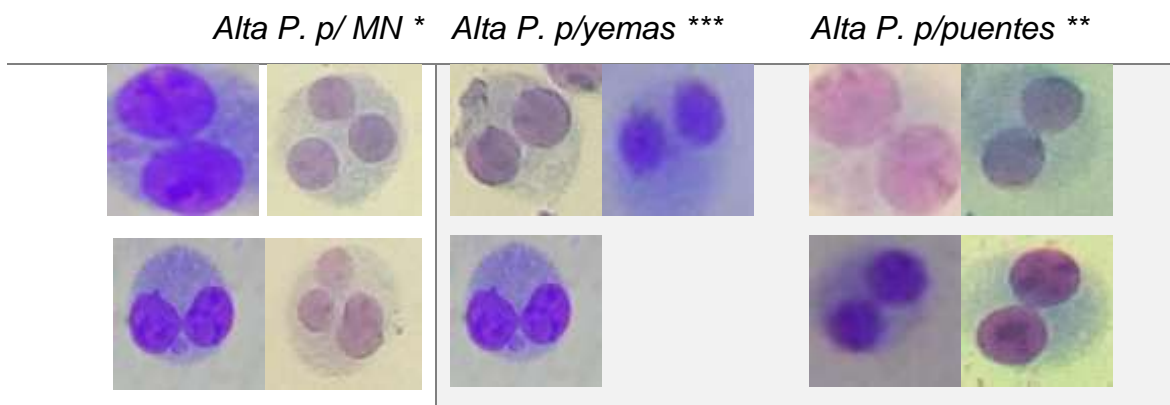
Perdida del modelo en testeo	$Kfold\ training = \frac{predicción\ de\ imágenes\ de\ testeo}{categoría\ imágenes\ reales\ de\ testeo}$	0.19
Número de capas de salida	Numero de categorías óptimas de separación de cuerpo	7
Tamaño de entrada de imagen	Tamaño de entrada de imagen individual para su clasificación	[227 227]

Donde: TN= Verdaderos negativos; TP= Verdaderos positivos; FP= falsos positivos; FN= Falsos negativos.

Clasificación por probabilidad de presentación de cuerpos de interés.

Ejemplos de la función secundaria de detección de cuerpos internos en los grupos de imágenes elegidos son mostrados en la **Tabla 6**.

Tabla 6. Tabla de eventos probables para presencia de micronúcleos.



*Presencia de subcuerpos redondos

**Presencia de cuerpos alargados unidos.

*** Presencia de irregularidades en el/los núcleos redondos

Se observa una buena identificación de cuerpos potenciales de interés, sin embargo, debido a la falta de agrupación de cuerpos multinucleados, el sistema identifica algunos de estos cuerpos como cuerpos binucleares con alta probabilidad de micronúcleos.

Finalmente, el conteo manual de células identificadas como potenciales portadores de las características de interés son mostradas en la **Tabla 7**, a continuación.

Tabla 7. Comparación entre eventos identificados como potenciales portadores de cuerpos de interés y su corroboración manual.

Identificación		Candidatos a puentes nucleares	Candidatos a MN	Candidatos a yemas
Manual	Eventos posibles	65	19	11
	Eventos poco identificables	48	17	6
		17	2	5
Automática	n	97	34	29
	identificado	23	13	2
	inidentificable	36	1	0
	negativo	38	21	27
Relación	Precisión respecto al número de eventos automáticos (%)	23.71	38.24	6.9
	Porcentaje de elementos potenciales, respecto al total de eventos celulares (%)	34.64	12.14	10.36
	Eventos micro nucleares positivos (%)	47.92	76.47	33.33

Se destaca la efectividad para contar eventos ´positivos MN. Siendo de importancia la cantidad reducida de eventos de necesidad de confirmación, en comparación con la cantidad total de eventos independientes que tendrían que ser contados de forma manual sin la intervención del sistema presentado.

DISCUSIÓN

Criterio de elección de método de clasificación

Como la resolución de las imágenes varían en resolución. No es posible basarse en una relación preexistente que asegure la equivalencia directa entre un determinado número de píxeles y un tamaño en micras. Razón por la cual, es necesario estandarizar las propiedades de los objetos obtenidos por el filtro inicial de características morfológicas de cada una de las imágenes, a fin de evitar una incompleta agrupación de cuerpos debido a las mediciones no normalizadas que pudieran provocar una clasificación errónea. Esto permite no depender de una clasificación basada únicamente de características directas, como, por ejemplo, una clasificación a partir de tamaños y circularidad únicamente.

En orden de cumplir con el objetivo planteado, así como a la limitación del banco de imágenes necesaria para desarrollar una red neuronal desde cero (Smith et al., 2018b) y dado que los cuerpos extraídos y preprocesados en los pasos de preprocesamiento y segmentación, tienen un tamaño promedio de 170 ± 80 pí. Se optó por la técnica de reentrenamiento de redes neuronales preexistentes, el cual es un sistema alternativo al entrenamiento completo de NN (Amin et al., 2021).

Con la finalidad de obtener una óptima clasificación de cuerpos a partir de sus características visuales. La extracción y preclasificación de cuerpos en función de criterios es necesaria para generar un punto de partida y definir el número óptimo de clases de cuerpos diferentes que será necesaria. Este paso inicial es crucial para el comienzo de la clasificación de imágenes de arquitecturas similares (Tareef et al., 2017; Zhao et al., 2016), debido a que la extracción de características de imágenes parte del entrenamiento indicando cuerpos aislados representativos, que son extraídos de las imágenes de entrada.

La clasificación automática en función de las propiedades medidas es un proceso ampliamente utilizado en la clasificación de morfologías de cuerpos no celulares (Mbogba et al., 2018) y celulares (Zhao et al., 2016) a partir de diferentes acercamientos y arquitecturas. La combinación de sistemas no supervisados con

sistemas supervisados para la clasificación de imágenes de cuerpos celulares es utilizada para identificar anomalías que asistan en el diagnóstico de cáncer de cérvix (Tareef et al., 2017; Zhao et al., 2016), mama (Gómez Flores et al., 2014), enfermedades hematológicas (Acevedo et al., 2019), entre otras (Adão et al., 2020).

El sistema de clasificación por grupos es aceptado en el análisis de información. El acercamiento a la clasificación no supervisada parte de la identificación de los componentes principales de los cuerpos de interés. Con la finalidad de identificar las variables que aporten mayor información e intervalos mejor marcados entre los diferentes cuerpos. Esta aproximación está inspirada en la metodología tradicional de detección de forma, y trata de lidiar con una problemática identificada en diferentes trabajos, la ineficiencia de la detección morfológica a partir de un número limitado de características como el área, la curvatura o la intensidad. Esto, sin la necesidad de recurrir a un nuevo canal de información adquirido directamente en el sistema de captura, tal como información del espectro en infrarrojo o de profundidad. Y parte de la posibilidad de que la baja precisión en la detección y clasificación de los cuerpos está asociada a la falta de mediciones que describan adecuadamente las diferencias entre cuerpos de características similares que superan los filtros de criterios de selección predefinidos por el programador.

La clasificación de cuerpos en función de sus componentes principales en lugar de las mediciones directas permite agrupar los cuerpos con un mayor grado de amplitud. Lo que genera una clasificación inicial con mayor diferenciación, que permite entrenar al sistema clasificador con cuerpos similares entre sí, y que, a su vez, favorece la precisión de la red neuronal empleada.

El uso de una validación múltiple de la clasificación de elementos ayuda a mejorar y validar de forma interna la clasificación final que el sistema generará, esto es debido a que se disminuye considerablemente la probabilidad de encontrar generar una clasificación parcial que resulte óptima sólo para un grupo de eventos, pero no óptima para clasificar todos los cuerpos de forma ideal a partir del punto de origen marcado por la clasificación parcial encontrado. Sin embargo, para definir un

punto de partida de forma ideal, es necesario conocer características *a priori* del sistema, lo cual, dependiendo de la complejidad del sistema, puede requerir un estudio completo previo. La validación cruzada permite generar un sistema de clasificación no supervisada más confiable partiendo de puntos aleatorios para generar la clasificación. Sin embargo, múltiples evaluaciones y reclasificaciones pueden provocar un sobre entrenamiento que aumente considerablemente el error en la clasificación de nuevos eventos. Por esta razón, se realizó una clasificación no supervisada cruzada y limitó el reentrenamiento sucesivo a solo 3 NN, siendo la clasificación generada por la última NN aquella reportada como resultado final.

La reproducibilidad local se valida a lo largo del proceso de reentrenamiento de los resultados de los clasificadores en serie. Se tomaron como criterios empíricos: reclasificar una baja porción de cuerpos de ensayo en una sola ronda y reclasificaciones sucesivas en categorías donde haya un número elevado de eventos pueden reducir la cantidad de cuerpos mal clasificados por el sistema no supervisado. La clasificación por redundancia permite validar un set de predicciones generadas por distintas NN empleadas para clasificar un mismo elemento, conservando aquella etiqueta que mayor presencia tenga. El sistema presentado aprovecha este acercamiento para validar la pertinencia de reclasificar las imágenes analizadas. El reentrenamiento de las clasificaciones se muestra en la **Figura 5**.

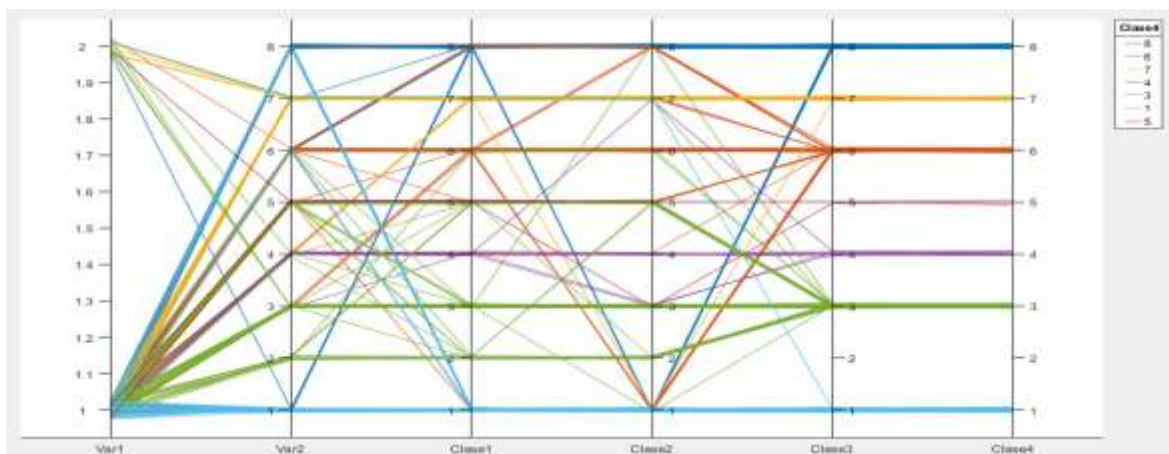


Figura 5. Cambio de etiquetados en el reentrenamiento de clases de clasificación basada en imágenes.

El desempeño final del sistema es medido en función de la capacidad para etiquetar correctamente cuerpos celulares mono y binucleados. En este sentido, la **Figura 4** mostró eventos representativos de ejemplares de imágenes promedio del desempeño final. La **Figura 6** muestra imágenes del banco donde las condiciones de captura son identificadas como la más óptima, y la de peores condiciones.

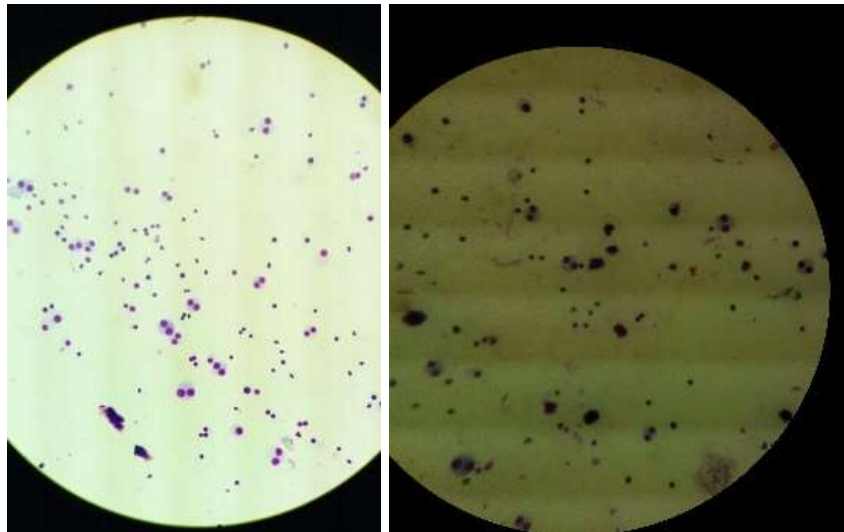


Figura 6. Imágenes representativas de imágenes con las condiciones más (a) y menos (b) óptimas.

De forma visual, se identifica que es en las imágenes de peores condiciones de captura es donde se encuentran una menor precisión en la clasificación de eventos, así mismo, es en ellas donde las condiciones de contraste, enfoque y retención de la imagen original son más irregulares, lo que genera una detección errónea de los elementos de interés.

Clasificación de grupos de importancia y detección de cuerpos de interés

La clasificación a partir de imágenes por medio de NN permite estudiar aquellos cuerpos con una pobre descripción a partir del análisis PCA de sus propiedades obtenidas durante la segmentación, al tiempo que ayuda a corroborar y detectar eventos mal agrupados durante la generación de los grupos iniciales a partir de la clasificación no supervisada, y con mayor razón en eventos como los que conforman este método, debido a no ser contribuciones representativas para la clasificación no supervisada (Ver **Figura 7**). Las NN extraen información de

importancia para cada grupo que es utilizada para determinar la clasificación de nuevos eventos, con una facilidad mayor en función de la diferencia entre los grupos a identificar.

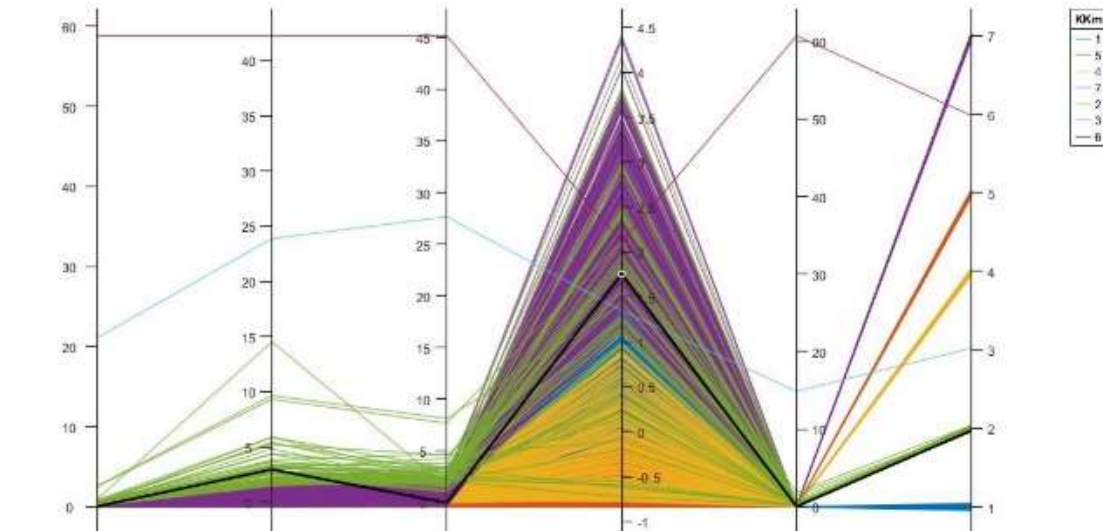


Figura 7.a) contribución a los primeros componentes principales del sistema de datos medidos. Cada sección vertical representa los valores de uno de los componentes principales.

Gracias al acercamiento dado por este trabajo, la NN entrenada permite trabajar con un set de cuerpos no identificados, para dirigirlo a la clasificación de cuerpos presentes en muestras de ensayo de MN, usando LSPH. En función de su potencialidad de ser o no, un cuerpo de interés, que pueda ser identificado como célula mononucleada o célula binucleada, de entre artefactos con características similares.

Cualidades de cuerpos clasificados

La diferencia presente entre los distintos grupos permitió una óptima identificación entre cuerpos con características similares en cuerpos celulares positivos respecto a cuerpos celulares negativos. De los cuerpos que presentan una alta similitud entre cuerpos celulares, destacan características como redondez o intensidad. Siendo diferenciables de forma visual principalmente por las

irregularidades del contorno. La **Tabla 8** muestra una comparación de eventos positivos y negativos en cuerpos celulares muy similares. Cabe señalar que aquellos cuerpos multinucleados no fueron objeto de identificación del presente trabajo debido a su variabilidad y similitud con artefactos no celulares presentes en el banco de imágenes de trabajo.

Tabla 8. Comparación entre eventos positivos y negativos muy similares a cuerpos celulares.

Tipo de célula	Clasificado positivamente	Clasificado negativamente
Mononucleada		
Binucleada		

Los cuerpos monos y binucleados que presentan algún cuerpo de interés, como micronúcleos, yemas y/o puentes celulares, tienen una alta similitud con los cuerpos mono y binucleados que no presentan dichas características, además de presentarse con una menor frecuencia respecto a la cantidad de eventos donde no se presentan, lo cual ha sido tema de discusión entre trabajos relacionados con la detección de células positivas a linfocitos y a linfocitos con micronúcleos. Del mismo modo, la identificación directa de estos eventos a partir de los criterios de inclusión y las variables comúnmente utilizadas (mostradas en la sección de antecedentes, en la **Tabla 2**) resulta inadecuada, debido a la probabilidad de clasificar cuerpos falsos positivos únicamente a partir de los criterios de inclusión. Por lo anterior, un acercamiento basado únicamente en la generación de una NN o una detección morfológica que identifique directamente los cuerpos celulares positivos que presenten dichas características en este caso particular resultaría inadecuado debido a las características que hay que resolver para cada uno de los enfoques.

La alternativa de clasificación por probabilidad es utilizada en diferentes trabajos para identificar eventos muy similares difíciles de aislar, y cuya clasificación a partir de los criterios de inclusión o exclusión es de alta complejidad. El uso de una segmentación directa a los cuerpos celulares identificados puede sustancialmente identificar cuerpos a partir de sus características propias, o ser enfocada para determinar la probabilidad de existencia de dichas características dentro de los cuerpos identificados.

La **Figura 8** muestra un esbozo del funcionamiento del sistema de detección de MN.



Figura 8. Separación de capas de citoplasma y cuerpos nucleares presentes. Se identifican irregularidades en los contornos del núcleo izquierdo asociables a yemas, en las partes centrales, se encuentran irregularidades y una unión entre cuerpos nucleares principales.

Comparativa con trabajos similares.

Este sistema cuenta con un porcentaje de acierto de 82.6% en la mejor clasificación de eventos celulares (células binucleadas. Este porcentaje de acierto es comparable con trabajos similares reportados en la literatura. Y es esta la clase de importancia para el post procesamiento.

Existen múltiples trabajos donde se apoyan en el uso de redes neuronales para resolver problemas de clasificación de imágenes de cuerpos celulares (Acevedo et al., 2019; Aich et al., 2020; Hollandi et al., 2020; Mbogba et al., 2018; Rahmat et al., 2020; Q. Wang et al., 2016; L. Zhang et al., 2014), algunos de los cuales especialmente para la distinción de células circulatorias de sangre periférica humana (Acevedo et al., 2019; Q. Wang et al., 2016). No hay un acercamiento con

sistemas de clasificación específicamente para distinción de aberraciones nucleares en linfocitos aislados de sangre periférica humana, en cuestión de micronúcleos, puentes celulares y yemas que al mismo tiempo permita distinguir eventos positivos a cuerpos celulares de eventos negativos con características muy similares a los cuerpos de interés (ver **Tabla 1**).

Trabajos similares comparten características en común tales como: contar con un pre etiquetado realizado por un profesional experimentado o un aislado inicial de los cuerpos celulares de interés, bien definidos y sin la presencia de cuerpos que no sean los cuerpos que se pretenden clasificar, cuyo banco de imágenes cuenta con una ligera varianza o nula variación en las condiciones de captura; el banco de imágenes inicial fue generado y la clasificación mejorada a partir del apoyo de un sistema especializado de captura de imágenes tales como cámaras acopladas de *alta* o *super alta* definición, sistemas de captura de información en infrarrojo(Q. Wang et al., 2016), ultrasonido (Deepthy Mary Alex et al., 2020; Gómez Flores et al., 2014; Zhou et al., 2019), resonancia magnética(Liu et al., 2020; Rahmat et al., 2020), sistemas de captura de información tridimensional(Delannoy et al., 2020), microscopios con sistemas acoplados y software embebido (Frieauff et al., 1998; Zhao et al., 2016); además de contar con una gran cantidad de imágenes en total, que fácilmente alcanzan una galería de entre 1,000 a 10,000 unidades por cada tipo de cuerpo de interés (Acevedo et al., 2019).

Las características mencionadas resultan deseables durante el proceso de la generación y entrenamiento de los algoritmos empleados para resolver una tarea específica, sin embargo, las hace susceptibles a una baja reproducibilidad en cuestión del desempeño frente a nuevos bancos de imágenes a pesar incluso de tratarse del mismo tipo de modelos celulares. Razón por la cual, diferentes trabajos publicados se basan en reportar modificaciones directas a trabajos previos que mejoran su reproducibilidad.

De las ventajas que presenta el sistema desarrollado son, la falta de necesidad de un banco de imágenes aislado de forma manual; la baja cantidad de imágenes de entrada necesarias para alimentar el sistema; así como el diseño

enfocado a la fácil implementación. Estas características, que comúnmente se identifican como no deseadas para estos fines, juegan un papel importante para el cumplimiento de los objetivos planteados en este trabajo, ya que se logra obtener una clasificación con una precisión que se encuentra dentro del intervalo para sistemas similares [80%-90%], donde se trabajan con bancos de imágenes más homogéneos, corroborando así la baja contribución de las condiciones de la imagen inicial en el sistema de clasificación final.

El rendimiento medido en el tiempo empleado en la totalidad de los procesos se puede medir en función de los módulos de trabajo, La **Tabla 9** muestra el tiempo promedio medido entre el inicio y termino de cada uno de los módulos de trabajo que conforman al algoritmo.

TABLA 9. Tiempos promedio empleado en cada uno de los procesos del algoritmo.

#	Descripción	Tiempo (segundos)
1	Detección de campo de interés	125.45
2	Segmentación de cuerpos de interés	1781.25
3	Clasificación no supervisada	319.42
4	Clasificación supervisada	5640 (1er) 6480 (2ndo) 2,000.0 (3er)
5	Evaluación de cuerpos para el post procesamiento para la detección de aberraciones genéticas	463.34

La tabla de tiempos posiciona a los procesos de rentrenamiento de las redes neuronales como los puntos de mayor tiempo. Del mismo modo, se identifica que los procesos de clasificación una vez entrenados los sistemas de clasificación son de los procesos más rápidos y que los procesos de entrenamiento de clasificación. El tiempo empleado por el sistema de inicio a fin, validando un total de 3926 elementos individuales identificados en 49 imágenes individuales, es de 2h43min,

lo que implica que, en promedio, el sistema requiere de aproximadamente 2.78 segundos para completar su ejecución.

Cabe señalar que, del tiempo total del algoritmo, el mayor consumo de tiempo corresponde al tiempo de entrenamiento de las distintas redes neuronales entrenadas. No obstante, resulta importante recalcar el hecho de que, una vez entrenada la NN final, se obtiene una NN funcional capaz de clasificar nuevos eventos ajenos al banco de imágenes original. Por lo que, puede ser utilizada directamente para la clasificación inicial de nuevos cuerpos de interés. Lo que nos deja con dos posibilidades: alimentar directamente a la NN con imágenes de cuerpos aislados para generar una predicción confiable basada en el entrenamiento adquirido, o, ingresar imágenes del campo de visión del ocular de un microscopio óptico, de la misma forma como se ingresaron las imágenes inicialmente para realizar el preprocesamiento y el segmentado, pero omitiendo las secciones correspondientes al sistema de entrenamiento no supervisado de aprendizaje automático y el sistema de aprendizaje profundo de clasificación de imágenes.

Considerando estos dos posibles modos de aprovechamiento del sistema, se obtiene un tiempo efectivo, respectivamente, de 0.1 segundos y 0.45 segundos de trabajo neto por cada evento analizado en dichos modos de trabajo. Estos tiempos son similares a los obtenidos por otros autores.

Robustez

Si bien el sistema logra clasificar de forma satisfactoria aquellos cuerpos que cumplen los criterios de selección iniciales a pesar de las condiciones particulares de entrada asociadas a cada imagen inicial. La falta de identificación de cuerpos celulares está directamente relacionada con el cumplimiento de los criterios de selección definidos manualmente con base en la morfología conocida de los tipos celulares.

Una posible razón por la que estos cuerpos no sean clasificados es porque el criterio base requiere que el cuerpo a detectar presente un contraste marcado entre el fondo que rodea a la célula y el citoplasma y el/los núcleos del cuerpo, dicho

valor está delimitado por un *thresholding* definido en función de las variaciones de colores en escala de grises generado a partir de la imagen en el espacio RGB invariante de iluminación. La definición de este parámetro en función de las características mencionadas resulta óptimo para la mayoría de los elementos celulares, sin embargo, derivado de la transformación al espacio de grises a partir de la imagen RGB invariante parece resultar no ser una definición adecuada, debido a que tiende a no lograr detectar cuerpos donde la diferencias entre iluminaciones no logran generar una buena definición entre los cuerpos oscuros generados por los núcleos, por lo que no se logra superar el *thresholding* establecido por el criterio de selección. El porcentaje de pérdida de casos positivos es del 25.56% del total de cuerpos celulares, sin embargo, esta perspectiva logra descartar más de 3,000 eventos negativos de características similares presentes en las imágenes analizadas.

Simplicidad

El sistema está enfocado en reducir al mínimo indispensable la interacción necesaria del humano para el análisis de las imágenes. Debido al diseño pensado para su desempeño en la identificación de morfologías celulares específicas en imágenes capturadas directamente del campo óptico de visión del ocular del microscopio. Los criterios de identificación son laxos pero el sistema no se limita a dichos criterios para determinar casos positivos, lo que permite descartar posibles falsos positivos.

Una diferencia muy grande entre este sistema y sistemas de clasificación de cuerpos aislados existentes actualmente para el conteo de micronúcleos, yemas y puentes celulares, es la función de aislamiento de cuerpos con alta probabilidad de presentar dichos cuerpos. Este enfoque permite al usuario descartar una gran cantidad de cuerpos y enfocarse en aquellos de mayor relevancia, lo que disminuye considerablemente el número de eventos totales analizados manualmente, al mismo tiempo, evita una mala identificación de casos positivos y falsos positivos.

La clasificación de cuerpos de interés en función de la probabilidad en lugar de la detección directa es una técnica utilizada en sistemas especializados en la

evaluación de calidad de muestras histopatológicas. Como es el caso del análisis de muestras de cáncer, que consiste en descartar la mayor cantidad de muestras de imágenes celulares que tengan una baja o nula probabilidad de resultar representativos de muestra positiva (Araújo et al., 2019). Lo que permite al patólogo enfocar su análisis en una baja cantidad de muestras con alta relevancia para la detección de cáncer, en lugar de analizar todo el set de imágenes.

La inclusión de un acercamiento de descarte por probabilidad de presencia aquí planteado ayuda a entrenar de forma más precisa el sistema final de clasificación de imágenes que diferencia eventos positivos de eventos negativos. Y permite evitar problemas comunes de clasificación de imágenes por NN donde se cuenta con una baja cantidad de cuerpos representativos de la clase binucleada positivos con micronúcleo, yemas o puentes celulares.

El acercamiento por probabilidad permite entrenar el sistema para detectar casos positivos a células mononucleadas o binucleadas, lo que contribuye a mantener un modelo simple para su detección, un tiempo óptimo de entrenamiento para cada NN utilizada y evita errores asociados al conteo erróneo de cuerpos presentes dentro de una célula, o a un conteo automático inflado derivado de falsos positivos durante una segmentación manual.

Actualización respecto a acercamientos reportados

Desde la presentación de la competición ImageNet en 2012, la implementación de redes neuronales para clasificación de imágenes ha ido en aumento por su gran potencial de análisis de imágenes, el cual es aprovechado principalmente en el área biomédica (Amin et al., 2021). No obstante, en la revisión realizada en este trabajo, no se encontró su aplicación enfocada a la identificación de los cuerpos celulares aquí analizados. Una revisión inicial del estado del arte enfocado a identificar el desarrollo de sistemas basados en arquitecturas tanto en algoritmos generales como por técnicas de ML y DL, mostrado en la **Figura 9** muestra un número creciente en técnicas de aprendizaje profundo o sistemas combinados de aprendizaje profundo con algoritmos generales de decisión basados

en criterios predefinidos, que contrasta con la disminución de técnicas basadas meramente en identificación a partir de criterios predefinidos.

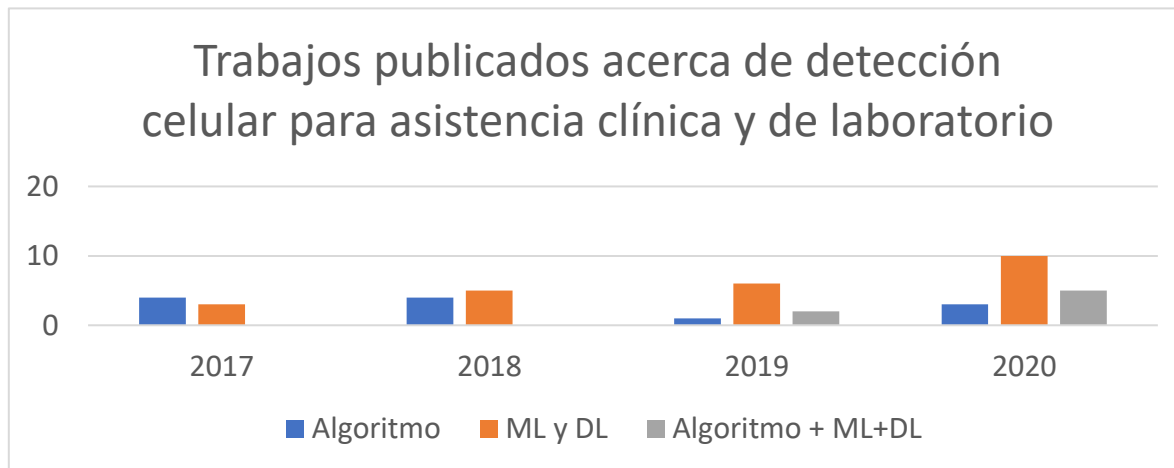
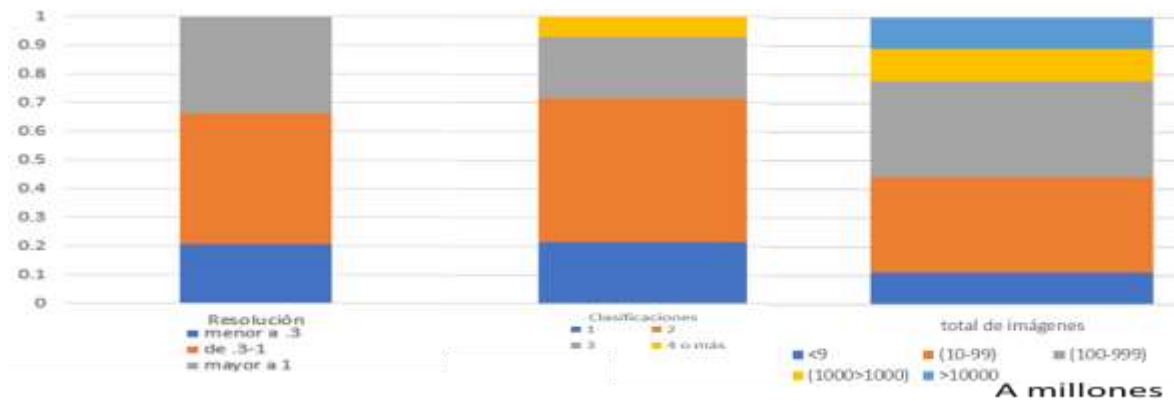


Figura 9. Tipos de arquitecturas presentes en trabajos publicados acerca de detección celular para asistencia clínica y de laboratorio en los últimos años.

El presente sistema pretende ser el primero en apoyarse en arquitecturas NN enfocadas en la clasificación de eventos en ensayo de micronúcleos específicamente en LSPH teñidos con colorantes no fluorescentes. Cuenta con la opción de reentrenar un modelo de clasificación, tiene la capacidad de identificar adecuadamente grupos de cuerpos basados en sus similitudes, y una vez identificados los cuerpos de interés, correspondientes a células binucleadas y mononucleadas, permite clasificarlos en función de la probabilidad de presentar cuerpos extraños identificados como micronúcleos y yemas en células mononucleares, así como también de puentes celulares presentes en células binucleadas. Estas ventajas se agrupan en la **Tabla10** mostrada a continuación.

Tabla 10. Comparación entre sistemas especializados previos, enfocados a LSPH y el sistema desarrollado.

Característica	(Frieauff et al., 1998)	(Bolognesi et al., 2011)	(Taghi et al., 2017)	(Hülber et al., 2020)	Presente trabajo
Tiempo de clasificación	-	-	-	-	34min
Capacidad para distinguir cuerpos extraños	-	-	-	X	X
Necesidad de pre selección manual	X	X	X	No	No
Necesidad de etiquetado inicial	No	No	X	No	No
Detección múltiple de eventos en una única imagen	X		X	X	X
Tiempo global del sistema	-	-	-	-	2h 18min
Distingue micronúcleos	x	x	x	x	X
Distingue células con yemas	No	No	No	No	X
Distingue presencia de puentes nucleares	No	No	No	No	x
Cantidad de bancos de imágenes	1	1	1	1	1
Robustez de banco de imágenes	Homogéneo	Banco de imágenes de luorescencia	Presencia de artefactos muy diferentes	No muy homogéneo	*condiciones de captura, *Tipo de tinción *cercanía de imagen
Precisión para el mejor evento	.	70%	99.6%	85%	82.6%
Uso de herramientas de ML o DL	No	No	Si	Si	Si
Requiere herramienta externa especializada	Si	Si	No	Si	No
Sistema de captura	Cámara especial	Sistema especial	Cámara acoplada	Sistema especial	Cámara celular convencional

Debido a que el presente sistema genera una clasificación de células con una efectividad superior al 80% en presencia de múltiples eventos únicos utilizando un sistema no homogéneo de propiedades tanto de captura, como de los cuerpos presentes. Se abre un enfoque directo para mejorar la clasificación de las imágenes ampliando el banco de imágenes utilizado en el entrenamiento de las NN, por lo que existe una alta posibilidad de mejorar la precisión de la clasificación. Así mismo, es importante resaltar que este trabajo se enfrenta a la clasificación de cuerpos usando un banco de imágenes no etiquetado inicialmente, lo cual es un punto de diferencia respecto a trabajos donde la precisión de clasificación reportada supera el 90%.

Se puede mejorar el sistema de filtrado de imágenes o entrenar al sistema de forma manual, sin embargo, para fines del objetivo de este trabajo, es de mayor interés obtener una mejor capacidad para trabajar en condiciones no óptimas obteniendo una precisión adecuada, que generar un sistema super eficiente, pero con una alta necesidad de preprocesamiento, filtrado, o etiquetado manual de cuerpos celulares. Puesto que son precisamente estas, las características identificadas como puntos de mejora en los sistemas reportados en la literatura, necesarios para aumentar su precisión y reproducibilidad (Tareef et al., 2017).

Si bien, la efectividad de trabajar en condiciones no óptimas permite sugerir una mejor reproducibilidad relacionada con las condiciones de captura e iluminación, se requiere evaluar un banco de imágenes totalmente independiente al actual para validar su rendimiento en condiciones donde las imágenes de entrada presenten mayor concentración celular o diferente cantidad de artefactos dentro de ellas.

Relevancia Actual

Dentro de los principales objetivos del presente trabajo se aborda la necesidad de generar un sistema automatizado que pueda ser directamente acoplado a las condiciones de muestras generadas en un laboratorio de investigación universitario. La buena efectividad determinada durante la evaluación

del sistema frente a una gran cantidad de eventos similares a eventos positivos ayuda a validar el proceso de identificación de cuerpos con alta fiabilidad de casos verdaderos. Por lo que el sistema generado cumple dicho objetivo de trabajo.

Cabe señalar que, en un análisis bibliográfico realizado con la herramienta SCOPUS. Existen pocas aportaciones de investigaciones atribuibles a la aportación de México en los últimos 10 años dentro del campo de la imagenología / procesamiento de imágenes, y son aún menos los aportes directamente relacionados con la aplicación de estas herramientas para el procesamiento de imágenes de ensayos *in vitro* o de diagnóstico (ver **Figura 10**).

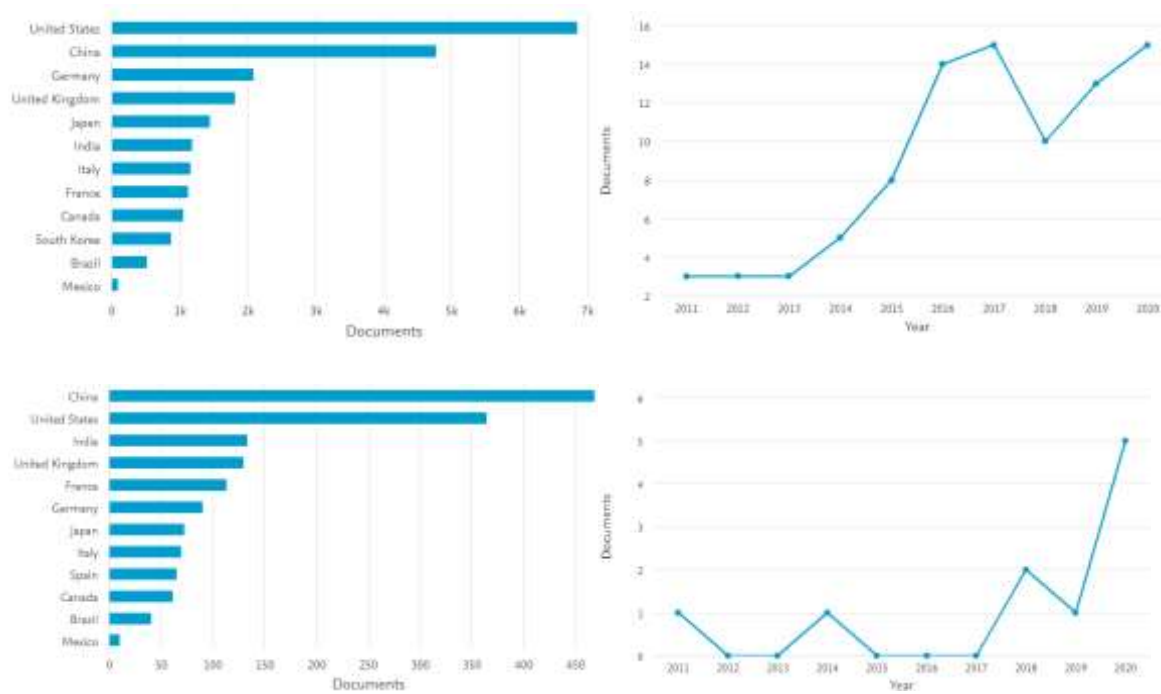


Figura 10. México ocupa el lugar 40 a nivel mundial de aportaciones en el campo de imagenología, mientras que una en temas relacionados con imagenología aplicada a cuerpos celulares, se ocupa la posición 34, de acuerdo con la búsqueda realizada.

En este sentido, México ha contribuido con solo 10 aportaciones en este campo, lo que puede estar relacionado con el hecho de ser un país en vías de desarrollo, lo que explicaría por qué 8 de las 10 contribuciones se realizaron entre 2018 y 2020.

CONCLUSIÓN

Se obtuvo una precisión de clasificación de 76.65% y 82.6% para detección de cuerpos celulares mononucleares y binucleares en presencia de cuerpos no celulares. Se identifica un 74.45% de los eventos validados manualmente y se presenta una efectividad para la determinación de micronúcleos, yemas y puentes celulares de 76.47%, 33.33% y 47.92%, respectivamente.

El acercamiento aquí presentado, clasifica a partir de cuerpos celulares no aislados ni pre etiquetados, teñidos con colorante no fluorescente y en múltiples condiciones de captura, de forma tal que se reduzca la necesidad de un preprocesamiento manual. Logra identificar los cuerpos celulares de interés respecto a artefactos de similares características, que se encuentran presentes comúnmente en los campos visuales a evaluar. La herramienta generada es lo suficientemente robusta para resolver una buena identificación y clasificación de cuerpos celulares dentro de un banco de imágenes de condiciones de captura variada, sin la necesidad de un amplio catálogo de imágenes pre procesadas o aisladas, logrando reducir a un número limitado de muestras con alta capacidad de identificar cuerpos micro nucleares menor al 35% del total de eventos positivos a células.

La precisión obtenida para identificar cuerpos micro nucleares de interés permite consolidar el primer paso en la generación de un sistema especializado capaz de discriminar cuerpos de interés y clasificarlos eventos celulares con una alta precisión. Cuya robustez le permite operar con imágenes generadas por equipos no especializados de acceso universal, y trabajar con imágenes de forma amigable al usuario sin la necesidad de especializarse con el uso de herramientas de análisis manual o semiautomático de imágenes.

PERSPECTIVAS

Dentro de las perspectivas del trabajo a futuro se encuentran:

- Generación de nueva biblioteca de imágenes más amplia de imágenes iniciales para mejorar el sistema de segmentado y validar su efectividad en campos con alta concentración de cuerpos celulares.
- Rentrenar el sistema para la identificación de nuevas morfologías celulares en diferentes bancos de imágenes.
- Implementación de red neuronal rentrenada para la identificación de morfologías de LSPH de forma directa.
- Mejorar el sistema de detección de probabilidad de aberraciones genéticas a un sistema de detección totalmente automático
- Mejorar el sistema de detección de aberraciones genéticas para su identificación directa por medio de una red neuronal
- Analizar la pertinencia de generar una red neuronal desde cero.

Así mismo, cabe señalar que una posible aplicación futura es la incorporación de un sistema para estimar el tamaño real de los cuerpos celulares en función de la imagen, para ello sería necesario la implementación de un soporte que mejore la homogeneidad en la distancia de captura de la imagen respecto al ocular.

REFERENCIAS

- Acevedo, A., Alférez, S., Merino, A., Puigví, L., & Rodellar, J. (2019). Recognition of peripheral blood cell images using convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 180. <https://doi.org/10.1016/j.cmpb.2019.105020>
- Adão, M. M., Guimarães, S. J. F., & Patrocínio, Z. K. G. (2020). Learning to realign hierarchy for image segmentation. *Pattern Recognition Letters*, 133, 287–294. <https://doi.org/10.1016/j.patrec.2020.03.010>
- Aich, S., Yamazaki, M., Taniguchi, Y., & Stavness, I. (2020). Multi-Scale Weight Sharing Network for Image Recognition. *Pattern Recognition Letters*, 131, 348–354. <https://doi.org/10.1016/j.patrec.2020.01.011>
- Amin, M., Borjali, A., & Del, G. (2021). A scoping review of transfer learning research on medical image analysis using ImageNet. *Computers in Biology and Medicine*, 128(November 2020), 104115. <https://doi.org/10.1016/j.compbiomed.2020.104115>
- Andrade, A. R., Vogado, L. H. S., Veras, R. de M. S., Silva, R. R. V., Araujo, F. H. D., & Medeiros, F. N. S. (2019). Recent computational methods for white blood cell nuclei segmentation: A comparative study. In *Computer Methods and Programs in Biomedicine* (Vol. 173, pp. 1–14). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.cmpb.2019.03.001>
- Andreini, P., Bonechi, S., Bianchini, M., Mecocci, A., & Scarselli, F. (2020). Image generation by GAN and style transfer for agar plate image segmentation. *Computer Methods and Programs in Biomedicine*, 184. <https://doi.org/10.1016/j.cmpb.2019.105268>
- Araújo, F. H. D., Silva, R. R. V., Ushizima, D. M., Rezende, M. T., Carneiro, C. M., Campos Bianchi, A. G., & Medeiros, F. N. S. (2019). Deep learning for cell image segmentation and ranking. *Computerized Medical Imaging and Graphics*, 72, 13–21. <https://doi.org/10.1016/j.compmedimag.2019.01.003>
- Baeyens, A., Swanson, R., Herd, O., Ainsbury, E., Willem, P., Thierens, H., Slabbert, J. P., Vral, A., Baeyens, A., Swanson, R., Herd, O., Ainsbury, E., Baeyens, A., Swanson, R., Herd, O., Ainsbury, E., & Mabhengu, T. (2011). A semi-automated micronucleus-centromere assay to assess low-dose radiation exposure in human lymphocytes A semi-automated micronucleus-centromere assay to assess low-dose radiation exposure in human lymphocytes. *International Journal of*

- Radiation Biology*, 3002(87:9), 923–931. <https://doi.org/10.3109/09553002.2011.577508>
- Biswas, S., & Hazra, R. (2020). A new binary level set model using L0 regularizer for image segmentation. *Signal Processing*, 174. <https://doi.org/10.1016/j.sigpro.2020.107603>
- Bolognesi, C., Balia, C., Roggeri, P., Cardinale, F., Bruzzi, P., Sorcinelli, F., Lista, F., Amelio, R. D., & Righi, E. (2011). Micronucleus test for radiation biodosimetry in mass casualty events : Evaluation of visual and automated scoring. *Radiation Measurements*, 46(2), 169–175. <https://doi.org/10.1016/j.radmeas.2010.11.003>
- Bolognesi, C., Bruzzone, M., Ceppi, M., & Marcon, F. (2020). Micronuclei and upper body cancers (head, neck, breast cancers) a systematic review and meta-analysis. *Mutation Research-Reviews in Mutation Research*, 108358. <https://doi.org/10.1016/j.mrrev.2020.108358>
- Braiki, M., Benzinou, A., Nasreddine, K., & Hymery, N. (2020). Automatic Human Dendritic Cells Segmentation Using K-Means Clustering and Chan-Vese Active Contour Model. *Computer Methods and Programs in Biomedicine*, 195. <https://doi.org/10.1016/j.cmpb.2020.105520>
- Deepthy Mary Alex, Hepzibah Christinal, A., Abraham Chandy, D., Singh, A., & Pushkaran, M. (2020). Speckle noise suppression in 2D ultrasound kidney images using local pattern based topological derivative. *Pattern Recognition Letters*, 131, 49–55. <https://doi.org/10.1016/j.patrec.2019.12.005>
- Delannoy, Q., Pham, C. H., Cazorla, C., Tor-Díez, C., Dollé, G., Meunier, H., Bednarek, N., Fablet, R., Passat, N., & Rousseau, F. (2020). SegSRGAN: Super-resolution and segmentation using generative adversarial networks — Application to neonatal brain MRI. *Computers in Biology and Medicine*, 120. <https://doi.org/10.1016/j.combiomed.2020.103755>
- Fenech, M., Kirsch-volders, M., Rossnerova, A., Sram, R., Romm, H., Bolognesi, C., Ramakumar, A., Soussaline, F., Schunck, C., Elhajouji, A., Anwar, W., & Bonassi, S. (2013). project initiative and review of validation , quality control and prospects for further development of automated micronucleus assays using image cytometry systems. *International Journal of Hygiene and Environmental Health*, 216(5), 541–552. <https://doi.org/10.1016/j.ijheh.2013.01.008>
- Frieauff, W., Cordier, A., & Suter, W. (1998). Automatic analysis of the in vitro micronucleus test on V79 cells. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 413(1), 57–68. [https://doi.org/10.1016/S1383-5718\(98\)00016-3](https://doi.org/10.1016/S1383-5718(98)00016-3)

- Gómez Flores, W., Pereira, W. C. de A., & Infantosi, A. F. C. (2014). Breast Ultrasound Despeckling Using Anisotropic Diffusion Guided by Texture Descriptors. *Ultrasound in Medicine and Biology*, 40(11), 2609–2621. <https://doi.org/10.1016/j.ultrasmedbio.2014.06.005>
- Hollandi, R., Szkalitsy, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbe, M., Kovacs, M., Migh, E., Goodman, A., Balassa, T., Koos, K., Wang, W., Caicedo, J. C., Bara, N., Kovacs, F., ... Horvath, P. (2020). nucleAlzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. *Cell Systems*, 10(5), 453-458.e6. <https://doi.org/10.1016/j.cels.2020.04.003>
- Hülber, T., Kocsis, Z. S., Kis, E., Sáfrány, G., & Pesznyák, C. (2020). A scanning and image processing system with integrated design for automated micronucleus scoring. *International Journal of Radiation Biology*, 96(5), 628–641. <https://doi.org/10.1080/09553002.2020.1722863>
- Kaur, S., & Sahambi, J. S. (2016). Curvelet initialized level set cell segmentation for touching cells in low contrast images. *Computerized Medical Imaging and Graphics*, 49, 46–57. <https://doi.org/10.1016/j.compmedimag.2016.01.002>
- Kirsch-volders, M., Fenech, M., & Bolognesi, C. (2018). Validity of the Lymphocyte Cytokinesis-Block Micronucleus Assay (L-CBMN) as biomarker for human exposure to chemicals with different modes of action: A synthesis of systematic reviews. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 836(January), 47–52. <https://doi.org/10.1016/j.mrgentox.2018.05.010>
- Kleczek, P., Jaworek-Korjakowska, J., & Gorgon, M. (2020). A novel method for tissue segmentation in high-resolution H&E-stained histopathological whole-slide images. *Computerized Medical Imaging and Graphics*, 79. <https://doi.org/10.1016/j.compmedimag.2019.101686>
- Liu, L., Cheng, J., Quan, Q., Wu, F. X., Wang, Y. P., & Wang, J. (2020). A survey on U-shaped networks in medical image segmentations. *Neurocomputing*, 409, 244–258. <https://doi.org/10.1016/j.neucom.2020.05.070>
- Lovell, D. P., Fellow, M., Elhajouji, A., Farabaugh, C. S., Gilby, B. G., Hashimoto, K., Li, Y., Roy, S., Schuler, M., Whitwell, J., & Tanir, J. Y. (2019). Analysis of historical negative control group data from the in vitro micronucleus assay using human lymphocytes. *Mutat Res Gen Tox En*,

837(June 2018), 52–59. <https://doi.org/10.1016/j.mrgentox.2018.08.009>

Mbogba, M. K., Haider, Z., Hossain, S. M. C., Huang, D., Memon, K., Panhwar, F., Lei, Z., & Zhao, G. (2018). The application of convolution neural network based cell segmentation during cryopreservation. *Cryobiology*, *85*, 95–104. <https://doi.org/10.1016/j.cryobiol.2018.09.003>

Murphy, S., Ellis-hutchings, R., Finch, L., Welz, S., & Wiench, K. (2018). In vitro genotoxicity studies : n-Butyl acrylate L5178Y mouse lymphoma (TK p / À locus assay), 2-Ethylhexyl acrylate gene mutation assay in Chinese hamster V79 cells , and 2-Ethylhexyl acrylate micronucleus test in human lymphocytes. *Data in Brief*, *20*, 316–325. <https://doi.org/10.1016/j.dib.2018.06.008>

OCDE, O. (2016). *OECD/OCDE 487. July*. <https://doi.org/10.1787/9789264264861-en>

Rahmat, R., Saednia, K., Haji Hosseini Khani, M. R., Rahmati, M., Jena, R., & Price, S. J. (2020). Multi-scale segmentation in GBM treatment using diffusion tensor imaging. *Computers in Biology and Medicine*, *123*. <https://doi.org/10.1016/j.combiomed.2020.103815>

Roemer, E., Zenzen, V., Conroy, L. L., Luedemann, K., Schunck, C., Sticken, E. T., Roemer, E., Zenzen, V., Conroy, L. L., Luedemann, K., Dempsey, R., Schunck, C., Trelles, E., Automation, S., Roemer, E., Zenzen, V., Conroy, L. L., Luedemann, K., Dempsey, R., & Schunck, C. (2015). Automation of the in vitro micronucleus and chromosome aberration assay for the assessment of the genotoxicity of the particulate and gas – vapor phase of cigarette smoke Automation of the in vitro micronucleus and chromosome aberration assay for the asse. *Toxicology Mechanisms and Methods*, *6516(25:4)*, 320–333. <https://doi.org/10.3109/15376516.2015.1037413>

Sert, E., Özyurt, F., & Doğantekin, A. (2019). A new approach for brain tumor diagnosis system: Single image super resolution based maximum fuzzy entropy segmentation and convolutional neural network. *Medical Hypotheses*, *133*. <https://doi.org/10.1016/j.mehy.2019.109413>

Shen, Y., Chen, J., Xiao, L., & Pan, D. (2019). Optimizing multiscale segmentation with local spectral heterogeneity measure for high resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *157*, 13–25. <https://doi.org/10.1016/j.isprsjprs.2019.08.014>

- Smith, K., Piccinini, F., Balassa, T., Koos, K., Danka, T., Azizpour, H., & Horvath, P. (2018a). Phenotypic Image Analysis Software Tools for Exploring and Understanding Big Image Data from Cell-Based Assays. *Cell Systems*, 6(6), 636–653.
<https://doi.org/10.1016/J.CELS.2018.06.001>
- Smith, K., Piccinini, F., Balassa, T., Koos, K., Danka, T., Azizpour, H., & Horvath, P. (2018b). Phenotypic Image Analysis Software Tools for Exploring and Understanding Big Image Data from Cell-Based Assays. *Cell Systems*, 6(6), 636–653.
<https://doi.org/10.1016/J.CELS.2018.06.001>
- Sram, R. J., Svecova, V., & Rossnerova, A. (2016). Systematic review of the use of the lymphocyte cytokinesis-block micronucleus assay to measure DNA damage induced by exposure to polycyclic aromatic hydrocarbons. *Mutation Research-Reviews in Mutation Research*, 770, 162–169. <https://doi.org/10.1016/j.mrrev.2016.07.009>
- Swee, V., Goh, T., Nakayama, R., Blakely, W. F., Abe, Y., En, C., Chua, L., Chew, Z. H., Nakata, A., Fujishima, Y., Yoshida, M. A., Kasai, K., Ariyoshi, K., & Miura, T. (2021). Improved harvest and fixation methodology for isolated human peripheral blood mononuclear cells in cytokinesis-block micronucleus assay. *International Journal of Radiation Biology*, 0(0), 1–14.
<https://doi.org/10.1080/09553002.2021.1844338>
- Taghi, M., Toossi, B., Azimian, H., Sarrafzadeh, O., Mohebbi, S., & Soleymanifard, S. (2017). Automatic detection of micronuclei by cell microscopic image processing. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 806(July), 9–18.
<https://doi.org/10.1016/j.mrfmmm.2017.07.012>
- Tareef, A., Song, Y., Cai, W., Huang, H., Chang, H., Wang, Y., Fulham, M., Feng, D., & Chen, M. (2017). Automatic segmentation of overlapping cervical smear cells based on local distinctive features and guided shape deformation. *Neurocomputing*, 221, 94–107.
<https://doi.org/10.1016/j.neucom.2016.09.070>
- Wang, Q., Chang, L., Zhou, M., Li, Q., Liu, H., & Guo, F. (2016). A spectral and morphologic method for white blood cell classification. *Optics and Laser Technology*, 84, 144–148.
<https://doi.org/10.1016/j.optlastec.2016.05.013>
- Wang, W., Taft, D. A., Chen, Y. J., Zhang, J., Wallace, C. T., Xu, M., Watkins, S. C., & Xing, J. (2019).

Learn to segment single cells with deep distance estimator and deep cell detector.

Computers in Biology and Medicine, 108, 133–141.

<https://doi.org/10.1016/j.compbiomed.2019.04.006>

Zhang, L., Kong, H., Chin, C. T., Liu, S., Chen, Z., Wang, T., & Chen, S. (2014). Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts.

Computerized Medical Imaging and Graphics, 38(5), 369–380.

<https://doi.org/10.1016/j.compmedimag.2014.02.001>

Zhang, X., Xiao, P., & Feng, X. (2020). Object-specific optimization of hierarchical multiscale segmentations for high-spatial resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 308–321.

ISPRS Journal of Photogrammetry and Remote Sensing, 159, 308–321.

<https://doi.org/10.1016/j.isprsjprs.2019.11.009>

Zhao, L., Li, K., Wang, M., Yin, J., Zhu, E., Wu, C., Wang, S., & Zhu, C. (2016). Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search MRF.

Computers in Biology and Medicine, 71, 46–56.

<https://doi.org/10.1016/j.compbiomed.2016.01.025>

Zhou, Y., Zang, H., Xu, S., He, H., Lu, J., & Fang, H. (2019). An iterative speckle filtering algorithm for ultrasound images based on bayesian nonlocal means filter model. *Biomedical Signal Processing and Control*, 48, 104–117. <https://doi.org/10.1016/j.bspc.2018.09.011>

Biomedical Signal Processing and Control, 48, 104–117. <https://doi.org/10.1016/j.bspc.2018.09.011>