

PROCESAMIENTO DIGITAL DE IMÁGENES CITOGÉNÉTICAS PARA SU CLASIFICACIÓN SEGÚN LA TÉCNICA ENSAYO COMETA PARA LA DETECCIÓN DE DAÑOS EN ELADN.

Vera Laceiras María Silvia ¹, Pisarello María Inés ², Caffetti Jacqueline Diana ^{1,3,4}.

¹Facultad de Exactas Químicas y Naturales. Universidad Nacional de Misiones. Posadas, Misiones.

²Facultad de Ciencias Exactas y Naturales y Agrimensura. Universidad Nacional del Nordeste. Corrientes. ³Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de Córdoba. Córdoba.

⁴Becaria Postdoctoral CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Instituto de Biología Subtropical (UNaM-IBS-CONICET).

vlhsilvia@gmail.com mainespisarello@exa.unne.edu.ar jacqui_caffetti@fceqyn.unam.edu.a

RESUMEN:

En este trabajo se estudian y analizan diferentes técnicas de clasificación, segmentación y extracción de contornos de imágenes digitales de origen biológico. En particular se centra la atención en los procesos que utilizan funciones con algoritmos neuro-fuzzy y de redes neuronales. Se presentan aquí algunos resultados logrados hasta el momento a través del desarrollo e implementación de un prototipo aplicado a imágenes obtenidas mediante microscopios, específicamente a aquellas parametrizadas según la técnica de ensayo cometa empleada por el Laboratorio de Citogenética General y Monitoreo Ambiental de la UNaM-IBS-CONICET para la detección de daños en el ADN con validación visual y numérica que avalan lo logrado.

Palabras claves: imágenes microscópicas, patrones de imágenes, algoritmos de segmentación, lógica difusa, extracción de contornos, redes neuronales, clasificación de imágenes.

CONTEXTO

La línea de investigación en que se enmarca el proyecto corresponde a la propuesta de tesis para acceder al título de Magíster en Tecnologías de la Información dictado conjuntamente por la Universidad Nacional de Misiones (UNaM) y la Universidad Nacional del Nordeste (UNNE). Las muestras pertenecen al Laboratorio de Citogenética General y Monitoreo Ambiental que se emplaza en la FCEQyN UNaM. El área de Monitoreo Ambiental analiza el daño genético en organismos acuáticos y su impacto potencial sobre la salud humana, como consecuencia de la exposición a contaminantes urbanos e industriales presentes en los ríos y arroyos de la Provincia de Misiones. Se desarrolla el procesamiento

digital de imágenes citogenéticas en Ensayo Cometa usando para este estudio, imágenes de moléculas de ADN donde el nucleóide de ADN se presenta como una cabeza y/o cola de cometa. Se busca distinguir cabeza, cola y fondo de la molécula de ADN y medirlas para su clasificación, a fin de facilitar el trabajo del Laboratorio. La investigación se inicia con la posibilidad de aplicar técnicas de clasificación, segmentación, algoritmos neuro fuzzy, matemáticos y computacionales a las imágenes digitales producidas por microscopio y registradas a través de una cámara de fotografía.

INTRODUCCIÓN

Descripción de actividades del Laboratorio de Citogenética General y Monitoreo Ambiental, Instituto de Biología Subtropical; Facultad de Ciencias Exactas, Químicas y Naturales; Universidad Nacional de Misiones; CONICET (UNAM-IBS-CONICET): El Laboratorio de Citogenética General y Monitoreo Ambiental inició sus actividades en el año 1989, dedicándose a estudios citogenético-evolutivos en diferentes modelos animales con principal énfasis en peces neotropicales de agua dulce. Luego se incorporó la línea de citogenética humana, prestando servicios de diagnóstico en convenio entre la UNaM y el Instituto de Previsión Social de la Provincia de Misiones (IPS). Y a partir del año 1993 se incluyó la línea de mutagénesis y monitoreo ambiental, que tiene como objetivos estudiar el impacto de contaminantes en ambientes naturales y en bioensayos de laboratorio a través de técnicas citogenético-moleculares. Actualmente, el laboratorio está constituido por su director, el Dr. Alberto Fenocchio; docentes-investigadores: Dra. Jacqueline Caffetti y Lic. Héctor Roncati; El laboratorio en su línea de investigación referida a agentes contaminantes acuáticos, aplica técnicas específicas de genética toxicológica, como el Ensayo Cometa, a diferentes organismos expuestos a muestras de agua tomadas de los Ríos Paraná y Uruguay. En el área de

Monitoreo Ambiental del laboratorio se analiza el daño genético en organismos acuáticos y su impacto sobre la salud humana, como consecuencia de la exposición a contaminantes urbanos e industriales presentes en los ríos y arroyos de la Provincia de Misiones.

Para ello se aplican técnicas específicas de genética toxicológica, como el Ensayo Cometa, uno de los “biomarcadores genéticos” [1] que son útiles como señales de alerta temprana en cursos de agua contaminados y, por lo tanto, su análisis resulta de interés predictivo en evaluaciones de estado de calidad de cursos hídricos y manejo de cuencas.

El Ensayo Cometa es una técnica donde se cuentan 100 células por cada individuo (de un total de 10-15 individuos por tratamiento). Estas 100 células se clasifican en distintos subtipos de acuerdo al daño que presenten en su ADN. Las 5 clases, que se muestran en la figura 1, dependen de la intensidad de fluorescencia y el largo de las “colas de los cometas” que son equivalentes a la cantidad de ADN fragmentado o con daño: Clase 0 (sin daño, es decir, no tiene cola); Clase 1 (tamaño de la cola hasta una vez el diámetro de la cabeza); clase 2 (tamaño de la cola hasta dos veces el diámetro de la cabeza); clase 3 (tamaño de la cola hasta tres veces el diámetro de la cabeza) y clase 4 (casi todo el ADN aparece fragmentado en la cola).

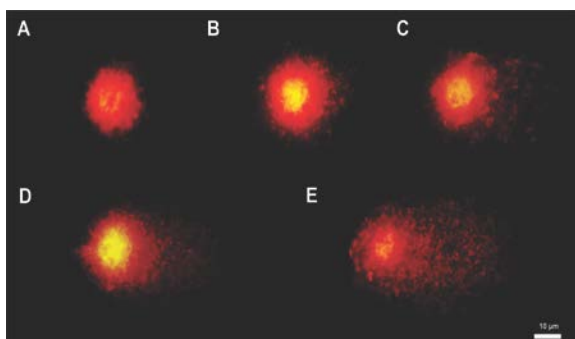


Figura 1. Clasificación de células en el ensayo cometa. A: Clase 0; B: clase 1; C: Clase 2; D: Clase 3; E: Clase 4.

El registro y clasificación de estos tipos celulares se hace de manera artesanal a través de procedimientos visuales y luego se realiza un registro manual. Una vez que se clasifican todas las células en las 5 clases, debe calcularse un índice de daño promedio (ID) o escore de daño por cada individuo y, a su vez, por

cada tratamiento (promediando los datos de los 10-15 individuos).

El registro manual que lleva adelante el Laboratorio, incluye anotaciones en cuadernos del número de células en cada clase y luego pasar los datos a una planilla de Excel para el cálculo del score o ID. Recién a partir de esta instancia los datos son cargados al programa Infostat® para su análisis estadístico.

De la bibliografía consultada se pueden citar un gran número de técnicas y desarrollos que se encuentran en constante perfeccionamiento, entre ellos la clasificación de una imagen tratándola como un objeto [2]; la segmentación por color [3]; la extracción de contornos en imágenes digitales es una operación que facilita también los procesos de segmentación e identificación de patrones [4], ya sea para tareas de reconocimiento e interpretación, como también de clasificación de objetos [5]. Las imágenes se capturan con cámara fotográfica digital y microscopio en el Laboratorio de Citogenética General y Monitoreo Ambiental. El objetivo de este trabajo es el procesamiento digital de dichas imágenes, para extraer información útil que pueda parametrizarse según la técnica de Ensayo Cometa, a fin de clasificar las moléculas ADN y realizar un dataset válido a fin de tener registros históricos de los datos recabados y lograr nueva información y nuevos conocimientos.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Actualmente, las líneas de investigación están basadas en la utilización de algoritmos matemáticos y computacionales, redes neuronales [6], [7] y algoritmos neurofuzzy [5], [8] que nos permiten manejar las características del material biológico, logrando:

- Almacenar cada imagen en un archivo para su posterior procesamiento digital.
- Identificación y detección de imágenes a través de algoritmos de aprendizaje supervisado [8], [9].
- Determinar las diferentes zonas: Fondo, cabeza y cola a través de filtros digitales, de segmentación de color que se logra al

transformar las imágenes por medio de funciones matemáticas y algoritmos computacionales combinados con algoritmos neuro fuzzy y redes neuronales

- Para la mejora del análisis de los resultados obtenidos mediante ensayo cometa (captura, clasificación, conteo y gestión de los resultados). Se usarán técnicas digitales que combinan algoritmos matemáticos y computacionales que incluyen fuzzy logic [10] y redes neuronales [11],[12].

RESULTADOS OBTENIDOS

Actualmente se estudiaron y analizaron más de 200 imágenes obtenidas a partir de células de diferentes especies luego de la aplicación del Ensayo cometa, logrando realizar primeramente un dataset que nos permite tener un histórico para nuevos estudios y una clasificación a través de un prototipo creado y validado con las imágenes de prueba.

Con la aplicación de técnicas de clasificación de imágenes que tienen un muy bajo margen de error en cuanto a la detección de clases (a través de la detección y distinción de fondo, cabeza y cola, su medición y posterior clasificación) se logra menor intervención humana a la hora de procesarlos. En este caso el desarrollo del prototipo funcional se realiza en Matlab y se puede ejecutar en Octave. La posibilidad de ejecución de los algoritmos en Octave, software libre de distribución gratuita, es lo que permite avanzar con el prototipo y aportar una mejor performance en la clasificación y los recursos más económicamente convenientes a la hora de evaluar la implementación en el laboratorio de Citogenética (UNaM-IBS- CONICET).

El desempeño final del sistema se evalúa mediante inspección visual y métodos de validación cruzada de datos. Utilizando K grupos (K-Fold) de datos en los que estén representados diferentes tipos de clases y de diferentes especies en cada grupo. En (K-Fold) donde K representa las interacciones, el error se calcula de la siguiente manera:

$$E_i = \frac{\text{numero de clasificaciones incorrectas}}{\text{total de datos del grupo test}}$$

El error final de la validación es:

$$Error = \frac{1}{K} \sum_{i=1}^k E_i$$

Se obtiene el siguiente error promedio: **Error 7.4%**

Conformado de la siguiente manera en cuanto a la precisión:

Clase	Error de Precisión
Clase 0	0,00%
Clase 1	8.18%
Clase 2	14.3%
Clase 3	14,00%
Clase 4	6.2%

Se espera finalizar la investigación con el desarrollo de una aplicación para celulares que nos permita capturar, gestionar, clasificar y guardar las imágenes para lograr una manipulación de imágenes que se adapte a las especificaciones del Laboratorio de Citogenética General y Monitoreo Ambiental (UNaM-IBS-CONICET) en la detección de daños en el material genético, planteando como línea de investigación futura la posibilidad de trabajar en otros proyectos que actualmente se llevan a cabo en dicho Laboratorio.

FORMACIÓN DE RECURSOS HUMANOS

Este proyecto de investigación forma parte del desarrollo de la tesis de posgrado, correspondiente a la carrera de Magíster en Tecnologías de la Información dictada por la Universidad Nacional de Misiones y la Universidad Nacional del Nordeste. La tesis tiene como estudio de caso una técnica de análisis utilizada en el Laboratorio, el Ensayo Cometa, que se encuentra bajo la dirección de la Dra. Maria Inés Pisarello y se articula con la línea de investigación de la Docente Investigadora en el Instituto de Biología Subtropical (UNaM-IBS-CONICET) que lleva adelante la codirección de la misma. Esta tesis de maestría no solo aporta soluciones reales al laboratorio de citogenética en su línea de investigación y a las cátedras donde se incluye el aprendizaje de la técnica de ensayo cometa; sino que además en el campo profesional aportan formación de recursos humanos especializados en técnicas de inteligencia artificial.

Bibliografía:

- [1] G. N. A. Furnus, J. D. Caffetti, E. M. García, M. F. Benítez, M. C. Pastori, y A. S. Fenocchio, «Frequências basais de micrónúcleos e anormalidades nucleares
- [2] A. Kaplan y M. Haenlein, «Rulers of the world, unite! The challenges and opportunities of artificial intelligence», *Bus. Horiz.*, 2019.
- [3] P. Kickingeder *et al.*, «Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study», *Lancet Oncol.*, vol. 20, n.º 5, pp. 728-740, 2019.
- [4] D. Sebasti Comas y G. J. Meschino, «Segmentación de Imágenes mediante Reconocimiento de Patrones», *Esc. y Work. Argentino en Ciencias las Imagenes*, 2014.
- [5] S. K. Oh, W. Pedrycz, y B. J. Park, «Self-organizing neurofuzzy networks based on evolutionary fuzzy granulation», *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans.*, vol. 33, n.º 2, pp. 271-277, mar. 2003.
- [6] A. P. Boyle, J. Guinney, G. E. Crawford, T. S. Furey, y A. Valencia, «F-Seq: a feature density estimator for high-throughput sequence tags», *Bioinforma. Appl. NOTE*, vol. 24, n.º 21, pp. 2537-2538, 2008.
- [7] «Redes neuronales profundas - Tipos y Características - Código Fuente». [En línea]. Disponible en: <https://www.codigofuente.org/redes-neuronales-profundas-tipos-caracteristicas/>. [Accedido: 20-dic-2019].
- [8] D. Ciregan, U. Meier, y J. Schmidhuber, «Multi-column deep neural networks for image classification», en *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642-3649.
- [9] «Learning machine, los usos del aprendizaje supervisado». [En línea]. Disponible en: <https://blog.es.logicalis.com/analytics/learning-machine-los-usos-del-aprendizaje-supervisado>. [Accedido: 20-dic-2019].
- [10] M. Landín, R. C. Rowe, y P. York, «Advantages of neurofuzzy logic against conventional experimental design and statistical analysis in studying and developing direct compression formulations», *Eur. J. Pharm. Sci.*, vol. 38, n.º 4, pp. 325-331, nov. 2009.
- [11] C. Quintero, F. Merchán, A. Cornejo, y J. S. Galán, «Uso de Redes Neuronales Convolucionales para el Reconocimiento Automático de Imágenes de Macroinvertebrados para el Biomonitorio Participativo», *KnE Eng.*, vol. 3, n.º 1, p. 585, feb. 2018.
- [12] M. Polycarpou, A. C. P. L. F. de Carvalho, J.-S. Pan, M. Woźniak, H. Quintian, y E. Corchado, Eds., *Hybrid Artificial Intelligence Systems*, vol. 8480. Cham: Springer International Publishing, 2014.