

Universidad Nacional de Misiones
Facultad de Ciencias Exactas Químicas y Naturales

**“Trabajo Final de Maestría en Tecnologías de la
información”**

**Procesamiento digital de imágenes
citogenéticas para su clasificación.**

Autor: Lic. María Silvia Vera Laceiras

Director: Dra. María Ines Pisarello
Co-Director: Dra. Jacqueline Caffetti

Año 2020

Procesamiento digital de imágenes citogenéticas para su clasificación

Dedico este trabajo a mi gente querida que siempre me impulsa a dar lo mejor.

Resumen

Este trabajo final de maestría tiene como objetivo el desarrollo de una herramienta bioinformática que permita detectar a través de reconocimiento de imágenes, nucleoides de ADN (con cabezas y colas) en fluorescencia, para su posterior clasificación de acuerdo con la técnica de ensayo cometa. Se busca automatizar y gestionar su almacenamiento, y optimizar los procesos en el laboratorio de citogenética.

Para tal fin, se evalúan posibles métodos a aplicar y se seleccionan los que combinan algoritmos matemáticos, computacionales, redes neuronales y sistema neuro difuso.

El método desarrollado realiza la detección y segmentación en tres pasos:

- Primero: se lleva a cabo un pre proceso y una segmentación inicial de la imagen en bruto.
- Segundo: los fragmentos obtenidos se clasifican mediante redes neuronales en tres grupos: cabeza, cola y fondo.
- Tercero: las cabezas o núcleos y colas se miden y son re-analizadas, clasificándolas según su relación de proporción.

Este proceso de detección, segmentación y clasificación se comprobó mediante el caso de estudio con imágenes citogenéticas en ensayo cometa del Laboratorio de Citogenética General y Monitoreo Ambiental del Instituto de Biología Subtropical UNaM-IBS-CONICET.

Los resultados expuestos a través de la evaluación del modelo dio un Promedio de aciertos (*Accuracy*) 0,92% y media armónica (*Fmeasure*) 0,96% fueron alcanzados a través de modelos que combinan métodos de clasificación mediante algoritmos híbridos, redes neuronales, sistema neurodifuso y funciones de clasificación.

Palabras claves: Ensayo Cometa, redes neuronales, algoritmos neurodifusos, Logica difusa, algoritmos de clasificación, procesamiento digital de imágenes, citogenética.

Abstract

The objective of this thesis is the development of a bioinformatics tool that allows to detect DNA nucleoides (with heads and tails) through the recognition of images in fluorescence, for its subsequent classification according to comet assay technique. It is sought to automate and manage their storage, and to optimize processes in the cytogenetic laboratory.

For this purpose, possible methods are evaluated and the ones selected combine mathematical and computer algorithms, neural networks and fuzzy.

The developed method performs detection and segmentation in three steps:

- First: a pre-process and initial segmentation of the raw image is carried out.
- Second: the fragments obtained are classified by means of a convolutional neural network (CNN) into three groups: head, tail and bottom.
- Third: heads or cores and tails are measured, re-analyzed, and classified according to its proportional ratio.

This detection, segmentation and classification process was verified by means of case study with cytogenetic images during the comet assay in General Cytogenetic and Environmental Monitoring Laboratory of Subtropical Biology Institute, UNaM-IBS-CONICET.

The results presented through the evaluation of the model gave accuracy of 0.92 and Fmeasure of 0.96 were models that combine classification methods that make use of hybrid algorithms, neuronal networks, neuro-fuzzy algorithms and classification functions.

Keywords: Comet assay, neural networks, neurofuzzy algorithms, fuzzy logic, classification algorithms, digital image processing, cytogenetics.

Agradecimientos

Agradezco a mis directoras de tesis, Dra María Inés Pisarello y Dra Jacqueline Caffetti por su tiempo y por asesorarme con calidez y precisión compartiendo sus saberes para lograr un buen trabajo.

A las autoridades y docentes de la maestría de la Facultad de Ciencias Exactas Químicas y Naturales de la UNaM y Facultad de Ciencias Exactas Naturales y Agrimensura de la UNNE que permitieron mi perfeccionamiento y crearon esta motivación en mí.

A mi gente querida, amigos y compañeros de estudio que hicieron este camino muy grato.

A mi familia que acompañó cada paso con amor y comprensión.

Y a Dios por darme paz y fortaleza para lograr mis objetivos a fin de que estos conocimientos puedan ser útiles a la sociedad.

Indice

Indice:

PORTADA.....	I
RESUMEN	III
ABSTRACT	V
AGRADECIMIENTOS	VI
INDICE.....	VII
CAPITULO 1.....	XIII
INTRODUCCIÓN	- 11 -
1.1 CONTEXTO Y JUSTIFICACIÓN:	- 11 -
1.2 OBJETIVOS DEL TRABAJO	- 13 -
1.3 ENFOQUE Y MÉTODO	- 14 -
1.4 BREVE DESCRIPCIÓN DE LOS OTROS CAPÍTULOS DE LA MEMORIA.....	- 16 -
CAPITULO 2- 18 - ESTADO DEL ARTE.....	- 18 -
2.1 ESTADO DEL ARTE DEL ENSAYO COMETA	19
2.2 ESTADO DEL ARTE DE IMÁGENES EN EL MONITOREO AMBIENTAL	22
CAPITULO 3 INTELIGENCIA ARTIFICIAL EN EL RECONOCIMIENTO DE PATRONES DE IMÁGENES.	27

Procesamiento digital de imágenes citogenéticas para su clasificación

3.1 PROBLEMAS PARA RECONOCER UNA IMAGEN.....	28
3.2 REDES NEURONALES	36
3.3 LOGICA DIFUSA	43
CAPITULO 4 DESARROLLO MATEMÁTICO Y RAZONAMIENTO DE LAS BASES DEL PROTOTIPO DE LA HERRAMIENTA BIOINFORMÁTICA	51
4.1 PANORAMA GENERAL:.....	52
4.2 DESARROLLO MATEMÁTICO DEL PREDICADO	61
4.3 CLASIFICADOR	63
4.4 AGRUPAMIENTO O CLASIFICACIÓN	67
4.5 REALCE DE CONTRASTE.....	69
4.6 CÓDIGO DEL CLASIFICADOR PARA MATLAB® Y OCTAVE.....	71
CAPITULO 5 RESULTADOS Y DISCUSIÓN:	75
5.1 RESULTADOS	76
5.2 PROPUESTA PARA INTEGRAR EQUIPAMIENTO COMPLEMENTARIO	90
CAPITULO 6 CONCLUSIONES	93
BIBLIOGRAFIA	99

ÍNDICE DE FIGURAS:

FIGURA 1: REPRESENTACIÓN GRÁFICA DE LOS PROCESOS CONECTIVOS DE LA NEURONA BIOLÓGICA MOSTRANDO LAS SINAPSIS	31
FIGURA 2: DIFERENTES CAPAS QUE ALIMENTAN A UNA NEURONA Y COMO SE ACUMULAN LAS SEÑALES.	32
FIGURA 3: FUNCIÓN DE ACTIVACIÓN SIGMOIDEA O LOGÍSTICA DE UNA NEURONA DONDE A ES LA SUMA DE LAS ENTRADAS POR SUS PESOS Y P LA FORMA DE LA FUNCIÓN.	33
FIGURA 4 FUNCIÓN DE UNIDAD LINEAL RECTIFICADA.	34
FIGURA 5 NEURONA Y LAS FUNCIONES ACTIVADORAS DE ACCIONES.....	34
FIGURA 6: DESARROLLO SONN BASADO EN DISCRIMINACIÓN DE CARACTERÍSTICAS. FUENTE: [37] "NEURAL NETWORK DESIGNED ON APPROXIMATE REASONING ARCHITECTURE AND THEIR APPLICATIONS HIDEYUKI TAKAGI E ISAO HAYASHI MATSUSHITA ELECTRIC INDUSTRIAL Co. SONN ADAPTA AUTOMÁTICA Y SIMULTÁNEAMENTE LA TOPOLOGÍA DEL NN Y TODOS SUS PESOS EN EL PROCESO DETERMINISTA	35
FIGURA 7: RAZONAMIENTO DIFUSO [4] H TAKAGI Y I. HAYASHI "NN DRIVEN FUZZY REASONNIG"	36
FIGURA 8: RED NEURONAL, DE PREALIMENTACION.	38
FIGURA 9: MODIFICACIÓN DE UNA SOLA NEURONA DESDE LOS VALORES DE ENTRADA, DE LOS PESOS Y LOS VALORES DE SALIDA LUEGO DE APLICAR LA FUNCIÓN.....	39
FIGURA 10: RED NEURONAL: CÁLCULOS PREVIOS AL ENTRENAMIENTO.	39

Procesamiento digital de imágenes citogenéticas para su clasificación

FIGURA 11: IMAGEN ILUSTRATIVA DE RED NEURONAL CON TRES CAPAS Y PESOS REPRESENTADOS CON PROPAGACIÓN HACIA ATRÁS.	40
FIGURA 12: NN DRIVEN FUZZY REASONING, H.TAKAGIY I.HAYASHI [4].	44
FIGURA 13: FUNCIÓN SIGMOIDEA EN REDES NEURONALES, RAZONAMIENTO DIFUSO (FUZZY) [4].	44
FIGURA 14: SIMULACIÓN DEL MODELO CON ACTIVACIÓN SIGMOIDEA	52
FIGURA 15: SIMULACIÓN CON 2 CAPAS OCULTAS PROBLEMA CON ACTIVACIÓN A)RELU Y ACTIVACIÓN B)TANH USANDO SIMULADOR DE REDES NEURONALES DE PLAYGROUNDTENSORFLOW	53
FIGURA 16: IMAGEN DE LA MOLÉCULA DE ADN LUEGO DE LA PREPARACIÓN EN ENSAYO COMETA Y EL MODELO DE ESTRATEGIA DE RECONOCIMIENTO DE PATRONES DE IMÁGENES. FUENTE PROPIA.	57
FIGURA 17: EJEMPLO DE RECONOCIMIENTO DE PATRONES DE IMÁGENES; LA IMAGEN CRUDA Y TRANSFORMADA PARA LOGRAR UN HISTOGRAMA UNIFORME. FUENTE PROPIA.	57
FIGURA 18: RECORRIDO DE LA IMAGEN A TRAVÉS DEL MODELO DE ENTRENAMIENTO DE RECONOCIMIENTO ADAPTATIVO PARA SU CLASIFICACIÓN. FUENTE PROPIA.	58
FIGURA 19: ADAPTACIÓN DEL MODELO. FUENTE PROPIA.	58
FIGURA 20: PRIMERA APROXIMACIÓN DE DETECCIÓN DE NÚCLEOS O NUCLEOIDES DE ADN. SE OBSERVA LA COMPLEJIDAD PARA DEFINIR LAS FRONTERAS Ó LIMITES.	59
FIGURA 21: OBSÉRVESE DOS IMÁGENES DE COMETA QUE PERTENECEN A LA MISMA CLASE.	60
FIGURA 22: CLASIFICACIÓN DE ACUERDO A LA COMPOSICIÓN MORFOLÓGICA CON RESPECTO AL MODELO. FUENTE PROPIA	60
FIGURA 23: DIFERENTES FOTOGRAFÍAS DE NUCLEOIDES DE ADN ANTES DEL PRE PROCESAMIENTO. SE PRESENTAN CON DIFERENCIAS NO SOLO DE CANTIDAD DE NÚCLEOS SINO TAMBIÉN DE LUZ Y POSICIÓN.	61
FIGURA 24 : MEDICIÓN DEL NÚCLEO DEL COMETA ANTES DE TRANSFORMAR A NIVELES DE GRIS. FOTOGRAFÍA TOMADA DEL PROGRAMA LUEGO DE DETERMINAR EL CENTROIDE FUENTE PROPIA.	62
FIGURA 25: COMO RECORRE LA IMAGEN LA MATRIZ DE CONVOLUCIÓN.	63
FIGURA 26: IMAGEN CLASE 0.EL INDIVIDUO A CLASIFICAR SE PRESENTA COMO UN COMETA SIN COLA.SE OBSERVA EL NUCLEO CON MUCHO BRILLO, LA COLA ALREDEDOR CON MENOR BRILLO Y EL FONDO DE COLOR NARANJA	64
FIGURA 27: IMAGEN CLASE 1, SE PUEDEN OBSERVAR LOS NÚCLEOS CON UN COLOR MÁS CLARO Y LA COLA TIENE TAMAÑO DEL DIÁMETRO DE LA CABEZA.	64
FIGURA 28: IMAGEN CLASE 2 LA COLA SE EXTIENDE MÁS DE DOS VECES EL DIÁMETRO DEL NUCLEOIDE.	65
FIGURA 29: IMAGEN CLASE 3 SE OBSERVA LA COLA QUE SE EXTIENDE 3 VECES EL DIÁMETRO DEL NUCLEOIDE.....	65
FIGURA 30: IMAGEN CLASE 4 SE OBSERVA EL NÚCLEO Y LA COLA LARGAMENTE DISGREGADA.....	65
FIGURA 31: MUESTRA DE LAS DIFERENTES CLASES DE COMETAS. LAS LETRAS A,B,C,D,E, SE CORRESPONDEN CON LAS CLASES 0, CLASE 1, CLASE 2, CLASE 3 Y CLASE 4 DONDE TAMBIÉN SE ENCUENTRAN LOS ANÓMALOS. FUENTE: FOTOGRAFÍA EDITADA LABORATORIO DE CITOGENÉTICA (UNAM-IBS-CONICET).....	66
FIGURA 32 APRENDIZAJE SUPERVISADO.	66
FIGURA 33 EJEMPLO DE EMPLEO DE UN VECTOR DE CARACTERÍSTICAS MULTIRESOLUCIÓN.	68
FIGURA 34: FOTOGRAFÍA TOMADA SOBRE EL MICROSCOPIO EN CONDICIONES DE TOTAL OSCURIDAD. IMAGEN CRUDA. FUENTE PROPIA.	68

Procesamiento digital de imágenes citogenéticas para su clasificación

FIGURA 35: IMAGEN LUEGO DEL PRE PROCESAMIENTO (IMAGEN PREPARADA EN ESCALA DE GRISES).FUENTE: IMAGEN MODIFICADA DE FOTOGRAFÍA TOMADA EN LABORATORIO DE CITOGENÉTICA (UNAM-IBS-CONICET) COMO RESULTADO DEL ALGORITMO DESARROLLADO.....	69
FIGURA 36: IMAGEN BINARIA (BLANCO Y NEGRO) DE LA MOLÉCULA DE ADN LUEGO DE APLICAR LA FUNCIÓN IMCLOSE.	70
FIGURA 37: IMAGEN BINARIA DE LA CÉLULA. SE PUEDE OBSERVAR ADEMÁS LA LÍNEA DEFINIDA POR EL ALGORITMO QUE MIDE EL DIÁMETRO DEL NUCLEOIDE MAS LA COLA.	71
FIGURA 38: IMÁGENES DE NUCLEOIDES DE ADN IMÁGENES LOGRADAS A TRAVÉS DE LAS FOTOGRAFÍAS.	80
FIGURA 39: IMÁGENES DE NUCLEOIDES DE ADN CON PREPROCESO OBTENEMOS IMÁGENES ECUALIZADAS.	80
FIGURA 40: SE OBSERVAN LAS IMÁGENES DE NUCLEOIDES DE ADN ANTES Y DESPUÉS DE LA TRANSFORMACIÓN A ESCALA DE GRISES.	81
FIGURA 41: IMÁGENES DE CABEZA Y COLA DEL NUCLEOIDE EN BLANCO Y NEGRO (BINARIO).....	81
FIGURA 42 IMÁGENES DEL MATERIAL BIOLÓGICO, PUEDE VERSE MUY DIFERENTE Y SIN EMBARGO PERTENECER A UNA MISMA CLASE.	82
FIGURA 43 CLASIFICACIÓN DE UN REGISTRO FOTOGRÁFICO DE UNA MUESTRA PERTENECIENTE A UN ANÓMALO. REALIZA EL MISMO PROCEDIMIENTO Y LO CLASIFICA COMO CLASE 4	83
FIGURA 44: GRÁFICO DE PRECISIÓN POR CLASES.....	85
FIGURA 45: GRAFICO DE EXHAUSTIVIDAD POR CLASES.	86
FIGURA 46: GRAFICO DE MATRIZ DE CONFUSIÓN.	87
FIGURA 47: SE OBSERVA EL COMPORTAMIENTO DEL ERROR DE PRUEBA Y DEL ERROR DE ENTRENAMIENTO CON CURVAS BASTANTES SIMILARES EN LAS QUE EL ERROR DECRECE.....	88
FIGURA 48: EJEMPLO DE DESARROLLO DE LA APLICACIÓN PARA EL MÓVIL DESDE LA MUESTRA FÍSICA USANDO EL PROTOTIPO DE DE CLASIFICACIÓN Y LLEGANDO A LA CLASIFICACIÓN INTELIGENTE DE LAS MUESTRAS. FUENTE PROPIA. ..	92

ÍNDICE DE TABLAS:

TABLA 1: FUNCIONES DE ACTIVACIÓN:	32
TABLA 2 VALORES DE FUNCIÓN PARA EFECTUAR LAS OPERACIONES DE RED NEURONAL.	37
TABLA 3 CUADRO DE ERRORES DE PRECISIÓN EN CADA CLASE.....	84
TABLA 4 CUADRO DE PRECISIÓN	85
TABLA 5 CUADRO DE EXHAUSTIVIDAD	85
TABLA 6 MATRIZ DE CONFUSIÓN	86
TABLA 7: RECALL Y PRECISIÓN	88

Índice de ecuaciones:

EC 1 $fz = 11 + e - z$	33
------------------------------	----

Procesamiento digital de imágenes citogenéticas para su clasificación

EC 2 $Rz = \text{Max}0, z$	34
EC 3 $F(X,Y,Z)=XY+z \leftarrow$	37
EC 4 $Z = wy0 \ xy0$	38
EC 5 $F(z) = 11 + e - z$	39
EC 6 $\mu_A^s (X_i) \in W_l^s, l=1,\dots,N.$	47
EC 7 $\theta ms = i = 1ncyi - usxi. \mu As xi^2$	48
EC 8 $\theta ms = i = 1nc\mu Axxiyi - usxi. \mu Asxi^2$	48
EC 9 $l^s=ncnts + nci = 1ntsyi - usxi^2 + ntsnts + ncj = 1ncyj - usxj. \mu Asxj^2$	48
EC 10 $\theta m - 1sp = i = 1ncyi - usxi. \mu Asxi^2, P=1,2,\dots,M$	48
EC 11 $yi * = s = 1r\mu Asxi. uixis = 1r\mu Asxi, l=1,2,\dots,N.$	49
EC 12 $Ei = \text{num de clasif. incorrectas total de datos del grupo test}$	84
EC 13 $Error = 1Ki = 1kEi$	84
EC 14 $TP + TNTP + TN + FP + FN$	87
EC 15 $\text{RECALL } TP + FN$	88
EC 16 $\text{PRECISION } TP + FP$	88
EC 17 $\text{FMEASURE} = TP + 12FP + FN$	89

1

Introducción

Introducción

1.1 Contexto y justificación:

El Laboratorio de Citogenética General y Monitoreo Ambiental del Instituto de Biología Subtropical en la Facultad de Ciencias Exactas, Químicas y Naturales; Universidad Nacional de Misiones; CONICET (UNAM-IBS-CONICET) 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 en distintos puntos de la Provincia de Misiones. Para tales fines, actualmente se emplean dos técnicas específicas, una es el Test de Ensayo Cometa o Electroforesis en gel de células individuales (single cell gel electrophoresis-SCGE) y el Test de Micronúcleos.

El Test que se amplía y desarrolla es el de Ensayo Cometa o Electroforesis en gel de células individuales, aplicado a organismos bioindicadores expuestos a muestras de agua tomadas del Río Paraná, en diversos tramos o puntos de la provincia.

El análisis y la clasificación visual artesanal de las imágenes obtenidas puede volverse tedioso y dependiente del genetista experto, además de ser susceptible de sesgos de interpretación cuando interviene más de un investigador en el proceso de observación. .

Para facilitar el trabajo del Laboratorio, se propone investigar la posibilidad de aplicar técnicas de clasificación y segmentación de las imágenes citogenéticas obtenidas mediante microscopio de epifluorescencia para el ensayo cometa, de manera automática, realizando además evaluación comparativa (*benchmarker*) con las diferentes opciones.

El Laboratorio de Citogenética General y Monitoreo Ambiental (UNAM-IBS-CONICET) 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). 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-

Procesamiento digital de imágenes citogenéticas para su clasificación

investigadores: Dra. Jacqueline Caffetti y Lic. Héctor Roncati; así como tesistas, becarios de grado y postgrado: Angemara Rau, Sergio Müller y Mari Florencia Rivero

El laboratorio en su línea de investigación referida a agentes contaminantes acuáticos, aplica técnicas específicas de genética toxicológica, como el Ensayo Cometa.

Estos denominados “biomarcadores genéticos”, 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[1].

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 5 clases dependiendo de la intensidad de fluorescencia del largo de las “colas de los cometas”. La fluorescencia que “tiñe” el ácido desoxirribonucleico (*ADN*) permite observar las manifestaciones 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, en esta clase se incluyen los anómalos que presentan características idénticas).

El registro y clasificación de estos tipos celulares se hace visualmente y de forma artesanal y luego se registra manualmente. Una vez que se clasifican todas las células en las 5 clases, debe calcularse el índice de daño en el *ADN*, (*score o ID*) 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, comienza en registros fotográficos tomados a través de cámaras adosadas al microscopio. Incluye anotaciones en cuadernos del número de células en cada clase y luego el pasaje de los datos a una planilla de cálculos para el cómputo del índice de daño de *ADN* (*score o ID*). Actualmente cuando se habla de tratamiento de las imágenes digitales, se pueden citar un gran número de técnicas y desarrollos que se encuentran en constante perfeccionamiento, entre ellos la capacidad de segmentación y clasificación de una imagen tratándola como un objeto.

La segmentación semántica a nivel de instancia o segmentación de instancia, por ejemplo, es la tarea de detectar y segmentar conjuntamente instancias individuales de objetos en una imagen [2].

Los sistemas neurodifusos [3], [4] permiten trabajar en imágenes digitales a través de una operación que facilita los procesos de segmentación e identificación de patrones, tanto para tareas de reconocimiento e interpretación [5], como también de clasificación de objetos [6].

Se inicia la investigación con la posibilidad de obtener imágenes digitales en el Laboratorio de Citogenética General y Monitoreo Ambiental a través del uso del microscopio de epifluorescencia y una cámara adosada.

El objetivo del procesamiento digital de dichas imágenes, que se plantea en esta investigación, es extraer información útil que pueda parametrizarse según la técnica de Ensayo Cometa.

1.2 Objetivos del trabajo

El objetivo principal de esta tesis es desarrollar un procedimiento que permita la captura, gestión y análisis de imágenes de muestras celulares, su auto clasificación en grupos de pertenencia y análisis de datos obtenidos en determinados lapsos (definidos por el usuario) para su posterior estudio en el Laboratorio de Citogenética del Instituto de Biología Subtropical de la Facultad de Ciencias Exactas, Químicas y Naturales, (IBS-UNaM-CONICET).

Para alcanzar el objetivo mencionado, se propone:

- Lograr la autclasificación de las imágenes obtenidas, para determinar los 5 grupos (considerando a los anómalos dentro de la clase 4), usando redes neuronales, algoritmos neuro difusos y finalmente funciones o algoritmos genéticos para su ajuste a fin de obtener datos estadísticos y precisos.
- Generar un procedimiento prototípico (desarrollo de una herramienta bioinformática con Matlab®) que permita definir claramente la manera de trabajar aplicando este tipo de reconocimiento de patrones de imágenes en las imágenes citogenéticas obtenidas al aplicar el ensayo cometa a través de microscopio de fluorescencia.

- Validar el prototipo a través del caso de estudio, en el Laboratorio de Citogenética General y Monitoreo Ambiental del Instituto de Biología Subtropical.

1.3 Enfoque y método

Para la mejora de los resultados de la observación de las imágenes obtenidas a través de Ensayo Cometa en microscopio de fluorescencia -captura, clasificación, y gestión de los resultados en el desarrollo del prototipo-, se proponen técnicas digitales que combinan algoritmos matemáticos y computacionales en una línea de estudio, que incluyen lógica difusa [7], [8], redes neuronales [9] [10], convolucionales [11], aprendizaje por transferencia, siguiendo el camino de inteligencia artificial (IA) [12], y aprendizaje profundo (*Deep Learning*) [2][9].

En esta implementación se seleccionará el método que realice el reconocimiento de patrones de imágenes de forma eficiente.

La misma técnica puede emplearse a otros casos de estudio, enriquecida con gestión del conocimiento [13]. Una vez adquirido el conocimiento, es posible aplicarlo para resolver otros problemas de la especialidad.

Hasta el momento, se estudiaron y analizaron diferentes fuentes de información para definir aquellas técnicas de clasificación de imágenes con menor intervención humana a la hora de procesarlos.

Se espera como producto de este trabajo final de maestría, el diseño y desarrollo de un prototipo basado en la técnica de manipulación de imágenes que más se adapte a las especificaciones del Laboratorio de Citogenética General y Monitoreo Ambiental de la UNaM-IBS-CONICET.

Para realizar este prototipo se evaluaron muchas técnicas, algunas fueron desestimadas porque el desarrollo serviría para un caso muy particular, otras porque eran aplicables a casos de perfil industrial, otras por ser muy costosa su implementación o muy engorroso el desarrollo. Finalmente el uso de las redes neuronales y sistemas de lógica difusa fueron la elección más simple de implementar y de llevar adelante mediante un prototipo muy funcional desarrollado en Matlab®.

Procesamiento digital de imágenes citogenéticas para su clasificación

Esta tecnología, de Inteligencia Artificial nos permite:

- Un *Aprendizaje Adaptativo*, o sea aprender basado en una experiencia inicial. Los ejemplos de entrenamiento son muy importantes.
- *Auto organización*: una red neuronal puede crear su propia organización mediante el aprendizaje.
- *Tolerancia a fallos*: si alguna parte de la red es dañada (aun si es grande el daño) se pueden recuperar igualmente algunas funciones ya aprendidas y solo se degrada el sistema, no se pierde todo.
- *Operación en tiempo real y fácil inserción en sistemas modulares*: dado que existen chips y hardware especializados para redes neuronales que mejoren su capacidad en la tarea.

Los Sistemas Difusos se pueden definir como algoritmos precisos, que describen relaciones imprecisas entre objetos. A través de descripciones lingüísticas se definen las relaciones entre los datos de entrada y los datos de salida esperados, eliminando el exceso de precisión inherente al modelado matemático.

Esta característica nos ayuda a tratar conceptos de vaguedad, imprecisión e incerteza. La segunda característica que mejora es el aprendizaje de bajo costo para adaptarse a preferencias y entornos individuales.

Algunas características que podemos mencionar sobre los sistemas difusos son:

- Representan formalmente la ambigüedad y subjetividad del pensamiento humano.
- Modelan problemas complejos reales en aplicaciones relevantes:
 - Reconocimiento de patrones.
 - Predicción.
 - Control.
 - Asistencia en la toma de decisiones.

1.4 Breve descripción de los otros capítulos de la memoria

El trabajo final está organizado de la siguiente manera:

En la primera parte, (capítulos 1 al 3) se incorporan los conceptos necesarios para el desarrollo del trabajo final de maestría. Se desarrolla el marco teórico evaluando diferentes métodos que pueden resolver el problema conociéndolos en profundidad y destacando sus aspectos diferenciadores y justificación conceptual. Así también se presenta una breve mención de los fundamentos matemáticos necesarios para el desarrollo de la teoría y la descripción detallada de los aspectos relevantes de esta herramienta.

El abordaje y selección de los métodos permite el desarrollo de:

La resolución a través de algoritmos híbridos neuro difusos (*Fuzzy*) mejorados a través de funciones y desarrollados en MATLAB® y GNU Octave.

A partir del capítulo 4, se describe el desarrollo de la herramienta, es decir, la solución propuesta.

Como líneas futuras a seguir, se deja planteada la solución final para producirla a través de una aplicación en Smartphone, mejorada con filtros específicos que permiten tratar la fluorescencia y también hacer extensivo al uso en células humanas.

El trabajo final de maestría tiene la siguiente organización:

Estado del Arte: Se analizan los conocimientos actuales con respecto a la temática que se abordará tanto en el problema como en la solución. Se incorporan conceptos con los que se desarrolla el trabajo final y cuales representan ventajas y desventajas con respecto a su utilización en la solución final.

Inteligencia Artificial en el Reconocimiento de Patrones de Imágenes: Teoría que enmarca y sostiene la solución elegida. Desarrollo de modelos y algoritmos elegidos. Redes Neuronales Artificiales, sus estructuras, aprendizajes, entrenamiento y otros aspectos importantes como la convergencia de los procesos y descripción de arquitecturas y sistema neurodifuso.

Procesamiento digital de imágenes citogenéticas para su clasificación

Modelos elegidos: justificación de la elección de métodos y algoritmos para desarrollo de la solución.

Desarrollo de la herramienta: Descripción del desarrollo, modelos, algoritmos matemáticos, imágenes de las diferentes etapas.

Resultados y Discusión: La realización del prototipo nos deja como resultado un conocimiento acerca de las nuevas tecnologías y la capacidad de reproducir con mucha precisión la selección y clasificación de imágenes que sólo podía ser realizado por la experticia de un genetista. Luego de un primer entrenamiento con un conjunto de datos específico (*dataset*) y clasificado a mano, seguido por un entrenamiento con un conjunto de prueba (*test set*) constituido por varias especies. La programación del algoritmo nos permite la automatización en la clasificación, además de tener un registro histórico y la creación de un dataset específico.

Conclusiones: Con la propuesta de este trabajo final de maestría se buscó automatizar la clasificación de células animales (peces y bivalvos) en test de fluorescencia mejorando los tiempos invertidos por el especialista en el Laboratorio de Citogenética y Monitoreo Ambiental y se logró una optimización muy efectiva en la clasificación de diferentes tipos de células. A través del estudio de diferentes metodologías se evaluó su aplicabilidad, con pruebas primero de característica empírica y luego validándolos con el caso de estudio. Con la aplicación de los algoritmos neuro difusos y se logró una optimización muy efectiva en la clasificación de diferentes tipos de células. Ese era uno de los problemas que generaba una limitación muy importante dado que muchos modelos solo permitían clasificar ADN de una sola especie. Hay que tener en cuenta que en este estudio se preparó al modelo para clasificar el ADN de diferentes especies de peces y bivalvos con aciertos del 93%.

Líneas futuras: Descripción de las líneas para futuros análisis basados en mejoras para el estudio de caso y la posibilidad de expandirlo a otros estudios en células humanas y a manejo de imágenes en general.

2

Estado del Arte

Estado del Arte

2.1 Estado del Arte del ensayo cometa

Recién a partir del año 1960 la contaminación cobra relevancia a nivel internacional debido a los problemas causados al hombre y su ambiente. El interés por estas cuestiones se intensificó aún más en las últimas décadas debido a la ocurrencia de accidentes con productos químicos que adquirieron repercusión mundial. Surgen así las primeras acciones por conocer causas y efectos no deseados de distintos agentes tóxicos[15].

El surgimiento de disciplinas especializadas como Ecotoxicología y genética toxicológica que se desarrollaron en torno a objetivos científicos, tecnológicos como prácticos; y del interés político de las Naciones Unidas manifestado desde 1992 en la Conferencia Mundial sobre Desarrollo y Medio Ambiente, permitieron adoptar un plan de Desarrollo Sustentable para el siglo XXI, que incluyó aspectos sociales, económicos y ecológicos de todos los dominios públicos tanto a nivel local y regional, como nacional e internacional[16].

Estos diferentes escenarios plantearon el surgimiento de los estudios de Evaluación de Riesgo Ambiental (ERA), que se volvieron preponderantes a medida que los expertos y el público en general comprendieron que ciertos químicos no tóxicos para la salud humana, pueden tener efectos adversos sobre los recursos naturales. Estos estudios implican la identificación y caracterización del riesgo, como también la probabilidad de ocurrencia de efectos adversos sobre todos los componentes del ecosistema[16].

El monitoreo ambiental es uno de los principales pasos en la evaluación de riesgo ambiental, que consiste en la observación repetitiva de uno o más elementos químicos o biológicos, en base a objetivos definidos dentro de un esquema organizado en tiempo y espacio, utilizando métodos comparables y estandarizados. Para evaluar el riesgo que suponen los contaminantes ambientales sobre los organismos y clasificar la calidad de los distintos ambientes, el monitoreo puede centrarse en componentes fisicoquímicos

(Monitoreo Físicoquímico) o en componentes biológicos en todos los niveles (Monitoreo Biológico o Biomonitoreo)[16].

El Biomonitoreo se realiza a través de biomarcadores, que detectan y reflejan los cambios en las respuestas biológicas (variables) inducidas por la presencia de contaminantes ambientales, que se manifiestan a través de variaciones de componentes celulares y bioquímicos, como así también de variaciones en estructuras o funciones determinables en un organismo (bioindicador) o una muestra del mismo. Constituyen señales de alerta temprana que reflejan respuestas biológicas adversas ante la acción de un contaminante y se caracterizan por su sensibilidad, relevancia biológica y factibilidad[16].

Entre las técnicas disponibles para evaluar daños en el ADN, el Ensayo Cometa ó Electroforesis en Gel de Células Aisladas (*Single Cell Gel Electrophoresis*), se ha convertido en uno de los métodos de elección. Este ensayo fue descrito inicialmente como una técnica de electroforesis en microgel para detectar daños en el ADN a nivel de células aisladas [17][18].

La técnica de Ensayo Cometa que utiliza como medio y herramienta la microscopía de fluorescencia tiene un importante papel en las investigaciones y es una técnica muy utilizada. Este ensayo se fundamenta en la estructura y organización del ADN dentro del núcleo, donde las células objeto de estudio son expuestas a una solución química sobre portaobjetos, se remueven las membranas celulares, el citoplasma y el nucleoplasma. “La estructura resultante se denomina nucleoide y es el ADN desprovisto de proteínas histonas, del tamaño del núcleo original de la célula”. Ese ADN, es desnaturalizado en una solución alcalina por un periodo de tiempo suficiente, para exponer la presencia de quiebres o roturas en sus cadenas como resultado de la exposición a contaminantes, lo que provoca una relajación del superenrollamiento en su estructura. Cuanto mayor sea el número de roturas producidas en las hebras del ADN, mayor será el grado de relajación de las mismas. Al someter estos nucleoides a electroforesis, el ADN negativamente cargado migrará hacia el ánodo. Luego de neutralizar y fijar los preparados, la tinción con colorantes fluorescentes de unión específica al ADN, revelará imágenes similares a “cometas”, característica que le otorga el nombre a la

Procesamiento digital de imágenes citogenéticas para su clasificación

técnica. Por lo tanto, la habilidad del *ADN* para migrar y originar los cometas, dependerá tanto del tamaño como del número de roturas ocurridas en sus hebras. Cuanto mayor sea el daño, más larga e intensamente coloreada se verá la cola de los cometas en relación con la cabeza de los mismos [16][18].

Luego del ensayo se registran las imágenes resultantes del análisis de microscopía. En el laboratorio, estas tareas se realizan hasta el momento manualmente con cámaras fotográficas que realizan las capturas en condiciones de total oscuridad y el ojo entrenado del genetista clasifica las imágenes.

Uno de los principales objetivos de investigación del laboratorio es evaluar el impacto de los ambientes de agua dulce con distintos niveles de contaminación sobre el *ADN* de diferentes organismos bioindicadores [19].

Con esta técnica citogenético-molecular es posible además el análisis de diferentes tipos celulares en un mismo individuo, brindando una información más acabada del impacto ambiental ya que permite evaluar el daño en el *ADN* de varios tejidos y órganos, que reaccionan de manera diferente a los diversos contaminantes ambientales[20].

La problemática de la contaminación de los ríos ha llevado a especialistas a explorar herramientas innovadoras para la evaluación de efluentes en aguas superficiales. Una de las alternativas que toma relevancia es el empleo de macroinvertebrados (moluscos bivalvos) y vertebrados inferiores (peces, anfibios) como bioindicadores. Esto es debido a su distribución y sedentarismo, sensibilidad a perturbaciones ambientales, largos ciclos de vida en el agua (en el caso de los macroinvertebrados), y por sus características ecológicas y su posición en la cadena trófica. Por todo esto, resultan buenos bioacumuladores de sustancias del entorno (principalmente compuestos orgánicos) [16].

La bioconcentración y bioacumulación de sustancias ocurren cuando un organismo ingiere contaminantes químicos de su entorno y a través de la alimentación, respectivamente. Estos contaminantes se acumulan en los tejidos y a través de la cadena trófica, por un proceso de biomagnificación, se incrementa su concentración en los organismos de mayor jerarquía en la

cadena alimentaria, pudiendo repercutir en última instancia en los seres humanos que se alimentan de ellos [16].

Este estudio ambiental a través del biomonitoreo es una alternativa muy usada al ser simple, de bajo costo y con resultados de alta confiabilidad.

2.2 Estado del arte de imágenes en el monitoreo ambiental

La incorporación de redes neuronales de una manera jerárquica y particionada al tratamiento de imágenes del estudio ambiental, se viene realizando desde 2008 [21][22].

En 2009 se desarrolla un algoritmo de evolución diferencial para reconocimiento de macroinvertebrados. Este consiste en un esquema autoadaptativo con dos algoritmos de búsqueda locales, que detectan y otorgan un valor a una descendencia. Dichos procesos ayudan en la búsqueda global que otorgan soluciones mejoradas como antecedente a lo que podemos utilizar hoy [23]. Dan más flexibilidad al problema al tener dos algoritmos de búsqueda y no solo uno.

Estos algoritmos usan la técnica de máquinas de vectores soporte que es una representación de los datos como puntos dispersos en el espacio de forma que facilita una clasificación binaria, luego en 2009 se propone un sistema de reconocimiento de patrones en dos y tres dimensiones[24].

La implementación de redes neuronales convolucionales (*CNN*) se manifiesta fundamentalmente desde 1998 con un artículo seminal [25] que presenta las ideas más relevantes acerca de reconocimiento de imágenes utilizando redes neuronales multicapas entrenadas convolucionales y retropropagación aplicado al reconocimiento de patrones de escritura, que son ejemplo de una técnica exitosa de aprendizaje basada en gradiente. Dada una arquitectura de red adecuada, los algoritmos de aprendizaje basados en gradientes se pueden usar para sintetizar una superficie de decisión compleja que puede clasificar patrones de alta dimensión, como los caracteres escritos a mano, con un procesamiento previo mínimo. Este artículo revisa varios métodos aplicados al

Procesamiento digital de imágenes citogenéticas para su clasificación

reconocimiento de caracteres escritos a mano y los compara en un dígito escrito a mano estándar, cuyo éxito se basa en la evolución de la capacidad computacional[25].

En el año 2012, el desarrollo de *Alex net* una red neuronal, logró un 15% de error en la identificación de imágenes para el concurso de reconocimiento visual a gran escala Imagenet (*ILSVRC*). Con este procedimiento se logró un notable avance sobre su rival que marco un 25%, y determinó claramente su superioridad[26]. El desarrollo de técnicas de aprendizaje profundo de precisión en 2015 logró un 4,8% de performance, excediendo la precisión humana en la misma prueba[27].

El análisis manual, puede resultar tedioso y estar expuesto al sesgo del profesional que lo analiza, lo que hace considerar el uso de otros métodos que permitan obtener indicadores posteriores al análisis visual de las muestras, como el de inteligencia artificial o minería de datos [26][27][28].

El desarrollo de este prototipo de automatización para clasificar las imágenes citogenéticas según su daño en el ADN, tiene el fin de mejorar los tiempos de análisis y la posibilidad de crear un histórico de los resultados, teniendo en cuenta además la estandarización de las fotos de las muestras para que al compararlas todas estén clasificadas con un histograma ecualizado. Asimismo evaluar de qué forma se puede optimizar este procedimiento usando algunos algoritmos Genéticos [29], Neuro difusos[30], o Redes Neuro Convolucionales (CNN) [24].

En este momento los métodos que se dispone para reconocer las imágenes que presenta el microscopio, para que esas imágenes puedan ser captadas, procesadas y analizadas a través de mecanismos digitales, son diversos[31].

En la bibliografía analizada, se encontraron desarrollos acerca del procesamiento de imágenes digitales en diferentes ámbitos, como ingeniería, biología, medicina, etc. Sin embargo, no hay muchas referencias sobre genética o citogenética con algoritmos de aprendizaje.

Procesamiento digital de imágenes citogenéticas para su clasificación

El procesamiento de imágenes digitales es el que se realiza a través de una computadora y son las imágenes obtenidas dentro del rango visible del espectro electromagnético, imágenes acústicas, electrónicas y sintéticas [32].

El avance en esta área no se da por sí sola sino que siempre va de la mano de otras disciplinas con las que está asociada como las matemáticas y la computación y se ve reflejado en biología, mecánica, medicina, geología, minería, astronomía etc. [33][34].

Los tratamientos de imágenes son multivariados. Por ejemplo, en histopatología se incorpora nueva tecnología al análisis de imágenes, no solo a través de computadoras sino también a través de Smartphone [35] que ofrece como ventaja la recopilación y almacenamiento de los registros fotográficos, y su clasificación posterior inmediata a través del uso de la aplicación desarrollada para tal fin. Hay software desarrollado como "Komet 7®" que captura y hace análisis de cometas, calcula las medidas aceptadas internacionalmente (posee certificación) y genera estadísticas, o Capslab que es un proyecto de software de ensayo cometa, que cumple con las funciones propuestas en esta investigación, algunas son de distribución libre y otras son desarrollos propietarios.

En las bases de todos los desarrollos y procedimientos son usados algoritmos matemáticos para dar sustento a las redes neuronales, sistemas neuro difusos e Inteligencia Artificial [33][36].

Se elige el razonamiento difuso impulsado por NN Hideyuki Takagi e Isao Hayashi Matsushita Electric Industrial Co., [4] porque puede resolver dos problemas del razonamiento difuso convencional combinando una red neuronal artificial y un razonamiento difuso, lo que permite además fijar las reglas de inferencias y el ajuste de acuerdo al entorno de razonamiento variable en el tiempo. Este método propuesto se puede aplicar al modelado de la red e inteligencia artificial, y se considera de incorporación de conocimiento en la estructura de red neuronal [4].

El razonamiento difuso imita por su flexibilidad al razonamiento de la lógica humana. Sus aplicaciones se ven en campos de control, en Inteligencia

Artificial e investigación operativa. No son un conjunto de reglas que logran un control fluido con suaves variaciones, sino que aporta además cierta flexibilidad como en la lógica humana en sí misma.

El control difuso es una regla de control del razonamiento humano expresada en palabras humanas, Logra representar en una regla difusa o borrosa conceptos como “la velocidad debe reducirse gradualmente”, que además disminuye la cantidad de reglas y para ejecutarla se debe determinar la relación entre los datos de entrada obtenidos a través de sensores y los diferentes conjuntos difusos. Esta relación de mapeo es función de membresía. La principal fortaleza del método es el logro de reunir las reglas lógicas y contextualizarlas en conjuntos difusos a través de funciones de membresía (la función se encarga de clasificar los elementos de confusión entrelazados en la lógica, o dicho de otra manera. clasifica a través de reglas claras representando formalmente la ambigüedad del pensamiento humano a través del modelado de problemas complejos reales) [37].

La segunda característica que mejora es el aprendizaje de bajo costo para adaptarse a preferencias y entornos individuales.

Las redes neuronales pueden manejar dos problemas del razonamiento difuso. La diferencia entre estas dos tecnologías tiene que ver con si la lógica es explícita o implícita. Para el razonamiento difuso, siempre se puede lograr un razonamiento estable a pesar de las desviaciones de datos, porque la lógica de la red troncal se manifiesta como una regla de la forma SI-ENTONCES (IF-THEN) y no se puede expresar la regla si no se identifica la lógica. Como la red neuronal autoorganiza la relación del mapeo durante el aprendizaje, se puede aplicar a relaciones no conocidas, y no lineales porque es en sí misma no lineal. A la vez requiere gran cantidad de datos. Si en el aprendizaje, los datos son erróneos, pueden desviar el aprendizaje [4].

Entonces el razonamiento difuso se utiliza para casos lógicos bien identificados como el control y las redes neuronales para reglas de reconocimiento no identificadas, como el reconocimiento de patrones [6].

Procesamiento digital de imágenes citogenéticas para su clasificación

Los métodos elegidos en este trabajo final de maestría son redes neuronales y sistemas neuro difusos, debido a que al momento de la clasificación de las imágenes de las células obtenidas a través de Ensayo Cometa lo difuso de la cabeza y de la cola no siempre se ajustará exactamente a una u otra clase por más detalles que se tenga al definirlo y al examinarlo. El hecho de ser un material biológico hace que tenga sus propias características, incluso que varíen de individuo en individuo, son características que invitan a utilizar un sistema neurodifuso. El ajuste de valores con algoritmos genéticos, o a través de funciones permite mejorar la performance global y llegar a resultados más exactos[4].

Con esta elección, se logra que el algoritmo pueda detectar núcleos y clasificarlos más allá de la especie que se estudie (en este caso dos especies de peces y bivalvos), lo que permite manejar cierta flexibilidad muy conveniente en este tipo de análisis citogenético, de modo que los resultados obtenidos puedan aplicarse a otros casos.

3

Inteligencia Artificial en el reconocimiento de Patrones de Imágenes.

3.1 Problemas para reconocer una imagen

Los problemas a los que hay que enfrentarse cuando hay que diseñar una aplicación que utiliza alguna tecnología de Visión por computadora, algoritmos o Inteligencia Artificial, son complejos. Dado que en el desarrollo de algoritmos, las imágenes constituyen una matriz de números, las aplicaciones tendrán que traducir esos números a objetos.

También hay que tener en cuenta que los objetos de estudio van a presentar un conjunto de apariencias muy distintas en la imagen ya que dependerá, por ejemplo, del punto de vista que lo hará aparecer rotados, trasladados a una zona de la imagen o a otra, o más o menos grandes. Y a pesar de esos cambios el algoritmo tiene que saber que se trata del mismo tipo de objeto.

Además, se puede tener el caso de que la imagen que se busca no tenga una forma con contornos simétricos, o continuos o claramente definidos, por ejemplo una célula, por lo que presenta cambios debidos a la posición o simplemente por no ser exactamente igual una a otra. Otro factor muy importante es la iluminación, que puede no ser constante, y que no sólo va a producir que las imágenes se vean más claras u oscuras, sino que también podrán generar sombras, modificar el contraste entre diversas zonas, etcétera.

En una aplicación puede ser que el objeto de estudio se encuentre junto con otros objetos en la imagen analizada. La presencia de otros objetos puede producir oclusiones, con lo cual la imagen de interés no se verá globalmente sino sólo alguna de sus partes.

La búsqueda también se complica si se tienen diferencias en los posibles campos de observación de acuerdo a la especie que se estudia, con lo que es fundamental definir el criterio de búsqueda de forma concreta y qué características se quiere buscar en la imagen. A todo ello hay que añadir que, normalmente, existe una restricción temporal muy exigente.

Procesamiento digital de imágenes citogenéticas para su clasificación

Y en ese tiempo se examinan los posibles resultados en cada imagen, con lo cual, si un solo algoritmo funciona al 99,9% de acierto, se tendrán varios resultados descartados o erróneos por imagen.

Estos errores serán los falsos positivos, es decir, cuando se afirme que los objetos buscados están presentes en la imagen y no sea así. Y en caso contrario, los falsos negativos, que es cuando no se los detectan a pesar de que estén.

En la actualidad hay muchos métodos que dan soporte a estas acciones, tanto en Aprendizaje Automático (*Machine learning*) como en Aprendizaje profundo (*Deep learning*)[38][39].

La neurona humana dio origen a la creación de los procesos de inteligencia artificial a través del descubrimiento de los mecanismos que gobiernan la morfología y los procesos conectivos de las células nerviosas. [40]

Las redes neuronales proponen una forma de emular ciertas características propias de los humanos, como la capacidad de memorizar y de asociar hechos. Si se examinan con atención aquellos problemas que no pueden expresarse a través de un algoritmo, se observará que todos ellos tienen una característica en común: la experiencia. El hombre es capaz de resolver estas situaciones acudiendo a la experiencia acumulada. Así parece claro que una forma de aproximarse al problema consista en la construcción de sistemas que sean capaces de reproducir esta característica humana [41].

En definitiva, las redes neuronales no son más que un modelo artificial y simplificado del cerebro humano, que es el ejemplo más perfecto del que disponemos para un sistema que es capaz de adquirir conocimiento a través de la experiencia. Una red neuronal es “un nuevo sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: la neurona” [41].

El proceso que se da es que la neurona es estimulada o excitada a través de sus entradas y cuando se alcanza un umbral, la neurona se activa pasando una señal al axón.

Esto es debido a que se desarrolla además un proceso electroquímico [42].

La inteligencia se da en ese contexto de interrelación entre millones de neuronas y el aprendizaje como la experiencia de nuevos caminos o nuevas conexiones creadas a través de nuevos estímulos. Entonces las redes neuronales consisten en unidades de procesamiento que intercambian datos o información. A través de ellas se pueden reconocer patrones, incluyendo imágenes, señales, manuscritos y líneas de tiempo (tendencias y proyecciones) con la capacidad de aprender y mejorar los procesos [25].

Las definiciones varían de acuerdo al contexto en que se las definan, pero siempre surgen de la inspiración en modelos biológicos y permiten obtener resultados a través de la aplicación de procesos matemáticos, o estadísticos, en el que los datos se organizan por niveles y las salidas son resultados de las interacciones de elementos y procesos simples interconectados que además intentan relacionarse con el mundo real [43].

Así como el cerebro es capaz de aprender de la experiencia, en la Inteligencia Artificial hay características similares que permiten abstraer puntos de coincidencia esenciales, a partir de entradas que representan información irrelevante. Las redes neuronales pueden reconocer datos o imágenes que nunca antes habían visto, a través de generalizar la información de entrada y autorganizarse, reconociendo ciertas características armando patrones.[44]

Como presentan características similares al cerebro, (son capaces de aprender, de generalizar casos nuevos basados en casos anteriores, pueden abstraer características esenciales a partir de datos de entrada) también tienen características muy diferentes; Los circuitos del cerebro no implementan computación recursiva y por lo tanto no son algorítmicos. Aunque un algoritmo, por definición, puede definir funciones recursivas.

La unidad computacional básica de una red neuronal, la neurona, recibe una señal eléctrica de entrada y al llegar a cierto umbral se transforma generando una nueva señal eléctrica de salida como se observa (Fig.1). Esa transformación se realiza en el modelo a través de funciones de activación, las más comunes son funciones de paso, sigmoideas, RELU, gaussianas.

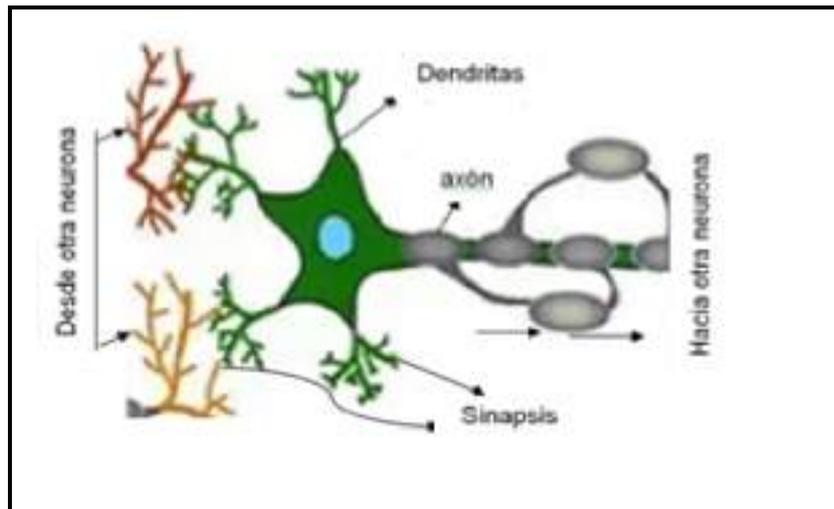


Figura 1: Representación gráfica de los procesos conectivos de la neurona biológica mostrando las sinapsis

La neurona, recibe información de algunos nodos, o de una fuente externa y genera una salida. Cada entrada tiene un peso asociado (w), que se asigna en función de su importancia relativa a otras entradas (Fig. 2). El nodo aplica una función de activación, por ejemplo, sigmoidea, Tangente Hiperbólica o RELU a la suma ponderada de sus entradas. Si la señal combinada no es lo suficientemente grande, el efecto de la función de umbral sigmoidea es suprimir la señal de salida y disparar de otro modo.

En una red neuronal biológica, las dendritas recogen las señales eléctricas y se combinan para formar una señal más fuerte. Si la señal es fuerte y supera el umbral, la neurona dispara una señal por el axón hacia las terminales para pasar a las dendritas de la próxima neurona. Es importante tener en cuenta que cada neurona recibe información de muchas anteriores y también proporciona señales a muchas más. Una forma de replicar este fenómeno de la naturaleza mediante un modelo artificial es tener capas de neuronas, cada una conectada entre sí en la capa anterior y posterior. El siguiente diagrama (Fig. 2) ilustra esta idea:

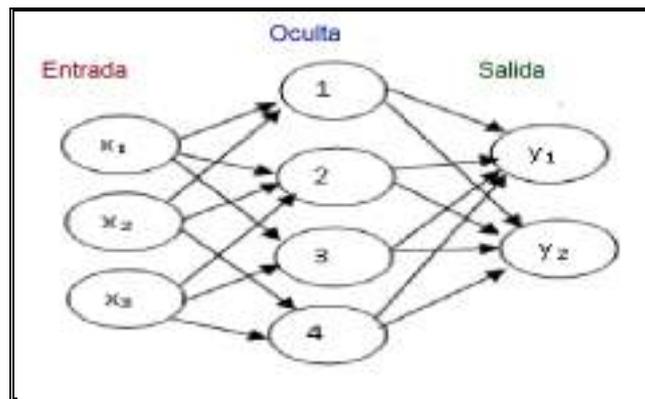


Figura 2: Diferentes capas que alimentan a una neurona y como se acumulan las señales.

En una arquitectura de Deep Learning una neurona simple produce transformaciones en los datos a través de la siguiente forma:

$X \rightarrow$  $\rightarrow y = f(\omega \cdot x + b)$ donde los datos entre paréntesis hacen la primera transformación y luego a través de las funciones de activación que no son lineales se provocan las salidas.

En la función lineal de la forma $Z = WX + b$ siendo W el peso asociado a las entradas y Z la salida o acción (esto quiere decir que las neuronas no reaccionan solo al entrar en contacto con el estímulo), se manifiesta que las neuronas no reaccionan con facilidad sino que suprimen la entrada hasta que la intensidad del estímulo ha crecido tanto que se dispara una salida.

Aquí aparece la función de activación. Es una función que toma la señal de entrada y genera la señal de salida pero teniendo en cuenta algún tipo de umbral.

Algunas funciones de activación son las siguientes:

Tabla 1: Funciones de activación:

Función	Formula	Rango
Identidad	$y = x$	$[-\infty, \infty]$
Escalón	$y = \begin{cases} +1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$	$[0, 1]$
	$y = \begin{cases} +1 & \text{si } x \geq 0 \\ -1 & \text{si } x < 0 \end{cases}$	$[-1, 1]$
Lineal a tramos	$y = \begin{cases} x & \text{si } -1 \leq x \leq 1 \\ +1 & \text{si } x > 1 \\ -1 & \text{si } x < -1 \end{cases}$	$[-1, 1]$
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$	$[0, 1]$
	$y = \tanh(x)$	$[-1, 1]$
Sinusoidal	$y = \text{Sen}(\omega \cdot x + \varphi)$	$[-1, 1]$

La función de **Escalón o Paso**, donde la salida es cero para valores de entrada bajos. Pero una vez que alcanza el umbral, la salida aumenta. Se puede mejorar la función de paso de muchas maneras, el umbral esta dado por la suma de los pesos de las entradas.

La **función en forma de S** que se muestra en la (Fig.3) se llama **sigmoidea** o función **logística**. Su forma es:

$$f(z) = \frac{1}{1+e^{-z}} \quad (1)$$

Esta función clasifica básicamente en dos clases, 0 y 1.

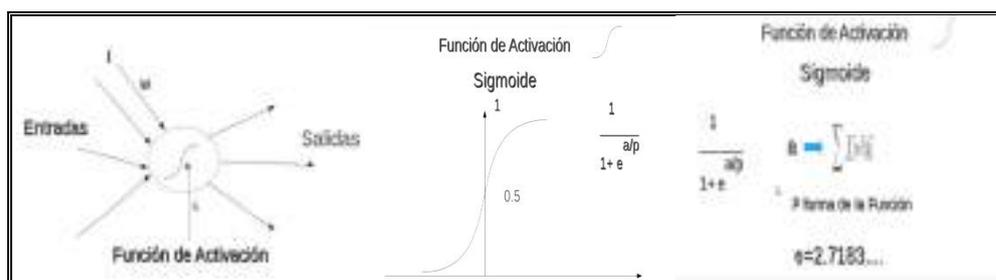


Figura 3: Función de activación sigmoidea o logística de una neurona donde a es la suma de las entradas por sus pesos y p la forma de la función.

Otra función de activación muy importante que se usa ampliamente es **ReLU** fig. 4 o función de activación de *unidad lineal rectificadora*, cuya ecuación es:

$$R(z) = \text{Max}(0, z) \quad (2)$$

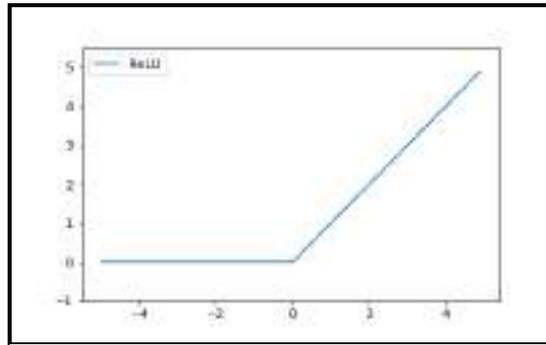


Figura 4 Función de unidad lineal rectificadora.

Hay diferentes tipos de funciones de activación. En el prototipo elaborado se ha utilizado la función sigmoidea para realizar la clasificación de las imágenes.

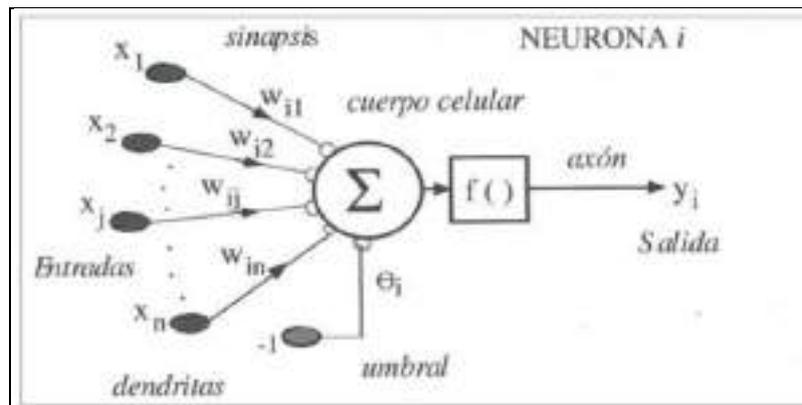


Figura 5 Neurona y las funciones activadoras de acciones.

En una primera instancia una de las líneas de investigación elegida fue la topología de redes neuronales polinómicas auto organizadas (SOPNN) basada en el perceptrón multicapa (MLP) genéticamente optimizado, una opción interesante a utilizar. En el diseño de las redes neuronales polinómicas auto organizadas (SOPNN) "convencional" se utiliza la técnica extendida del método grupal de manejo de datos (GMDH) para considerar los nodos fijos que componen la entrada de cada capa. Sin embargo, este aprendizaje no siempre da como resultado la arquitectura óptima, por lo que el procedimiento de diseño

aplicado en la construcción de cada capa teniendo en cuenta las características específicas y la optimización de parámetros, propone además un índice de rendimiento con un factor de ponderación que permita un equilibrio entre las aproximaciones y las generalizaciones o predicciones. Es un modelo con precisión y capacidad predictiva. Esta topología de red neuronal ontogénica supervisada llamada Redes Neuronales de Auto Optimización (SONN) (Fig.6) tiene redes neuronales multicapas características, con conexiones que se logran a través de un proceso de adaptación basados en los datos de entrenamiento proporcionados. El proceso es capaz de extraer las características más generales y discriminativas de los datos y crear solo las conexiones principales (neuronas), reduciéndose automáticamente los vectores de datos de entrada de entrenamiento, y evitando así problemas de dimensionalidad [33].

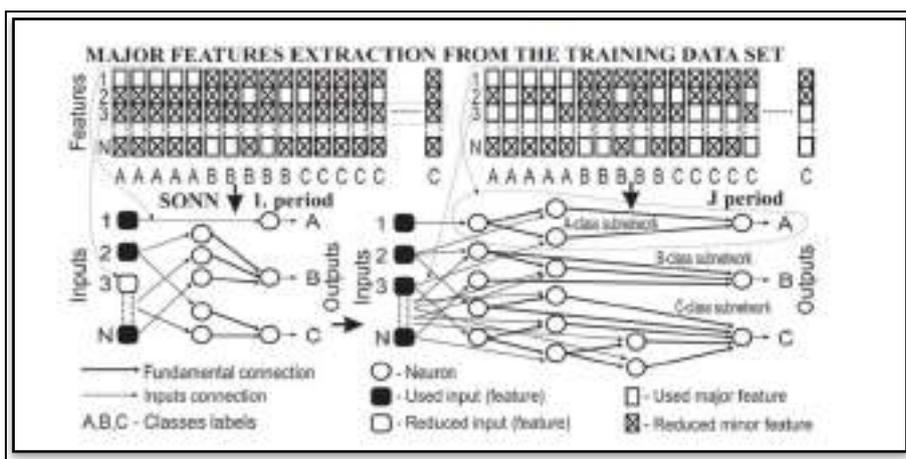


Figura 6: Desarrollo SONN basado en discriminación de características. Fuente: [37] "Neural network designed on approximate reasoning architecture and their Applications Hideyuki Takagi e Isao Hayashi Matsushita Electric Industrial Co. SONN adapta automática y simultáneamente la topología del NN y todos sus pesos en el proceso determinista .

El razonamiento difuso(Fig. 6) puede resolver los problemas del razonamiento difuso convencional a través de una red neuronal artificial y un razonamiento difuso, para lograr el diseño de la función de membresía y La flexibilidad de adaptación al entorno de razonamiento[4].

El enfoque de razonamiento difuso los resuelve mediante una función de aprendizaje y una red neuronal, siendo capaz de determinar reglas de

inferencias y el ajuste de acuerdo al entorno de razonamiento variable en el tiempo debido al uso de red neuronal en el razonamiento difuso. Este método propuesto se puede aplicar al modelado de la red neuronal y la inteligencia artificial, incorporando conocimiento a la estructura de red neuronal[4].

El razonamiento difuso imita al razonamiento de la lógica humana, esta flexibilidad -de fundamental importancia al aplicar sobre material vivo- fue definitoria a la hora de elegir el método para el algoritmo solución.

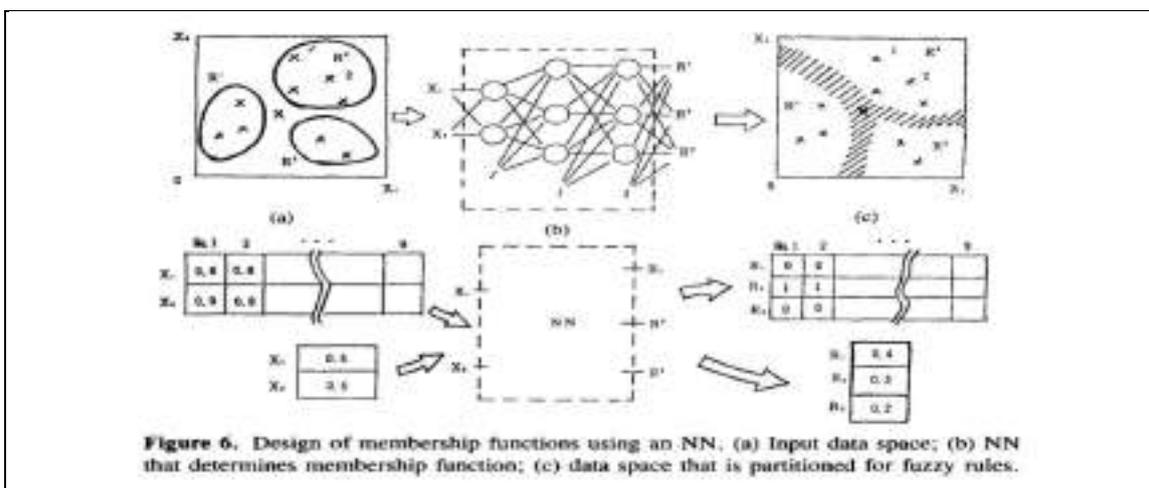


Figura 7: razonamiento Difuso [4] H Takagi y I. Hayashi "Nn Driven fuzzy reasonnig"

De esta manera, tal como se ve en la Fig.7, dentro de las redes neuronales y los conjuntos difusos se encuentra la solución al problema planteado en el Laboratorio de Citogenética (UNaM-IBS- CONICET). Se selecciona este método híbrido, neuro difuso por su flexibilidad, simpleza de ejecución y sobre todo por lo sencillo del prototipo luego de aplicar algunas funciones y algoritmos.

3.2 Redes Neuronales

Esta es una red neuronal con tres capas, cada una con varias neuronas o nodos artificiales, donde cada nodo está conectado a cualquier otro nodo en las capas anteriores y siguientes. De esta manera tomamos la idea del cerebro humano y la aplicamos para construir una arquitectura neuronal para

Procesamiento digital de imágenes citogenéticas para su clasificación

computadoras. El aprendizaje se realiza al ajustar la fuerza de las conexiones entre nodos a través de algoritmos matemáticos. Dentro de un nodo, se ajusta la suma de las entradas, o bien la forma de la función umbral sigmoidea, pero eso es más complicado que simplemente ajustar la fuerza de las conexiones entre los nodos. La Fig. 7 muestra los nodos conectados, pero esta vez se muestra el peso asociado con cada conexión. Un peso bajo desestimará una señal y un peso alto la amplificará.

La idea de calcular señales en una red neuronal desde las entradas a través de las diferentes capas para convertirse en la salida, se llama alimentación de red neuronal. Consiste en una función booleana representada por la fórmula :

$$F(x,y,z)=xy+z \quad (3)$$

Los valores de esta función que se dan a continuación, se utilizan para efectuar las operaciones de la red neuronal a través de la ecuación 3:

Tabla 2 Valores de función para efectuar las operaciones de red neuronal.

X	Y	Z	xy	zZ	$f(x,y,z)=xy+z$
1	1	1	1	0	1
1	1	0	1	1	1
1	0	1	0	0	0
1	0	0	0	1	1
0	1	1	0	0	0
0	1	0	0	1	1
0	0	1	0	0	0
0	0	0	0	1	1

Se utiliza la cuarta fila representada en la tabla 2 ((1, 0, 1) ≥ 0) para demostrar la prealimentación, donde (1, 0, 1) ≥ 0 muestra la alimentación directa.

Procesamiento digital de imágenes citogenéticas para su clasificación

La primera capa se llama capa de entrada y la última capa se llama capa de salida. Las capas intermedias se llaman capas ocultas. Las capas de entrada y ocultas contienen tres nodos y la capa de salida contiene un solo nodo. Si se asignan pesos a las sinapsis entre la entrada y la capa oculta, los pesos se toman aleatoriamente entre (0) y (1) ya que es la primera vez que se alimenta la red Fig. 8.

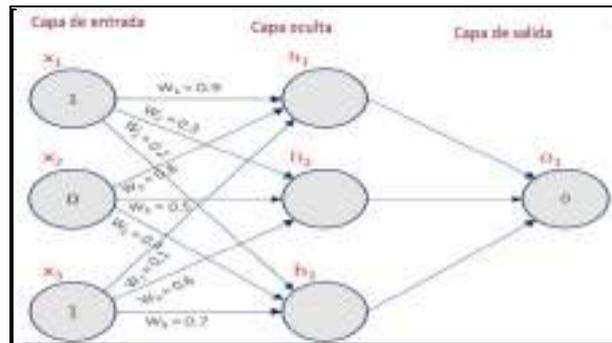


Figura 8: Red Neuronal, de prealimentación.

Como trabaja una sola neurona o nodo, se toman todas las entradas y las multiplicamos con los pesos asociados, finalmente se suman. Luego, el nodo aplica una función de activación, como la sigmoidea, a la suma ponderada de sus entradas para introducir la no linealidad.

Este proceso se repite para varias capas y para cada nodo en estas capas.

Todos los nodos en la capa de entrada están conectados a la capa oculta y esos nodos de entrada tienen valores brutos de los pesos asociados como se ve en la ecuación 4

$$Z = \sum w_{y0} x_{y0} \quad (4)$$

$$d(Z) = \frac{1}{1 + e^{-z}}$$

Procesamiento digital de imágenes citogenéticas para su clasificación

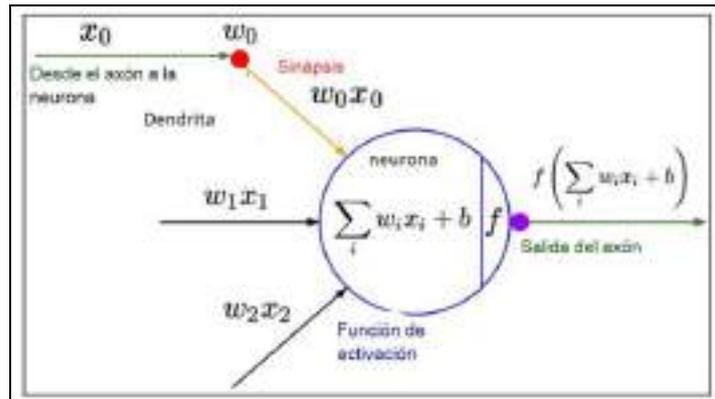


Figura 9: Modificación de una sola neurona desde los valores de entrada, de los pesos y los valores de salida luego de aplicar la función.

Las sumas pequeñas están en el círculo porque no son el valor final. Ese valor de nodo de salida se calcula usando la función de activación. Aplicada a las tres sumas ponderadas de capa oculta se obtiene el valor de capa oculta de red neuronal Fig.9 y luego aplicada a los resultados de la capa oculta con el segundo conjunto de pesos (determinado al azar) se obtiene la suma de salida,:

$$f(z) = \frac{1}{1+e^{-z}} \quad (5)$$

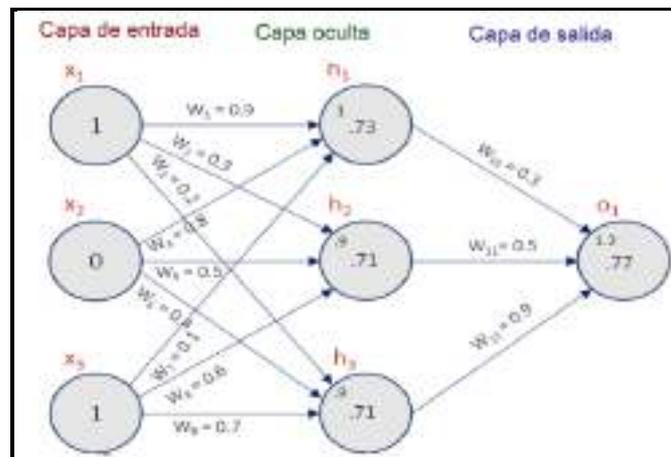


Figura 10: Red Neuronal: cálculos previos al entrenamiento.

Cuando se aplica la función de activación sigmoidea para obtener el resultado final de salida y al usar un conjunto aleatorio de pesos iniciales, el valor de la

Procesamiento digital de imágenes citogenéticas para su clasificación

neurona de salida está fuera de rango; en este caso por ejemplo, +0.77 (ya que el objetivo es 0).

$$\sigma(1.213) = 0.7708293339958$$

Entonces, lejos del valor objetivo hay que entrenar la red para calibrar los pesos o valores obtenidos Fig. 10, para calcular el proceso de alimentación directa o prealimentación (*feedforward*) también llamada red neuronal de avance, se usa el cálculo matricial. El proceso para calibrar los pesos es la retropropagación Fig.11.

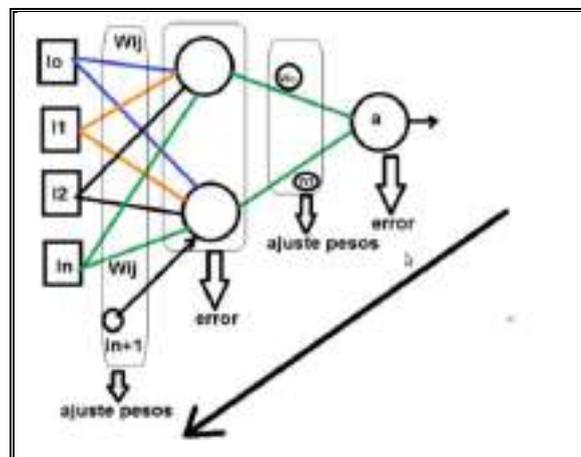


Figura 11: Imagen ilustrativa de Red Neuronal con Tres capas y pesos representados con propagación hacia atrás.

Matrices en la red neuronal

Una matriz tiene la siguiente forma de tabla con números, donde los valores de la matriz son los pesos de la red neuronal

W_{11}	W_{21}	W_{31}
W_{12}	W_{22}	W_{32}
W_{13}	W_{23}	W_{33}

Se puede representar las entradas de la red con otra matriz. Al multiplicar ambas matrices, se obtienen las entradas:

Input 1
Input 2
Input 3

Con la suma ponderada de la entrada y la capa oculta, entonces, se puede calcular la salida de la capa oculta: donde W es la matriz de pesos y x es la matriz de entrada.

Esta técnica se llama **vectorización**. La inteligencia artificial permite como se había enunciado antes (pág. 15), un aprendizaje adaptativo, auto organización, tolerancia a fallos, operación en tiempo real y fácil inserción en sistemas modulares.

En el *Aprendizaje Adaptativo* las redes aprenden a diferenciar patrones mediante ejemplos y entrenamientos siendo sistemas auto adaptativos. Son adaptables porque pueden ajustarse a los nuevos aprendizajes desarrollando nuevas jerarquías. Son dinámicos porque pueden estar cambiando constantemente y esto es fundamental ya que no siempre necesita un algoritmo que considere todas las circunstancias, sino que puede crear su distribución de pesos basado en el aprendizaje. Algunas redes aprenden toda la vida, incluso después de completado su período de entrenamiento, solo se debe proporcionar un entrenamiento con patrones.

De esta manera, hablamos de auto organización cuando las redes usan el aprendizaje adaptativo para auto organizar la información que reciben durante el aprendizaje, a fin de lograr un objetivo específico.

Retropropagación o *backpropagation*, genera su organización característica que le permite reconocer ciertos patrones y a la vez le permite generalizar comportamientos. Esto es útil cuando los datos de entrada están parcialmente definidos o definidos de forma incompleta. Este nombre se da porque el error es propagado desde la capa de salida hacia atrás (o sea va corrigiendo los errores hasta llegar a la capa de entrada, ajustando los valores de los datos de entrada). Esto permite que los pesos sobre las conexiones de las neuronas de las capas ocultas, cambien de estado durante el entrenamiento. El cambio de los pesos influye sobre la entrada global, activación y por consiguiente en la salida de la neurona.

Procesamiento digital de imágenes citogenéticas para su clasificación

Para llevar a cabo el diagnóstico del caso de estudio es necesaria la detección y segmentación del núcleo celular o de la célula en su conjunto. Y más allá de la forma utilizada para procesar la información, es difícil encontrar el método lo suficientemente robusto y preciso.

Los modelos de Redes Neuronales son neuronas conectadas entre sí que trabajan en conjunto, sin que haya una tarea concreta para cada una. Con la experiencia, las neuronas van creando y reforzando ciertas conexiones para aprender, y esas conexiones quedan fijas en el tejido. Si bien el sentido primero es biológico, tiene base en las matemáticas, estadísticas y algoritmos computacionales. Se basa en una idea sencilla: dados algunos parámetros, hay una forma de combinarlos para lograr cierto resultado.

Las redes neuronales permiten buscar la combinación de parámetros que mejor se ajusta a un determinado problema. Son un modelo para encontrar esa combinación de parámetros y aplicarla al mismo tiempo.

Para esto se debe entrenar la red neuronal, una red ya entrenada se puede usar para hacer clasificaciones o predicciones, es decir, para aplicar la combinación [44].

Se utilizan en la literatura varios métodos de detección de núcleos, que se pueden subdividir según el algoritmo principal utilizado. Se encuentran métodos que utilizan la transformada de la distancia operación morfológica, transformada H Máxima/mínima, filtro Laplaciano de la Guadiana [45], detección de la región extrema estable máxima, transformada de Hough [46], votación basada en simetría radial; así como también se pueden encontrar métodos de aprendizaje supervisado [47], Support vector machine [22], Reinforcement learning [48] y redes neuronales profundas [49]. Alguno de estos métodos (como la transformada de Hough), veremos que se utilizan en el prototipo desarrollado en esta investigación (en el capítulo 4 pág. 55).

Por otra parte, en la segmentación se intenta delimitar objetos a través de los límites. Tres estrategias se destacan para llevar a cabo la separación del núcleo de las células:

Primero: separar el fondo del plano.

Segundo: Identificar los indicios de núcleo para luego expandir sus límites.

Tercero: generar regiones candidatas y luego seleccionar las mejores para ser segmentadas. Los algoritmos de segmentación más utilizados actualmente son intensidad limite, transformación divisoria, Kmedias (K means) como método de agrupamiento (clustering) [48], cuya versión difusa es Fuzzy-C means clustering, maximización de la expectativa, métodos basados en grafos y clasificación supervisada (clasificación por píxeles: clasificación por súper píxeles) [49]. Muchas de las aproximaciones que presenta la bibliografía, no se incluye para testeo debido a la poca flexibilidad de estos métodos frente a la variabilidad intrínseca de las imágenes del material biológico en microscopía. El tipo de muestra y el modo de preparación, también la luz, puede influir y hacer que la segmentación de las células/núcleos que están parcialmente superpuestas o pegadas entre sí, resulten un reto científico a resolver.

3.3 Logica Difusa

El razonamiento difuso [4], es el enfoque de razonamiento propuesto, resuelve estos problemas mediante el uso de la función de aprendizaje y la no linealidad de una red neuronal. Primero, se identifican los problemas involucrados en el razonamiento difuso convencional y la red neuronal utilizada en este documento. Luego se formula el algoritmo propuesto y se desarrolla una explicación concreta utilizando datos reales.

El razonamiento difuso es capaz de determinar automáticamente las reglas de inferencia y el ajuste de acuerdo con el entorno de razonamiento variable en el tiempo Fig. 12 y 13. Con lo que se incorpora conocimiento a la estructura de la red neuronal.

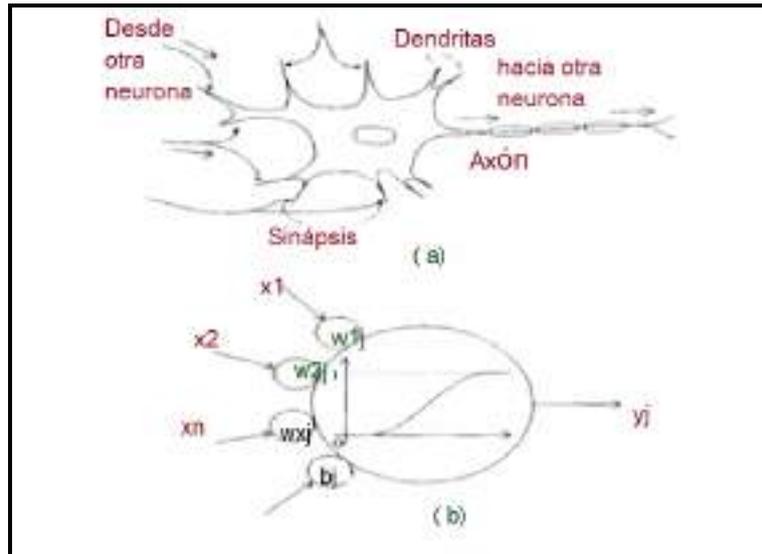


Figura 12: NN driven Fuzzy reasoning, H.Takagi y I.Hayashi [4].

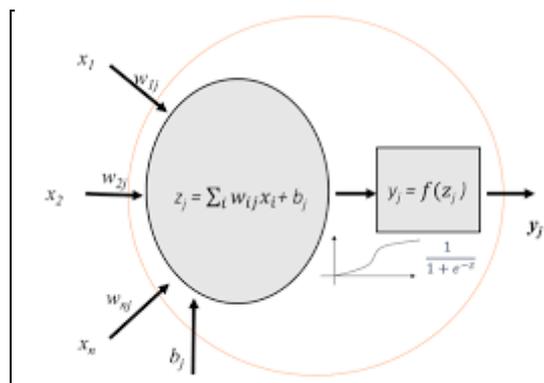


Figura 13: Función sigmoidea en redes neuronales, razonamiento difuso (Fuzzy) [4].

Los algoritmos de lógica difusa y números borrosos permiten perfilar y diferenciar áreas en las que pueden determinar la pertenencia a una clase, sin tener todas las características que se corresponden estrictamente a esta clase. En el método de agrupamiento (clustering), el número de grupos es igual a la cantidad de reglas que se infieren.

3.4 Método Neuro-difuso

El razonamiento difuso que se menciona anteriormente puede resolver dos problemas del razonamiento difuso convencional combinando una red neuronal artificial (NN) y un razonamiento Difuso [4]:

Procesamiento digital de imágenes citogenéticas para su clasificación

- El diseño para la función de membresía, solo enfoque heurístico.
- La falta de flexibilidad de adaptación al entorno de razonamiento.

El enfoque de razonamiento difuso los resuelve mediante una función de aprendizaje de una red neuronal, además es capaz de determinar reglas de inferencia y el ajuste de acuerdo al entorno de razonamiento variable en el tiempo, aplicado al modelado se lo considera como incorporación de conocimiento en la estructura de la red neuronal [4].

La principal fortaleza del método es que logra reunir las reglas lógicas y contextualizarlas en conjuntos difusos a través de funciones de membresía (la función se encarga de canalizar a través de estas reglas más flexibles, los datos con funciones entrelazadas en la lógica).

La segunda característica que mejora es el aprendizaje de bajo costo para adaptarse a preferencias y entornos individuales.

Entonces el razonamiento difuso se utiliza para casos lógicos bien identificados como el control y las redes neuronales para reglas de reconocimiento no identificadas, como el reconocimiento de patrones.

Al diseñar la función de membresía, la función de inferencia se puede derivar partiendo de reglas difusas de los datos de aprendizaje usando el aprendizaje de la red neuronal, lo que también permite una adaptación de la función.

El esquema del método se explica tomando como ejemplo el control realizado por dos entradas x_1 y x_2 , derivadas de dos sensores.

El algoritmo consta de tres partes principales:

1. Reglas de inferencia.
2. La identificación de SI (*If*) (determinación de una función de membresía).
3. La identificación de ENTONCES (*then*) (la determinación del evento de control para cada regla).

La primera parte es la determinación de reglas de inferencia difusas y la combinación de datos que pertenecen a cada regla. Estos datos se agrupan

por un método de agrupamiento y el número de grupos es igual al número de reglas (Ver Fig.7 pagina 36).

La segunda parte es la determinación de la entrada arbitraria para cada regla, que define la función de pertenencia para cada regla y corresponde a la identificación de las partes Si (condición de la regla) que combina las partes difusas. Como se ve en la Fig. 7, en la parte SI se constituye la membresía que utiliza las reglas explícitas y flexibles determinadas en la primera parte para cada clase definida. La red neuronal puede formar una función continua arbitraria, que puede tener incluso una capa oculta.

La tercera parte del algoritmo es la determinación del ENTONCES (partes de la conclusión).

La red neuronal supervisada por los datos de aprendizaje y el valor de control para cada regla como en la segunda parte. La formulación detallada de la red neuronal se explica en el siguiente ejemplo de modelado difuso [4] donde la parte ENTONCES en el control del sistema es responsable de inferir el valor exacto del control. Uno de estos métodos es el modelado difuso mediante la forma: si "Si x es A , ENTONCES $y = u(x)$ ", donde x es un vector de entrada, A es el conjunto difuso de los espacios de reglas particionadas y $u(x)$ designa una función de inferencia para la operación de control.

- Paso1: siendo las variables de entrada $x_j, j = 1,2 \dots k$. y el valor observado, el de salida y_i . donde el $x_j, j=1,2,\dots,m, m \leq k$ relacionado con el valor observado y son seleccionados por la red neuronal. Se realiza mediante el método de backpropagation o eliminación de errores hacia atrás, usando la suma de los errores al cuadrado como función de costo para eliminar las variables de entrada que tienen que ver con el ruido y diferenciar las que realmente son entradas.
- Paso 2: Los datos de entrada y salida (x_j, Y_j) se dividen en datos de entrenamiento (TRN de n_t) y datos de verificación (CHD de n_c) para la estimación del modelo, donde $n = n_t + n_c$.

- Paso 3: La partición de entrenamiento se encuentra mediante un método de agrupación. El mejor número se obtiene mediante un método de agrupación y se decide en relación a la distancia entre los grupos en un dendograma de grupo: (ver Fig. 2). Cada uno de los TRD divididos en grupos r se expresa como $R_s, s=1,2,\dots,r$, y los TRD de R^s se expresan por (x_i^s, y_i^s) donde $i=1,2,\dots,(n_t)^s$, and $(n_t)^s$ son números de TRD en cada R^s . La división del espacio n -dimensional en r aquí significa que el número de reglas de inferencia es r .
- Paso 4: Es la identificación de la constitución de cada parte SI (*IF*) en NN_{mem} (NN que genera las funciones de membrecía). Si x_i son los valores para la capa de entrada, w_i^s se asigna como datos supervisados para la capa de salida.

$$w_i^s = \begin{cases} 1, & \dots, x_i \notin \mathbf{R}^s \\ 0, & \dots, x_i \in \mathbf{R}^s \end{cases}$$

donde $i=1,2, \dots (n_t); s=1,\dots,r$.

El aprendizaje de NN_{mem} se lleva a cabo para que estos w_i^s puedan inferirse de la capa de entrada x , por lo que NN_{mem} se vuelve capaz de inferir el grado de atribución w_i^s de cada ítem de datos de entrenamiento x_i a R^s . Si la función de pertenencia de la parte IF es el valor inferido w_i^s que es el resultado del NN_{mem} aprendido, esto es,

$$\mu_A^s (X_i) \equiv w_i^s, i=1,\dots,n. \quad (6)$$

- Paso 5: Es la identificación de cada ENTONCES. La estructura de cada regla de inferencia se expresa mediante la relación entrada /salida. La entrada los TRD $x_{i1}^s, \dots, x_{im}^s$ y el valor de salida $y_t^s, i=1,2,\dots,(n_t)^s$ se asignan a la entrada y salida de los NN_s . Esta NN_s es la NN de la parte ENTONCES en R^s . El entrenamiento de NN_s se lleva a cabo de tal manera que se puede inferir el valor de control. Los valores de entrada

Procesamiento digital de imágenes citogenéticas para su clasificación

CHD $x_{i1}, \dots, x_{im} \ i=1,2,\dots, n_c$, se sustituyen en la NN para obtener la suma Θ^s_m de los errores al cuadrado.

$$\Theta^s_m = \sum_{i=1}^{n_c} \{y_i - u_s(x_i) \cdot \mu_A^s(x_i)\}^2 \quad (7)$$

Este valor estimado $u_s(x_i)$ se obtiene como la salida NN_s . Otra forma de calcular Θ^s es con el peso; esto significa

$$\theta^s_m = \sum_{i=1}^{n_c} \mu_A^s(x_i) \{y_i - u_s(x_i) \cdot \mu_A^s(x_i)\}^2 \quad (8)$$

Usándolo para decidir el mejor número de iteraciones de aprendizaje de la red neuronal para prevenir sobreentrenamiento.

$$I^s = \frac{n_c}{n_t^s + n_c} \sum_{i=1}^{n_t^s} \{y_i - u_s(x_i)\}^2 + \frac{n_t^s}{n_t^s + n_c} \sum_{j=1}^{n_c} \{y_j - u_s(x_j) \cdot \mu_A^s(x_j)\}^2 \quad (9)$$

Si la red neuronal tiene datos sobreaprendidos el error TRD se vuelve pequeño pero el error CHD se vuelve grande, por lo que el número de iteraciones I^s más pequeño es el mejor.

- Paso 6: Es la simplificación del ENTONCES con el método de eliminación hacia atrás. Entre las m variables de entradas de red neuronal ENTONCES para cada regla de inferencia, una variable de entrada x_p se elimina arbitrariamente y el NN de cada parte ENTONCES se entrena usando TRD como el paso 6 Ecuación 10 da el error al cuadrado θ_{m-1}^{sp} del valor del control. Del valor de control de la regla sth en el caso de eliminar x^p Este θ_{m-1}^{sp} puede estimarse utilizando el CHD

$$\theta_{m-1}^{sp} = \sum_{i=1}^{n_c} \{y_i - u_s(x_i) \cdot \mu_A^s(x_i)\}^2, \quad p=1,2,\dots,m \quad (10)$$

Por comparación ecuaciones (8) y (10) es

$$\Theta^s_m > \Theta^{sp}_{m-1}$$

La importancia de la eliminación de las variables de entrada x pueden considerarse mínimas y x_p puede descartarse.

Procesamiento digital de imágenes citogenéticas para su clasificación

- Paso 7: Las mismas operaciones que las realizadas en el paso 5 se llevan a cabo para las variables de entrada $m-1$. Los pasos 5 y 6 se repiten cíclicamente y en el siguiente paso de la ecuación 6 no se mantendrían para variables de entrada. El modelo que da el valor mínimo de O es la mejor NN. Por lo tanto los pasos 1-7 determinan las partes IF y ENTONCES partes de cada regla de inferencia. Luego se completa el proceso de identificación del sistema para el modelo difuso.
- Paso 8: La siguiente ecuación puede derivar el valor control final y :

$$y_i^* = \frac{\sum_{s=1}^r \mu_A^s(x_i) \cdot u_i(x_i)}{\sum_{s=1}^r \mu_A^s(x_i)}, \quad i=1,2,\dots,n. \quad (11)$$

Donde $u_s(x_i)$ es un valor obtenido cuando CHD se sustituye en la mejor NN del paso 7.

Este método es usado para determinar los límites del núcleo definiendo los puntos frontera y así determinar el diámetro del núcleo para la clasificación.

4

Desarrollo matemático y razonamiento de las bases del Prototipo de la herramienta bioinformática

4.1 Panorama general:

Para clasificar la imagen primero se debe conocer como se constituye la misma y como trabajarla. Para eso se organiza un *pipeline* y comienza el desarrollo.

Como se ha visto, la computadora recibe la información, desde una cámara que captura la imagen. Las imágenes para alimentar el modelo deben estandarizarse y limpiarse. El preprocesamiento permite reducir la complejidad y aumentar la precisión del algoritmo haciendo que las imágenes puedan ser procesadas por un algoritmo más general.

Cada imagen se estandariza a través de operaciones como cambiar el tamaño, rotar, transformarla de un sistema de color a otro, o de color a escala de grises, definir un contorno, etc. Solo al estandarizar se puede comparar adecuadamente; (pre procesamiento de datos).

Luego se extraen las características que se distinguen de la molécula de *ADN* en ensayo cometa y debemos distinguir (fondo, cabeza y cola). El resultado de este proceso es un vector de características, que es una lista de formas únicas que identifican el objeto buscado.

Para ver el comportamiento de las funciones, se realizaron algunas pruebas con el siguiente simulador de clasificación, viendo si lograba clasificar, con cuantas épocas y prueba de pérdida. Se puede observar el desempeño con activación sigmoidea en Fig. 14 y con Tanh y Relu en Fig. 15.



Figura 14: Simulación del modelo con activación sigmoidea

Procesamiento digital de imágenes citogenéticas para su clasificación



Figura 15: Simulación con 2 capas ocultas problema con activación a) RELU y activación b) Tanh usando simulador de redes neuronales de playgroundtensorflow.

Con estas características se crea un modelo de clasificación, utilizando la función sigmoidea fundamentalmente por su rango de salida entre 0 y 1. Como se ve en los modelos simulados la activación con RELU y Tanh clasifican en menor cantidad de épocas y en la prueba de perdidas también ofrecen valores inferiores. Si bien RELU es la función más frecuentemente utilizada en redes convolucionales y es no lineal, se elige la activación sigmoidea por el rango de salida, que en este caso de estudio se aplica perfectamente a la salida deseada y los tiempos que requieren las épocas extras son absolutamente despreciables con los GPU y TPU actuales.

Este paso analiza el vector de características y predice la clase a la que pertenece la imagen, lo que hace la experticia del genetista.

En el prototipo:

Procesamiento digital de imágenes citogenéticas para su clasificación

- Primero se observa y diferencia el fondo distinguiéndolo de la molécula de *ADN* y si es posible se mide el diámetro del núcleo. Para esto primero se ecualizan las imágenes con respecto al aporte de luz y rotación a fin de obtener comparaciones equalizadas, o con un entorno similar de análisis. En este punto se habla específicamente de la imagen porque las muestras se preparan siguiendo un procedimiento estricto. Además se las guarda en archivos independientes para ser tratados por Matlab® y GNU Octave.
- Luego se diferencia la cabeza y cola del nucleoide de *ADN*, se transforma la imagen color a una imagen en tonos de grises para observar las diferencias con respecto al fondo, para lo que se binariza la imagen a blanco y negro a los fines de facilitar la detección y medición. La imagen se vuelve binaria.
- Posteriormente la cabeza y cola ya diferenciadas del fondo se miden.
- Finalmente se mide el diámetro de la cabeza del nucleoide de *ADN* y se lo compara con la medida de la cola estableciendo las diferentes relaciones definidas en el modelo de clasificación.

Este proceso determina y clasifica las moléculas en clases de acuerdo a las características oportunamente definidas.

Aunque pueda clasificar cada clase, el conjunto de prueba (*test set*) en el primer ejemplo de prototipo, arroja una cantidad de errores, a medida que se va ajustando el algoritmo, se vuelve más preciso. El principal problema que se suscitó en este punto, además de determinar correctamente los umbrales, fue encontrar un número suficiente de imágenes. Conjunto de datos (*Dataset*) tan específicos de citogenética no se encuentran, porque el preparado, (específicamente la fluorescencia) es sensible a la exposición de la luz, en muy pocas ocasiones se toman algunas elegidas para publicaciones siendo incluso escasas las imágenes publicadas. Por lo tanto fue un reto el hallazgo.

El banco de datos de 200 imágenes, compuesto por imágenes de cometas registradas en el laboratorio de citogenética durante un lapso de tiempo

Procesamiento digital de imágenes citogenéticas para su clasificación

comprendido entre los años 2009 hasta la actualidad aproximadamente, donde en el caso de peces se refiere a la especie *Steindachnerina brevipinna* (sabalito), *Piaractus mesopotamicus* (pacú) y en moluscos pertenecen a las especies *Corbícula fluminea* y *Limnoperna fortunei*. Se trabajo con imágenes de cometa, procedente de varias especies, porque no se consiguen suficientes imágenes de las muestras de una sola especie, para integrar el conjunto de datos (*dataset*) con que se entrena y valida el modelo. Si bien para el estudio en el laboratorio son cientos de preparados que se estudian, los registros fotográficos se toman solamente a las mejores muestras dado que la exposición a la luz desgasta la fluorescencia, se apaga el cometa y no se puede seguir analizando. Por eso quedan las mejores imágenes de los registros fotográficos para uso de los investigadores en publicaciones, poster y revistas científicas. Se inicia la instancia de entrenamiento y pruebas de algoritmos. Al realizar las pruebas de clasificación, con un grupo de entrenamiento compuesto por 78 ejemplos, muy específicos primero, (el grupo de peces *Steindachnerina brevipinna* y *Piaractus mesopotamicus*), y luego con imágenes de los bivalvos, se logra considerar y evaluar el rendimiento y desempeño del modelo en la clasificación propuesta por el prototipo, y a la vez gran flexibilidad en cuanto se presentan conjunto de datos (*datasets*) con nuevos elementos, inclusive de especies diferentes. En el laboratorio se trabaja con esta técnica aplicada a distintos modelos biológicos y dentro de cada modelo biológico a distintos tejidos, donde cada tejido y cada animal (pez molusco, etc), tiene sus particularidades biológicas, o sea las células no son iguales por eso tampoco los tamaños de los cometas y su morfología tampoco van a ser iguales. Por esta razón también se necesita suficiente flexibilidad.

Al incorporar diferentes imágenes se debe hacer más pre- procesamiento para limpiar las imágenes y extraer las características más fácilmente. Para lograrlo se escalan y transforman las imágenes y se guardan en archivos independientes para luego ser tratadas por Matlab® u Octave. Operación que además, es valor agregado ya que permite tener un registro e histórico de las muestras clasificadas.

Procesamiento digital de imágenes citogenéticas para su clasificación

En este paso también, inicialmente se ajusta muchas veces el programa, cambiando el algoritmo clasificador en Matlab®, como así también buscando encontrar la detección del umbral más adecuado. En la investigación se probó Detección de región extrema y transformada de Hough, verificando las posiciones del acumulador con mayor valor, máximos locales en el espacio del acumulador. Matlab® utiliza *Houghpeack* para identificar los picos de la transformada y *Houghlines* extrae los segmentos de línea basados en la transformada. Aunque finalmente se utilizó *regionprops* que devuelve medidas para el conjunto de propiedades especificado para cada objeto en la imagen binaria que también puede trabajar la imagen volumétrica con parámetros para 3d e incluso devolver más estadísticas, y *struct array* que contiene una estructura para cada objeto de la imagen. Las primeras pruebas del funcionamiento se hicieron con un grupo de prueba compuesto por pocas imágenes 1 de cada clase en principio y luego 1 de cada clase de validación. A lo largo del proceso se prueba el prototipo con 30 imágenes, luego de las primeras pruebas, se hicieron pruebas con 48 más y se validó con 122 imágenes. Todas las tareas se suman para mejorar el rendimiento del modelo. En este prototipo las pruebas lograron clasificar con precisión, usando funciones sencillas como veremos más adelante. Las imágenes atraviesan procesos para lograr su clasificación digital, logrando un prototipo funcional. En este proceso se llevaron a cabo los siguientes pasos:

1. Selección de imagen de entrada.
2. Preprocesamiento de Imágen.
3. Extracción de características.
4. Clasificación.

En el proceso de reconocimiento de imágenes las fases fundamentales son:

- Extracción de rasgos o características del sistema físico.
- Selección de características.
- Obtención de la regla de decisión (diseño del clasificador).
- Utilización de la regla adaptativa.

Procesamiento digital de imágenes citogenéticas para su clasificación

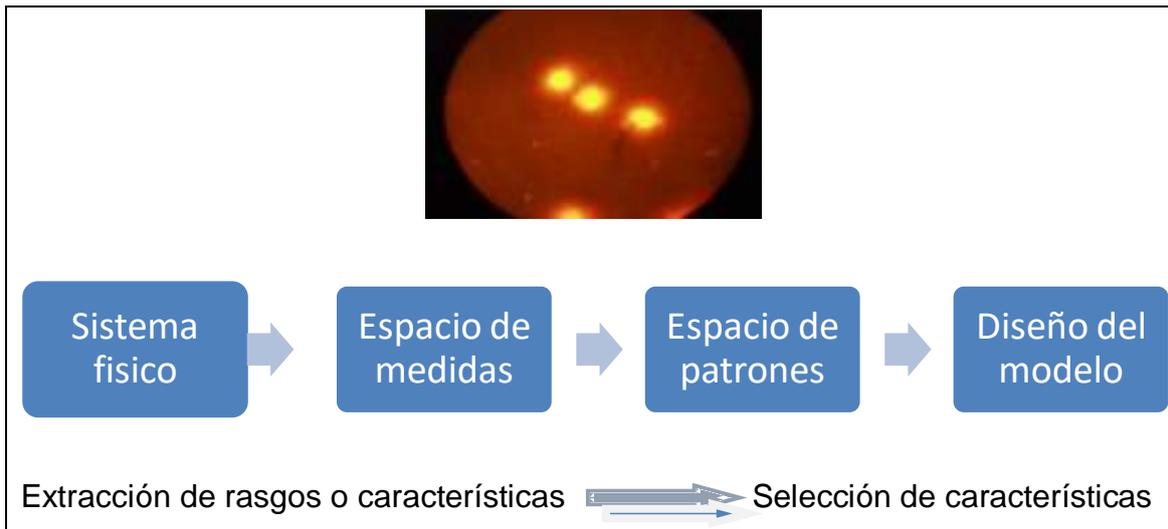


Figura 16: Imagen de la molécula de ADN luego de la preparación en Ensayo Cometa y el modelo de estrategia de reconocimiento de patrones de imágenes. Fuente propia.

En el esquema de la Fig16 se detalla la propuesta de *pipeline*, las muestras del medio son capturadas (Sistema Físico), medidas (Espacio de medidas) y transformadas a una imagen homogénea (Espacio de patrones), desde donde se extraen los descriptores de las características elegidas para luego pasar por un clasificador desarrollado a partir de las características (Diseño del modelo). Dicho de otra manera, para que el proceso que realiza el ojo humano distinguiendo imágenes y clasificándolas, se pueda realizar a través de un modelo de reconocimiento de patrones de imágenes, donde la computadora pueda ver y a través de filtros y modelos, reconocer y clasificar las imágenes imitando el comportamiento humano (Fig.17).

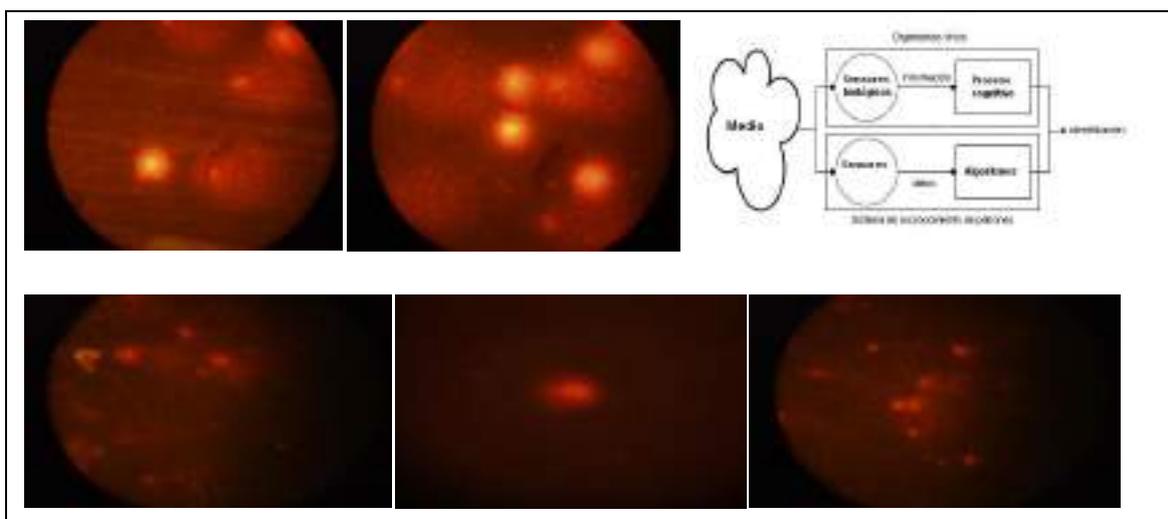


Figura 17: Ejemplo de Reconocimiento de patrones de imágenes; la imagen cruda y transformada para lograr un histograma uniforme. Fuente propia.

Procesamiento digital de imágenes citogenéticas para su clasificación

Para la elaboración del algoritmo solución se probaron diferentes estrategias, aplicando al caso de estudio la estrategia de reconocimiento adaptativo Fig18 y se midió el rendimiento entre clases.



Figura 18: Recorrido de la imagen a través del modelo de entrenamiento de reconocimiento adaptativo para su clasificación. Fuente propia.

En nuestro modelo se aplica la estrategia de reconocimiento adaptativo donde primero se homogeneizan los datos a fin de que el tratamiento tenga el mismo histograma de base como vemos en la Fig. 19.

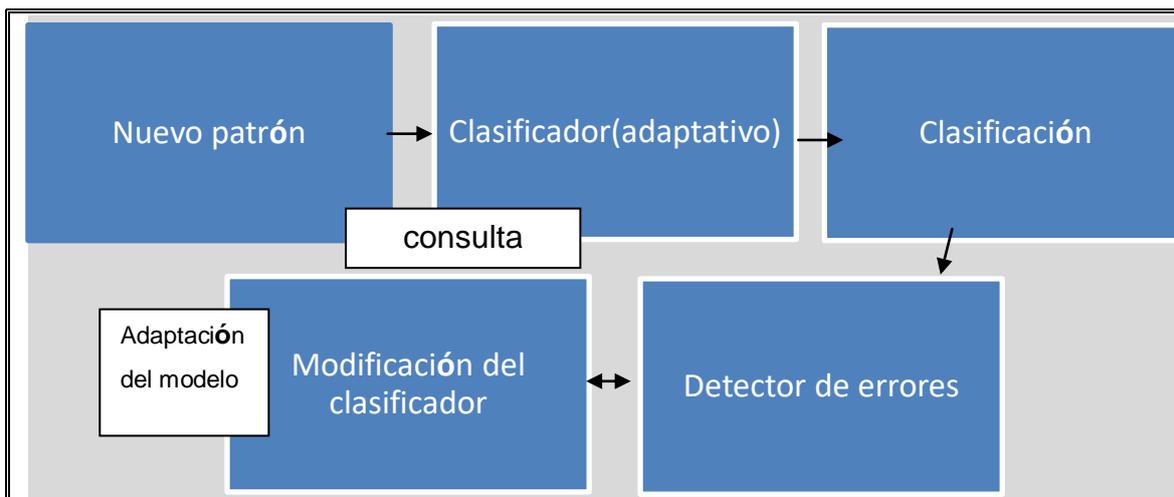


Figura 19: Adaptación del Modelo. Fuente propia.

Para llevar a cabo el diagnóstico es necesaria la detección y segmentación de la imagen de la molécula de ADN en su conjunto. Y más allá de la forma utilizada para procesar la información, es difícil encontrar un método lo suficientemente robusto, preciso y a la vez lo suficientemente flexible para que pueda adaptarse al cambio que se registra en las muestras.

Primero la segmentación de la molécula de ADN requiere identificar los diferentes componentes de la imagen, en nuestro caso de estudio corresponden a cabeza, cola y fondo del nucleoide de ADN. Para ello se realiza

Procesamiento digital de imágenes citogenéticas para su clasificación

una descomposición morfológica de la imagen usando funciones de segmentación y luego se homogeneizan datos a partir de filtros.

La intensidad y color de la cola se pueden confundir con el fondo y por lo tanto es posible que se obtengan errores en el agrupamiento debido a la dificultad para aislarlos. El mismo problema se produce al definir la cabeza del nucleoide y determinar cuáles son los límites.

El desafío en esta etapa es lograr la suficiente flexibilidad para determinar los límites de la cabeza del nucleoide y a la vez utilizar las mismas reglas para medir todas las cabezas de nucleoides de ADN, o bien el mismo sistema de decisión para determinar en la imagen, la región en que se observa la cabeza del nucleoide, cola y fondo y cuáles son sus límites (Fig. 20). El nucleoide es la parte central mas iluminada, a continuación se presenta la cola, es la parte brillante mas difuminada, y todo lo demás fondo, aunque a veces se presenta más de un nucleoide en la imagen.

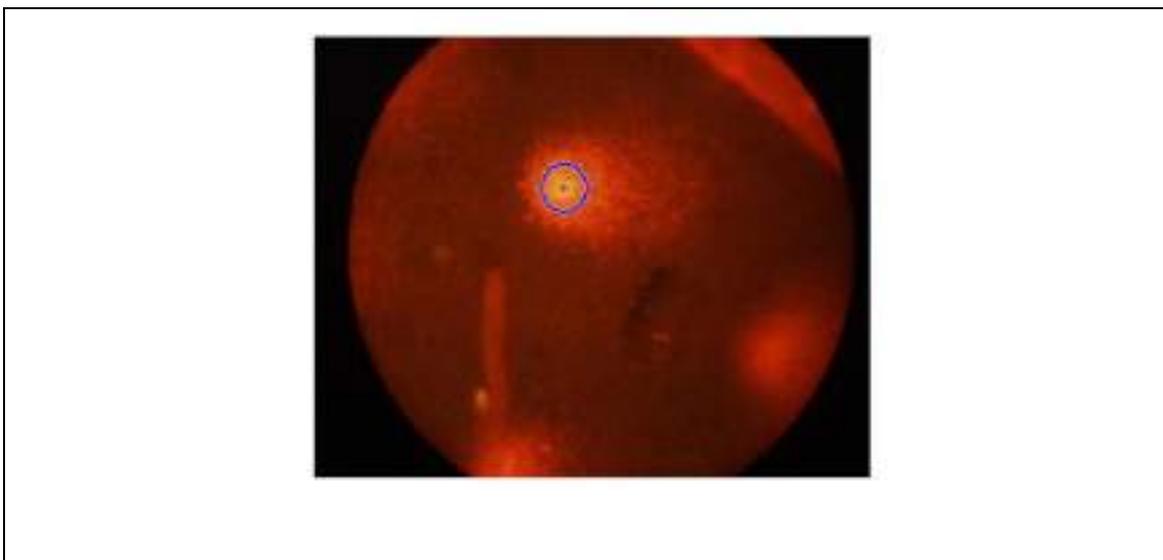


Figura 20: Primera aproximación de detección de núcleos o nucleoides de ADN. Se observa la complejidad para definir las fronteras ó limites.

Finalmente, A pesar de que las muestras sigan un mismo protocolo de preparación del material biológico, en el tratamiento experimental, las imágenes no siempre son homogéneas. Las imágenes del registro fotográfico de células

Procesamiento digital de imágenes citogenéticas para su clasificación

fenotípicamente diferentes entre si, que incluso se muestren diferentes, pueden pertenecer a una misma clase. (Fig. 21).

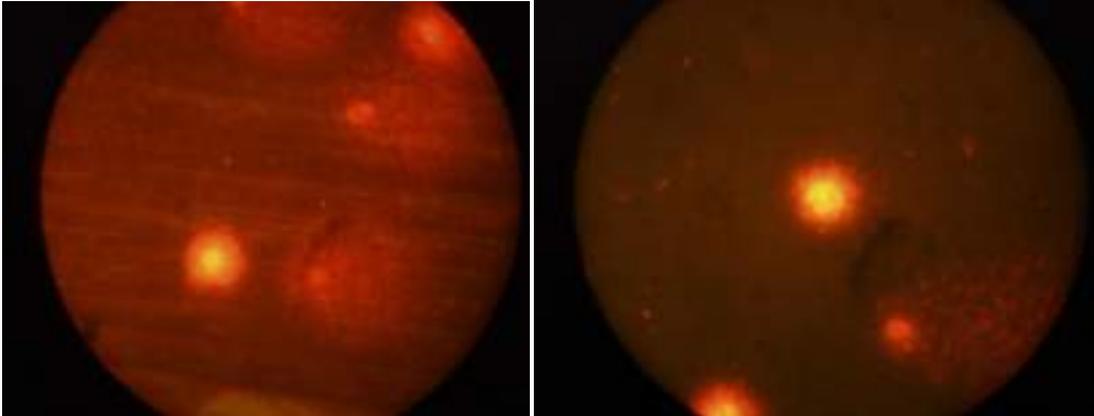


Figura 21: Obsérvese dos imágenes de Cometa que pertenecen a la misma clase.

La segmentación de la cabeza del nucleoide de la molécula de ADN se realiza en diferentes pasos, primero se detecta y localiza la cabeza y cola, posteriormente se lo delimita y mide.

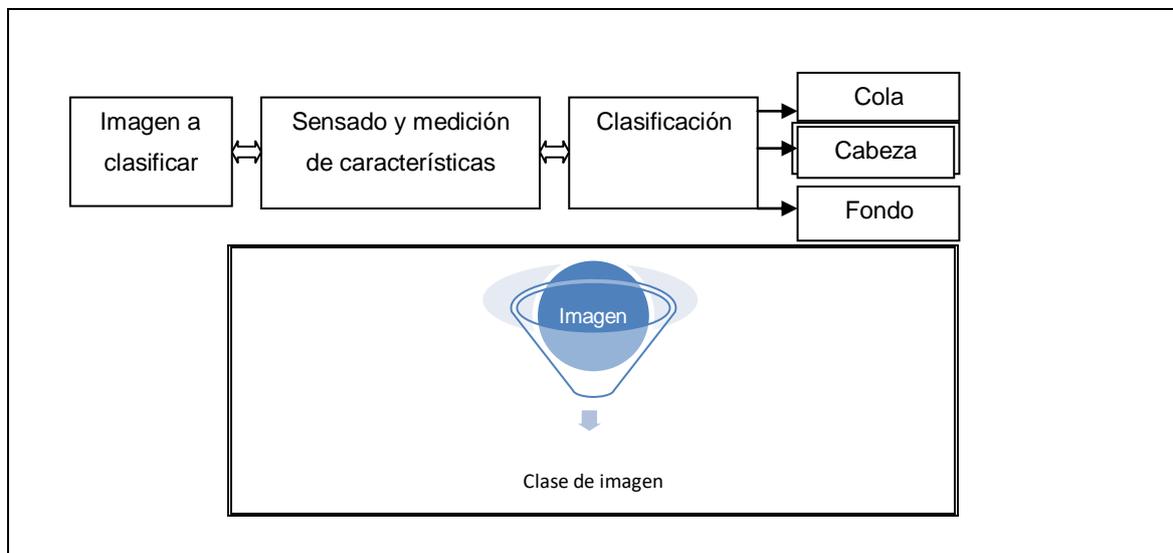


Figura 22: Clasificación de acuerdo a la composición morfológica con respecto al modelo.
Fuente propia

El procedimiento realizado, comienza con la obtención de la fotografía, fig. 23 a partir de la misma el diseño de reconocimientos de patrones debió resolver lo siguiente:

Procesamiento digital de imágenes citogenéticas para su clasificación

- Problemas de preprocesamiento o selección de características.
- Problemas de cálculo o medición de las características.
- Problemas de identificación y clasificación o agrupamiento.

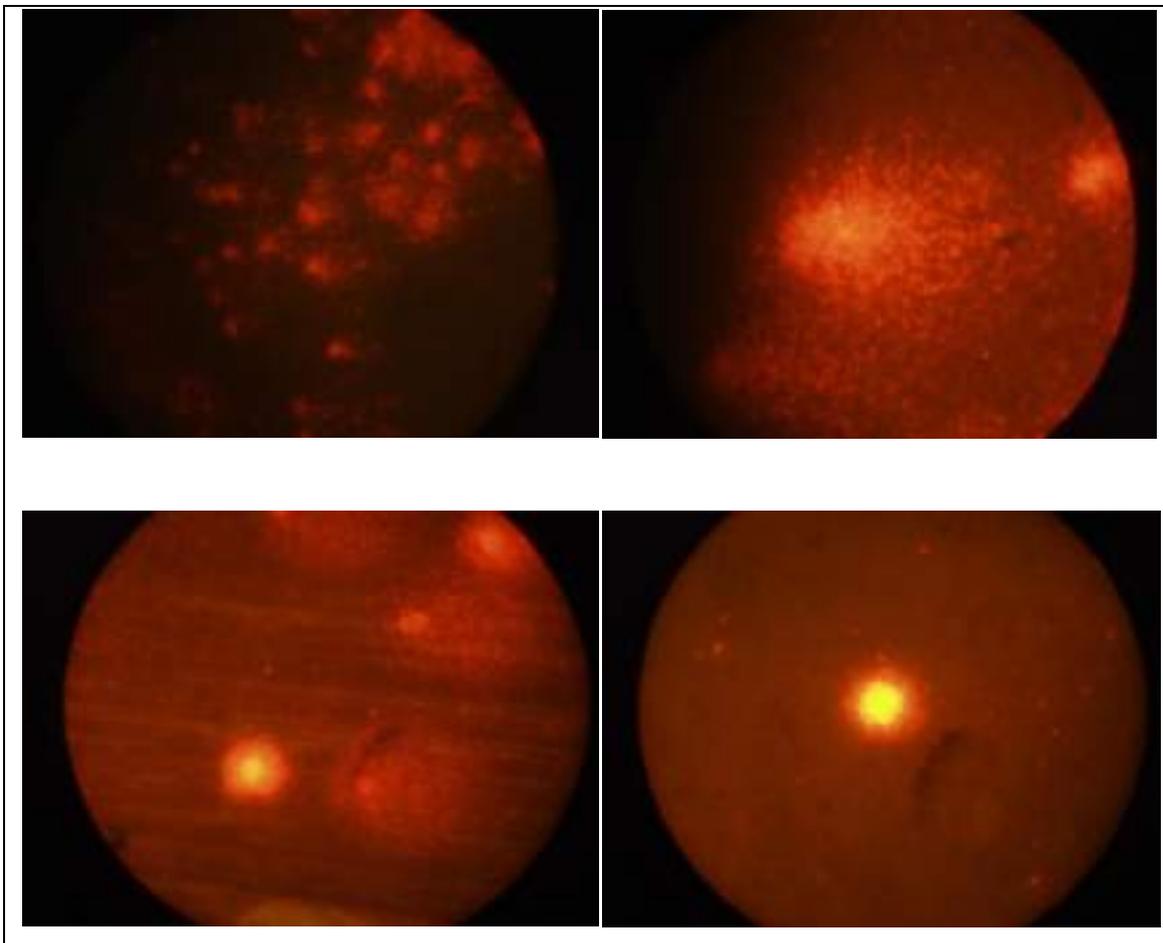


Figura 23: Diferentes fotografías de nucleoides de ADN antes del pre procesamiento. Se presentan con diferencias no solo de cantidad de núcleos sino también de luz y posición.

4.2 Desarrollo matemático del predicado

Primitiva de representación:

La primitiva de representación es el pixel. La representación discreta (imagen digital) es una aproximación, dado que se tiene un número finito de muestras. El pixel representa intensidad luminosa en niveles de gris, y puede representar cualquier magnitud física, química o biológica.

La imagen para clasificar se encuadra dentro de imágenes multi-nivel, L niveles de gris usualmente $L=256$. $U L$ Imágenes de niveles de gris (escalar).

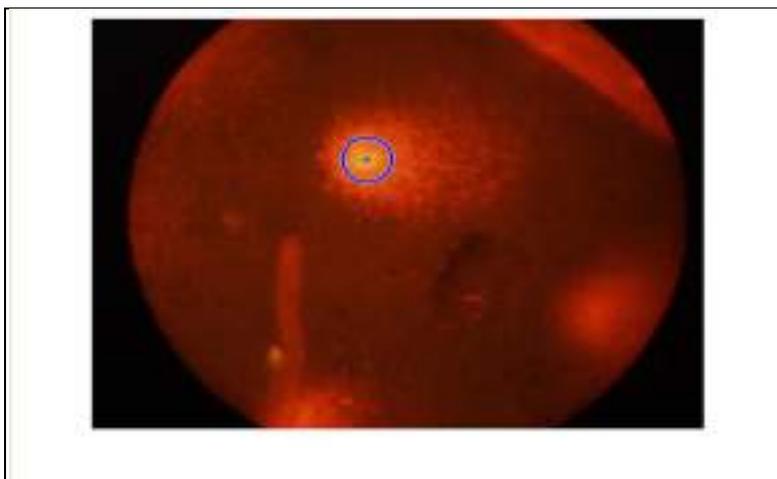


Figura 24 : Medición del núcleo del cometa antes de transformar a niveles de gris. Fotografía tomada del programa luego de determinar el centroide Fuente propia.

Como se ve en la Fig. 24 antes de las transformaciones es muy difícil determinar la frontera o límites a pesar de que a simple vista parecería sencillo. El límite depende mucho del criterio o la intuición, aunque como se ha mencionado antes en este caso de estudio se utilizaron funciones que permiten determinar las regiones y los umbrales. Así también se recurre a la transformación en escalas de grises como parte del preprocesamiento.

La capa de entrada toma los valores de todos los píxeles. Al pasar a la capa oculta se extraen las características de una imagen, y se realizan dos operaciones fundamentalmente, agrupación y la definición de un conjunto de filtros entrenables que procesan la imagen junto con los valores de la capa precedente. En la práctica estos valores son aprendidos para su activación, al encontrar determinadas características. Si son puestos en cascada se obtienen diferentes grados de abstracción o profundidad. Dicho de otra manera, estos filtros devuelven las características que se evalúan en una imagen y cada pasada de cada filtro toma alguna de ellas por eso es tan importante esta etapa. Luego a través de la agrupación se hace una matriz más pequeña aun, para que el número de conexiones sea lo suficientemente manejable.

Filtrado en el dominio espacial: filtros lineales basados en máscaras de convolución

Un filtro de convolución puede representarse como una matriz cuadrada o rectangular (matriz de convolución) de dimensiones mucho más pequeñas que la imagen sobre la que se aplicará. Esta matriz se desplaza sobre la imagen de manera que el elemento central de la misma coincida con cada uno de los píxeles de la imagen como vemos en la Fig. 25, definiendo al píxel por ejemplo como $p(x)$ =el píxel es gris oscuro.

En cada posición se multiplica el valor de cada píxel de la imagen por el valor de la posición del elemento de la matriz. El píxel de la imagen que coincide con el elemento central de la matriz es reemplazado luego por la suma de los productos. Posteriormente se realiza el agrupamiento y se obtiene la capa de convolución, y capa de agrupamientos intercalados, tantas veces como filtros tenemos. Las principales características en el proceso de segmentación son: color, forma e intensidad de gris o distinción de blanco y negro en nuestro caso.

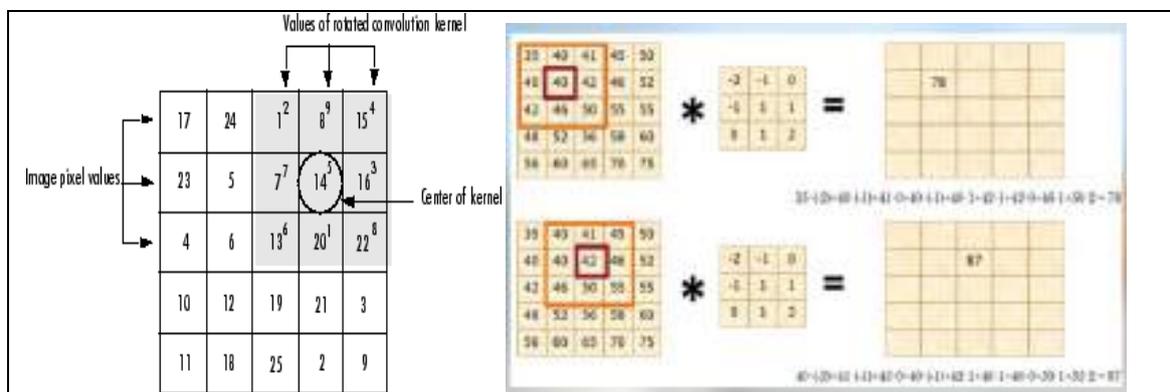


Figura 25: Como recorre la imagen la matriz de convolución.

4.3 Clasificador

Datos: Características que permiten pertenecer a la membresía de la clase. De un tratamiento experimental de 10 a 15 individuos por tratamiento, se contabilizan 100 células por individuo y las mismas se clasifican de acuerdo al largo e intensidad de la cola de los cometas, que son equivalentes al ADN fragmentado o con daño.

Individuo a clasificar: se presenta en las fotografías como un cometa, con un núcleo más claro y un halo alrededor, de diferentes formas de acuerdo al daño de *ADN* (como veremos más adelante) que determinará la clase a la que pertenece. El halo muy disgregado pertenece a clase 4 que incluye a los anómalos.

Clases:

Clase 0 = sin cola, solo cabeza ó núcleo (sin daño) Fig. 26

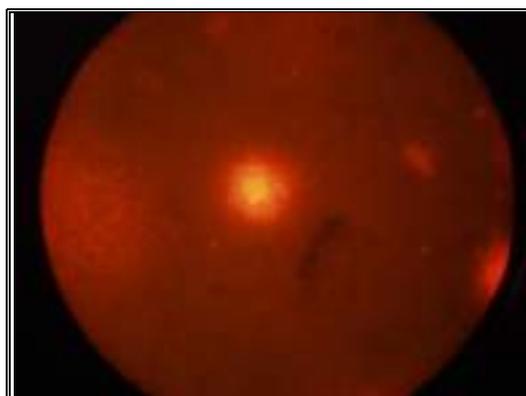


Figura 26: Imagen clase 0. El individuo a clasificar se presenta como un cometa sin cola. se observa el núcleo con mucho brillo, la cola alrededor con menor brillo y el fondo de color naranja

Clase 1 = tamaño de la cola hasta una vez el diámetro de la cabeza Fig. 27



Figura 27: Imagen clase 1, se pueden observar los núcleos con un color más claro y la cola tiene tamaño del diámetro de la cabeza.

Clase 2 = tamaño de la cola hasta dos veces el diámetro de la cabeza Fig. 28

Procesamiento digital de imágenes citogenéticas para su clasificación

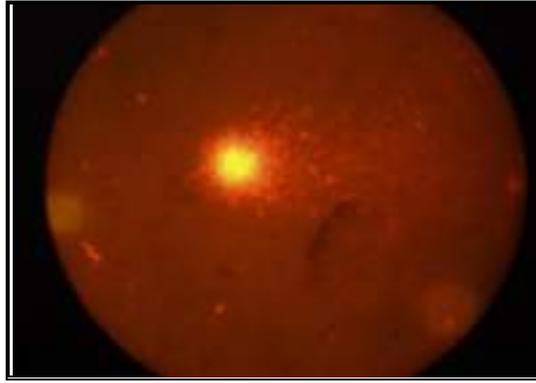


Figura 28: Imagen clase 2 la cola se extiende más de dos veces el diámetro del nucleoide.

Clase 3 = tamaño de la cola hasta tres veces el diámetro de la cabeza Fig.29

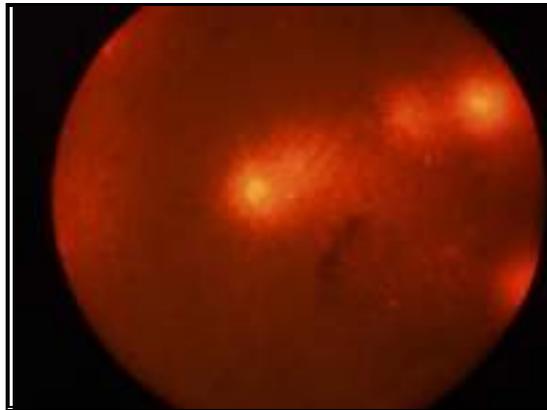


Figura 29: Imagen clase 3 se observa la cola que se extiende 3 veces el diámetro del nucleoide.

Clase 4 = casi todo el ADN aparece fragmentado, en este grupo se incluyen los anómalos Fig. 30.

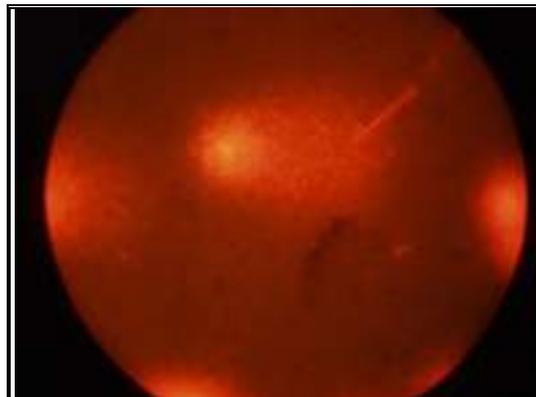


Figura 30: Imagen clase 4 se observa el núcleo y la cola largamente disgregada.

Procesamiento digital de imágenes citogenéticas para su clasificación

Objetivo: clasificar cada molécula de *ADN* dentro de la clase o categoría a la que pertenece Fig. 31.

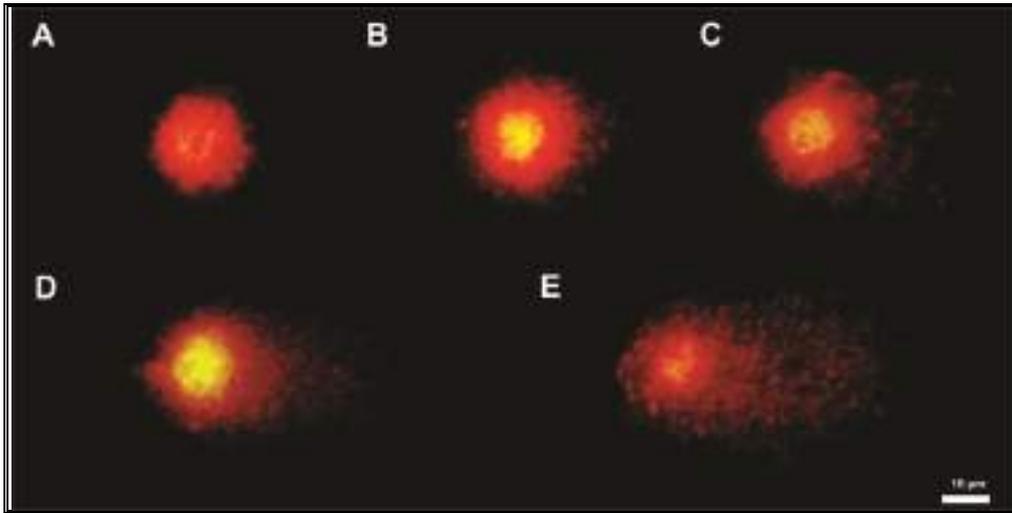


Figura 31: Muestra de las diferentes clases de cometas. Las letras A,B,C,D,E, se corresponden con las clases 0, clase 1, clase 2, clase 3 y clase 4 donde también se encuentran los anómalos. Fuente: Fotografía editada Laboratorio de Citogenética (UNaM-IBS-CONICET).

El proceso que se realiza es: al ver una imagen, el prototipo logre clasificarla y tenga la suficiente flexibilidad en la distinción de la cola del fondo a través de algoritmos de lógica difusa.

De la clasificación se obtienen las membresías a las clases que vemos en la fig. 31. También se usan algoritmos de aprendizaje supervisado teniendo como referencia en la Fig. 32 el tratamiento del error y el aprendizaje.

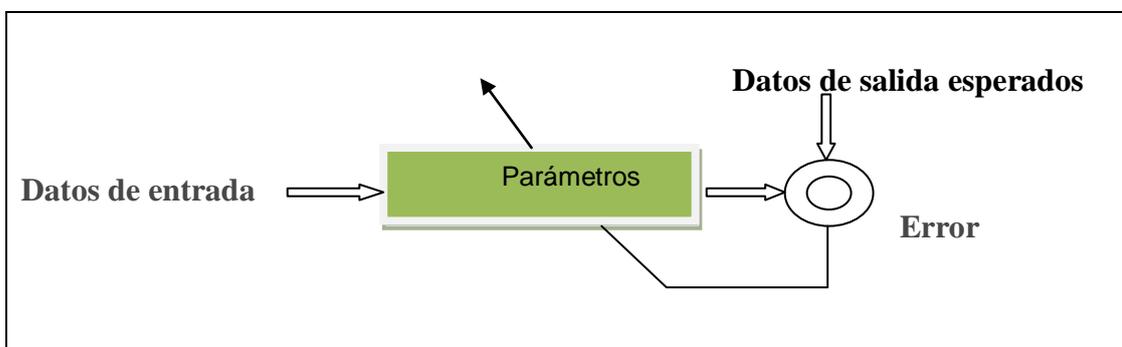


Figura 32 Aprendizaje supervisado.

4.4 Agrupamiento o clasificación

Con la Matriz de convolución se pueden trabajar diferentes filtros, por ejemplo:

Enfoque, Desenfoque, Realce de bordes, Detección de bordes, Filtro de tipo Sobel, Filtro de tipo Sharpen, y lograr un vector de características(Fig.33).En Matlab® se utiliza como predeterminado el filtro de Sobel `BW=edge(I)` devuelve una imagen binaria que contiene `s` donde la función encuentra bordes en la imagen de entrada y `s` en otro lugar.`BW` de forma predeterminada, utiliza el método de detección de bordes Sobel, que encuentra bordes en aquellos puntos donde el degradado de la imagen es máximo, utilizando la aproximación de Sobel a la derivada. Con el método “log” encuentra bordes buscando cruces cero después de filtrar con un filtro Laplacian de Gaussian (LoG).

Para los métodos de detección de borde de magnitud de gradiente (Sobel, Prewitt, Roberts), se utiliza como umbral la magnitud de gradiente **calculada.edgethreshold**.

Para los métodos de cruce cero, incluyendo Laplacian de Gaussian, se utiliza como umbral para los cruces **cero.edgethreshold**. En otras palabras, un gran salto a través de cero es un borde, mientras que un pequeño salto no lo es.

El método Canny aplica dos umbrales al degradado: un umbral alto para la sensibilidad de borde bajo y un umbral bajo para la sensibilidad de borde alta.

Comienza con el resultado de baja sensibilidad y luego lo crece para incluir píxeles de borde conectados del resultado de alta **sensibilidad.edge** Esto ayuda a rellenar huecos en los bordes detectados. De todas maneras en el algoritmo desarrollado se utilizo **Imclose** para rellenar los huecos en los bordes

En todos los casos, elige el umbral predeterminado heurísticamente, dependiendo de los datos de **entrada.edge**. La mejor manera de variar el umbral es ejecutarse una vez, capturando el umbral calculado como el segundo argumento de **salida.edge** A continuación, a partir del valor calculado por ajuste el umbral más alto para detectar menos píxeles de borde o más bajo para detectar más píxeles de **borde.edge**.

Procesamiento digital de imágenes citogenéticas para su clasificación

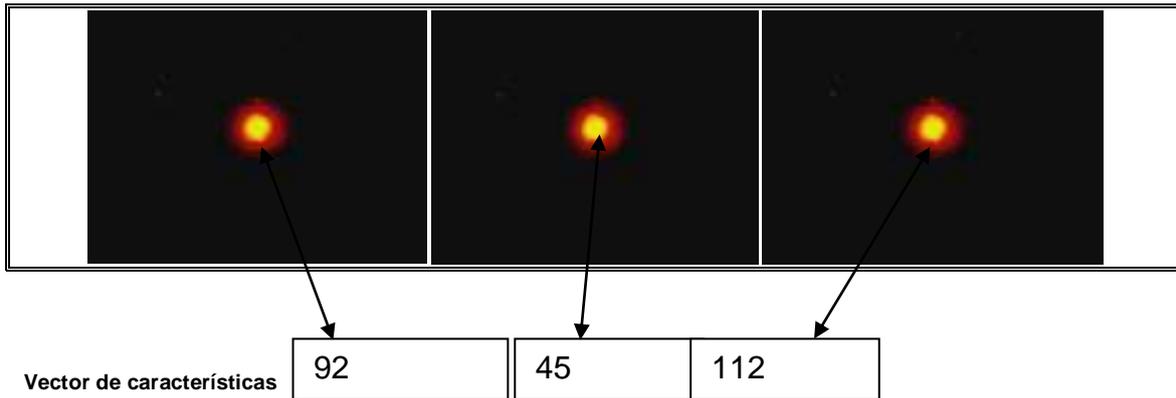


Figura 33 Ejemplo de empleo de un vector de características multiresolución.

Luego de aplicar los filtros a través de funciones en este caso, es fácilmente medible tanto la cabeza de nucleoides de *ADN* (se determina el centro del centroide y se traza una línea para determinar el diámetro, se utilizan también funciones seno y coseno para determinar las coordenadas) se mide luego el nucleoide y como diferencia con respecto a la cabeza se obtiene la cola, y además se pueden diferenciar claramente del fondo mediante la binarización. Se puede observar la fotografía cruda Fig34 antes del pre-procesamiento.

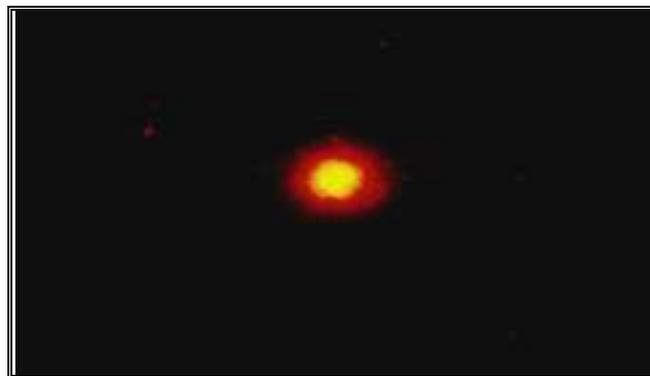


Figura 34: Fotografía tomada sobre el microscopio en condiciones de total oscuridad. Imagen cruda. Fuente propia.

Al realizar el cambio de escala de color a grises podemos observar con mayor claridad los diferentes aspectos del nucleoide de *ADN* y diferenciar sus características morfológicas, separar la cabeza y cola despegándolo del fondo, para posteriormente diferenciar la cola del fondo y distinguir la cabeza del

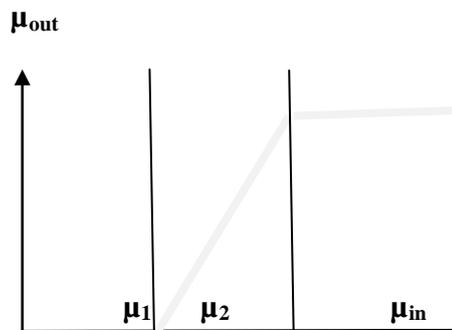
Procesamiento digital de imágenes citogenéticas para su clasificación

nucleoide de la cola para poder iniciar el proceso de clasificación. Para esto realizamos una serie de operaciones como por ejemplo:

- Función de transferencia para realce de contraste

Esta función se aplica para cambiar a escala de grises y luego se aplican las funciones para determinar el umbral como se ha explicado. Se observa con claridad las diferentes características que necesitamos aislar para poder medir y clasificar.

$$\mu_{out} = T(\mu_{in}) = \begin{cases} 0 & \mu_{in} \leq \mu_1 \\ 2.5 (\mu_{in} - \mu_1) & \mu_1 \leq \mu_{in} < \mu_2 \\ 255 & \mu_{in} \geq \mu_2 \end{cases}$$



4.5 Realce de contraste

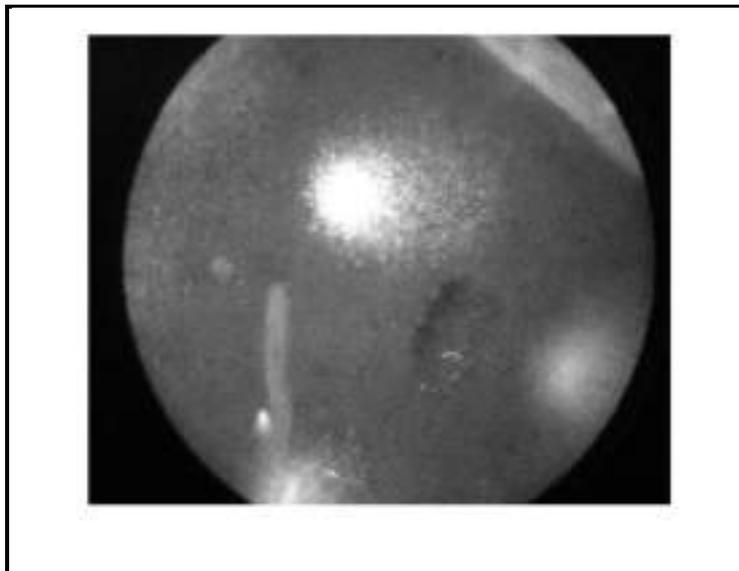


Figura 35: Imagen luego del pre procesamiento (imagen preparada en escala de grises). Fuente: imagen modificada de fotografía tomada en Laboratorio de citogenética (UNaM-IBS-CONICET) como resultado del algoritmo desarrollado.

Procesamiento digital de imágenes citogenéticas para su clasificación

Esta imagen es producto de la función **IMADJUST** en Matlab®, que nos permite lograr un histograma similar, como vemos en la Fig. 35, esta función hace que las luces y sombras se estandaricen y al homogeneizar la imagen, con los mismos parámetros de luz y sombra, se logra un histograma ecualizado para analizar cada imagen y así clasificarla.

Esta función es muy importante dado que en el material biológico, los nucleoides de *ADN*, como vimos anteriormente en las fotos, las imágenes del objeto de estudio pueden aparecer rotadas, superpuestas o con mayor aporte de luz o sombra de acuerdo al lugar de la muestra en que se encuentren: a pesar de que las muestras se preparen para que las colas migren hacia el mismo lado.

En la Fig. 36 se ve como la imagen se lleva a binario, blanco y negro. El umbral se establece a través de la función **regionprop** en `BW = imread('text.png');` que determina el centroide y posición del objeto de estudio, además establece regiones luego de completar los espacios haciendo solido el borde a través de **Imclose**. Se puede observar la imagen con los centroides superpuestos a través de la función: `imshow(BW) hold on plot(centroids(:,1),centroids(:,2),'b*')` hold off.

Este proceso nos permite establecer cabeza y cola con claridad, aunque la imagen aun tiene cierta porosidad con espacios en blanco que genera el material biológico.

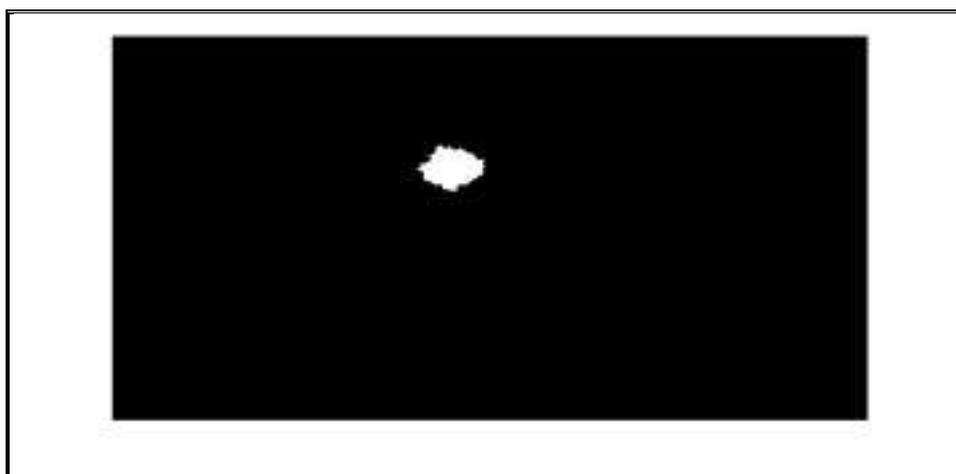


Figura 36: Imagen binaria (blanco y negro) de la molécula de ADN luego de aplicar la función `Imclose`.

Procesamiento digital de imágenes citogenéticas para su clasificación

Luego a través de la función `IMCLOSE` esos espacios en blanco o poros se van completando hasta lograr una imagen sólida que nos permite definir la frontera y determinar automáticamente a través de algunas operaciones matemáticas el límite. Lo interesante en esta parte del proceso es encontrar el valor adecuado para el cierre de la figura.

Luego de esta transformación de la imagen, es más fácil realizar las operaciones definiendo fronteras. Se calcula el centro estableciendo dos medidas de diámetros. Se define un centroide (se mide el largo y el ancho) a través de las funciones mencionadas anteriormente y se traza una línea midiendo además el diámetro como vemos en la Fig. 37.

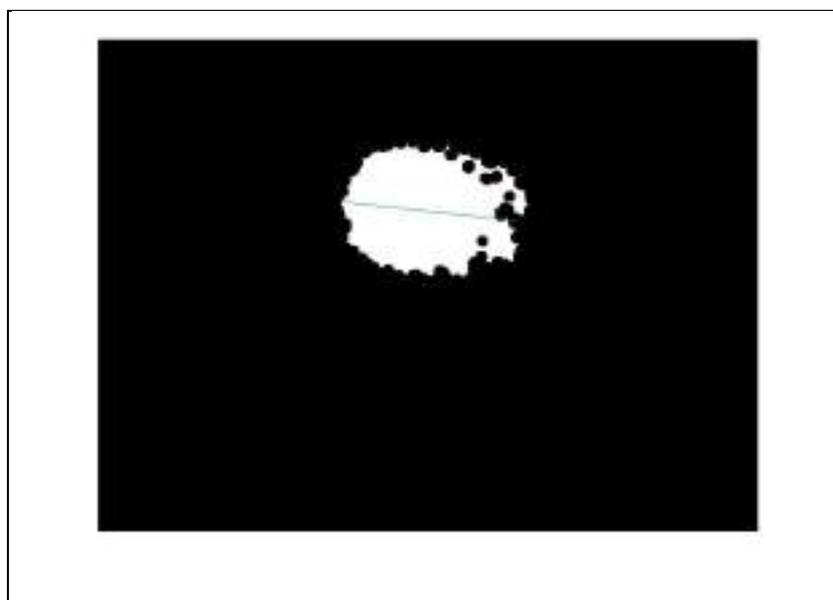


Figura 37: Imagen binaria de la célula. Se puede observar además la línea definida por el algoritmo que mide el diámetro del nucleoide mas la cola.

4.6 Código del clasificador para Matlab® y Octave

En el código del prototipo clasificador vemos comentadas las funciones descriptas anteriormente:

```
clc
clear all
I=imread('P1010174.jpg');
S = sum(I,3);
meanI= max(S(:)) %meanI=mean2(I)
```

Procesamiento digital de imágenes citogenéticas para su clasificación

% Carga las imágenes

```
A = im2bw(I,0.53);
```

```
se = strel('disk',5);
```

```
A = imclose(A,se);
```

```
B = im2bw(I,0.41);
```

```
se = strel('disk',13);
```

```
B = imclose(B,se);
```

% Halla la diferencia de las imágenes

```
closeBW=B-A;
```

```
figure
```

% Define los bordes de las dos imágenes

```
dim = size(A);
```

```
col = round(dim(2)/2)-90;
```

```
row = min(find(A(:,col)));
```

```
boundary1 = bwtraceboundary(A,[row, col],'N');
```

```
dim = size(B);
```

```
col = round(dim(2)/2)-90;
```

```
row = min(find(B(:,col)));
```

```
boundary2 = bwtraceboundary(B,[row, col],'N');
```

% Halla el centro de el núcleo

```
rx=mean(boundary1(:,2));
```

```
ry=mean(boundary1(:,1));
```

```
numofel=numel(boundary2(:,1));
```

```
maxx=0;
```

```
for i=1:numofel
```

```
    if sqrt((boundary2(i,2)-rx)^2+(boundary2(i,1)-ry)^2)>maxx
```

```
        maxx=sqrt((boundary2(i,2)-rx)^2+(boundary2(i,1)-ry)^2);
```

Procesamiento digital de imágenes citogenéticas para su clasificación

```
Mx=boundary2(i,2);
My=boundary2(i,1);
end
end

minn=999999;
for i=1:numofel
    if sqrt((boundary2(i,2)-rx)^2+(boundary2(i,1)-ry)^2)<minn
        minn=sqrt((boundary2(i,2)-rx)^2+(boundary2(i,1)-ry)^2);
        mx=boundary2(i,2);
        my=boundary2(i,1);
    end
end

%Muestra los resultados
imshow(I)
hold on;
plot(boundary1(:,2),boundary1(:,1),boundary2(:,2),boundary2(:,1),'g','LineWidth',1);
plot(rx,ry,'*b')
plot(Mx,My,'*b')
plot(mx,my,'*b')

maxx
minn
```

Algoritmo 2

lc

Procesamiento digital de imágenes citogenéticas para su clasificación

```
clear all
```

```
%Comentar esta línea para ejecutar en Matlab
```

```
pkg load image
```

```
%Carga la imagen
```

```
prompt = 'Introduzca el nombre del archivo:';
```

```
iname = input(prompt)
```

```
I=imread(iname);
```

```
res=size(I(:,:,1)); %Determina la resolución de la imagen de entrada
```

```
grayI=rgb2gray(I);
```

```
grayI=imadjust(grayI);
```

```
%A = im2bw(I,double(max(grayI(:)))/422); %Pasa a blanco y negro la imagen del nucleo
```

```
A = im2bw(I,0.53); %Pasa a blanco y negro la imagen del nucleo
```

```
se = strel('disk',round(mean(res(:))/200),0);
```

```
A = imclose(A,se); %Cierra agujeros en la imagen del nucleo
```

```
CCA = bwconncomp(A,8); %Detecta objetos conectados de la imagen del nucleo
```

```
SA = regionprops(CCA, 'Area');
```

```
L = labelmatrix(CCA);
```

```
A = ismember(L, find([SA.Area] >= 0.6*max([SA.Area]))); %Elimina todos los objetos menores en un 40% al mayor
```

```
SA = regionprops(A, 'Centroid','MajorAxisLength','MinorAxisLength');
```

```
centrosA = cat(1,SA.Centroid);
```

```
radiosA=mean([SA.MajorAxisLength SA.MinorAxisLength],2)/2;
```

```
figure(1);
```

Procesamiento digital de imágenes citogenéticas para su clasificación

```
imshow(A);
```

```
% Repite las mismas operaciones para la cola
```

```
% B = im2bw(I, mean(grayI(:))/104);
```

```
B = im2bw(I, 0.30);
```

```
se = strel('disk', round(mean(res(:))/100), 0);
```

```
B = imclose(B, se);
```

```
CCB = bwconncomp(B, 8);
```

```
SB = regionprops(CCB, 'Area');
```

```
L = labelmatrix(CCB);
```

```
B = ismember(L, find([SB.Area] >= 0.6 * max([SB.Area])));
```

```
SB = regionprops(B, 'Centroid', 'MajorAxisLength', 'Orientation');
```

```
centrosB = cat(1, SB.Centroid);
```

```
figure(2);
```

```
imshow(B);
```

```
% Dibuja una línea por el centroide
```

```
hlen = SB.MajorAxisLength/2;
```

```
xCentre = SB.Centroid(1);
```

```
yCentre = SB.Centroid(2);
```

```
cosOrient = cosd(SB.Orientation);
```

```
sinOrient = sind(SB.Orientation);
```

```
xcoords = xCentre + hlen * [cosOrient -sinOrient];
```

```
ycoords = yCentre + hlen * [-sinOrient cosOrient];
```

```
line(xcoords, ycoords);
```

Procesamiento digital de imágenes citogenéticas para su clasificación

```
rela=99;

figure(3);

imshow(I);

hold on;

if sum(centrosA(:))>0

    plot(centrosA(:,1),centrosA(:,2),'b*');

    viscircles(centrosA,radiosA,'Color','b');

    distR=sqrt((centrosA(:,1)-centrosB(:,1))^2+(centrosA(:,2)-centrosB(:,2))^2);

    %Calcula la relación entre la cola y el diámetro

    rela=(hlen+distR-radiosA)/(2*radiosA)

end

plot(centrosB(:,1),centrosB(:,2),'r*');

plot(xCentre ,yCentre, xcoords, ycoords );

hold off;

if rela<=0.5

    disp("Clase 0")

end

if rela>0.5 & rela<=1

    disp("Clase 1")

end

if rela>1 & rela<=2

    disp("Clase 2")

end
```

Procesamiento digital de imágenes citogenéticas para su clasificación

```
if rela>2 & rela<=3
```

```
    disp("Clase 3")
```

```
end
```

```
if rela>3
```

```
    disp("Clase 4")
```

```
end
```

```
%Muestra los resultados
```

```
hold on;
```

```
figure(4);
```

```
imshow(grayI);
```

5

Resultados y discusión:

5.1 Resultados

En esta tesis de maestría se substancian desarrollos científicos en el amplio campo del procesamiento de imágenes biológicas, en particular en el reconocimiento y clasificación de las imágenes obtenidas con microscopio de epifluorescencia y clasificadas a través de la técnica de Ensayo Cometa.

La mayoría de los procesos que involucran señales, en los que a la vez interviene material biológico, traen consigo una gran dificultad para manejar los datos (ruido, captura, no uniformidad de datos, variaciones del material biológico, entre otros). Otro aspecto complicado en el manejo de datos biológicos es la extracción de sus características, debido principalmente a las diferencias entre una clase y otra, e inclusive al analizar datos de una misma muestra. Una misma muestra puede tener diferentes enfoques de luz o incluso pueden aparecer varias células superpuestas. Por esto, la ecualización cobra relevancia, o dicho de otra manera, ofrece igualdad de condiciones al análisis de las muestras al ofrecer un histograma parejo. Para lograr esas características únicas en cada muestra se recurre a una serie de procesos, uso de funciones matemáticas, filtros y algoritmos que permiten esa ecualización.

Los algoritmos para la detección de parámetros de estas imágenes deben incorporar aspectos de generalización que conllevan pre-proceso para lograr la generalidad buscada. Este punto, en el caso de estudio pudo haber sido un desafío al tener varias especies diferentes para clasificar.

Así también clasificar las imágenes es una tarea complicada, si la detección de los parámetros que se han seleccionado como significativos no es adecuada.

Estos procesos son realizados con imágenes reales, de muestras reales que involucran algoritmos matemáticos y computacionales para la obtención de resultados. Los resultados obtenidos en los procesos aquí desarrollados predisponen a continuar con la investigación en esta rama de la Ingeniería Biomédica y también continuar por el camino de las ciencias de datos, posicionando a estos procesos como soporte para la toma de decisiones del especialista.

Procesamiento digital de imágenes citogenéticas para su clasificación.

Se han presentado diversas técnicas de procesamiento de señales desde el marco teórico. Explotando un enfoque adaptativo, es posible mejorar el desempeño de estos sistemas de procesamiento, en relación a las estrategias clásicas. A través de las técnicas abordadas en esta investigación, pueden tratarse exitosamente los problemas típicos del reconocimiento de objetos en imágenes, es decir, se proporciona invariancia a distorsiones geométricas de los objetos, ruido del fondo, ruido del sensor e iluminación no-homogénea en términos de su capacidad de discriminación y complejidad computacional.

Los tópicos desarrollados en el trabajo de investigación están relacionados entre sí, de manera que unos sirven de preproceso para otros, o bien post-proceso, como es el caso de la aplicación de algoritmos matemáticos para un ajuste de pesos de las redes neuronales.

La metodología de inteligencia computacional que más se adapta al caso es un híbrido neurodifuso, dado que, como se expuso anteriormente, las muestras biológicas no se adaptan exactamente a cada clase definida sino que a veces hay zonas borrosas en que hay que considerar los métodos difusos (por ejemplo, un umbral activo). De todas maneras se logra una clasificación adecuada con procedimientos funcionales sencillos, como se mencionó anteriormente. En el caso de estudio se logra clasificar con mucha precisión, porque se desarrolla el algoritmo en Matlab® para este caso específicamente.

En este caso, el desarrollo del prototipo funcional se realiza en Matlab® y se puede ejecutar en GNU Octave. La posibilidad de ejecución de los algoritmos en GNU Octave, es lo que permite avanzar con el prototipo.

Una de las propuestas a futuro incluye la realización de una aplicación que incorpore el procedimiento del prototipo a Smartphone equipado con filtros especiales para tratar la fluorescencia. Esta propuesta no disminuiría tecnología (ya que el laboratorio cuenta con microscopios especiales sobre el que se apoya la cámara para realizar los registros fotográficos de las muestras), sino que reemplazaría a la cámara digital incorporando nueva tecnología que además permitirá gestionar de manera más cómoda el almacenamiento, clasificación y conteo.

Procesamiento digital de imágenes citogenéticas para su clasificación.

Matlab® tiene un modulo que permite compilar el algoritmo para obtener una app que se puede correr en el móvil. Al ejecutar el software también se optimizan los tiempos de clasificación de las muestras.

Con esto se propone ampliar la gama de soluciones para diferentes casos posibles en el laboratorio y así encontrar, de una variedad de soluciones, el método que aporte una mejor performance en la clasificación y los recursos más convenientes económicamente a la hora de evaluar la implementación en el laboratorio de Citogenética (UNaM-IBS-CONICET). Para el laboratorio es un aporte genuino que le simplificaría mucho los tiempos de investigación/clasificación: como también sería un aporte educativo específicamente en la materia genética ecotoxicológica que corresponde al ciclo superior de orientación y flexibilización regional de la carrera Licenciatura en genética del nuevo plan de estudios UNaM 2017.

A lo largo del trabajo final de maestría, se presentan distintas herramientas para el proceso de imágenes digitales, en particular para las imágenes obtenidas con microscopio de fluorescencia a través de la técnica de Ensayo cometa.

Las técnicas de procesamiento digital usadas en este desarrollo, incorporan la innovación de su aplicación a las imágenes obtenidas.

También se presentaron como antecedentes de las líneas de investigación, el uso de algoritmos híbridos Neuro Fuzzy y Somn con algoritmos genéticos, incluyendo sus bases conceptuales cuidadosamente estudiadas y presentadas para ver su aplicación al caso. Los resultados preliminares fueron presentados en:

- WICC 2018 (Workshop de Ciencias de Computación desarrollado en la UNNE Ctes.).
Vera Laceiras M. Silvia y Caffetti Yanina; *“Reconocimiento de Patrones de imágenes digitales obtenidas mediante microscopio y parametrizadas según la técnica de Micronúcleos y la Técnica de Ensayo Cometa empleada por el Laboratorio de Citogenética General y Monitoreo Ambiental UNaM-IBS CONICET para la*

Procesamiento digital de imágenes citogenéticas para su clasificación.

detección de daños celulares.” 26 y 27 de Abril de 2018 Red UNCI - UNNE - ISBN 978-987-3619-27-4 Editorial Universidad Nacional del Nordeste (UNNE) 2018. Red de Universidades con Carreras en Informática (RedUNCI). <http://sedici.unlp.edu.ar/handle/10915/69205> [50].

- Y en el Primer Congreso Internacional de Ciencia, Diseño y Tecnología (UNAE 2019).

Vera Laceiras M. Silvia y Caffetti Yanina; “*Reconocimiento de Patrones de imágenes digitales obtenidas mediante microscopio y parametrizadas según la técnica de Micronúcleos y la Técnica de Ensayo Cometa empleada por el Laboratorio de Citogenética General y Monitoreo Ambiental UNaM-IBS CONICET para la detección de daños celulares.*” 1 y 2 de Octubre de 2019 Con Publicación en la SAETA Universitaria académica y científica del centro de documentación de la Universidad Autónoma de Encarnación ISSN: 2414 2506.

<https://www.unae.edu.py/tv/images/Memoria-del-I-Congreso-Internacional-de-ciencia-diseno-tecnologia-UNAE-2019.pdf> [51]

- WICC 2020 (Workshop de Ciencias de Computación desarrollado en forma virtual por la pandemia).

Vera Laceiras M. Silvia; “*Procesamiento digital de imágenes citogenéticas para su clasificación según la técnica ensayo cometa para la detección de daños en el ADN*”. Red UNCI - UNNE – artículo 12770 perteneciente al área “Agentes y Sistemas inteligentes”. <https://wicc2020.unpa.edu.ar/> [52].

Para realizar el preprocesamiento de las imágenes se recurrió al tratamiento de las imágenes cambiando a escalas de grises y luego binarización, segmentación y posterior almacenamiento en archivos independientes para ser analizados con Matlab® y Octave (Fig 38).

Procesamiento digital de imágenes citogenéticas para su clasificación.

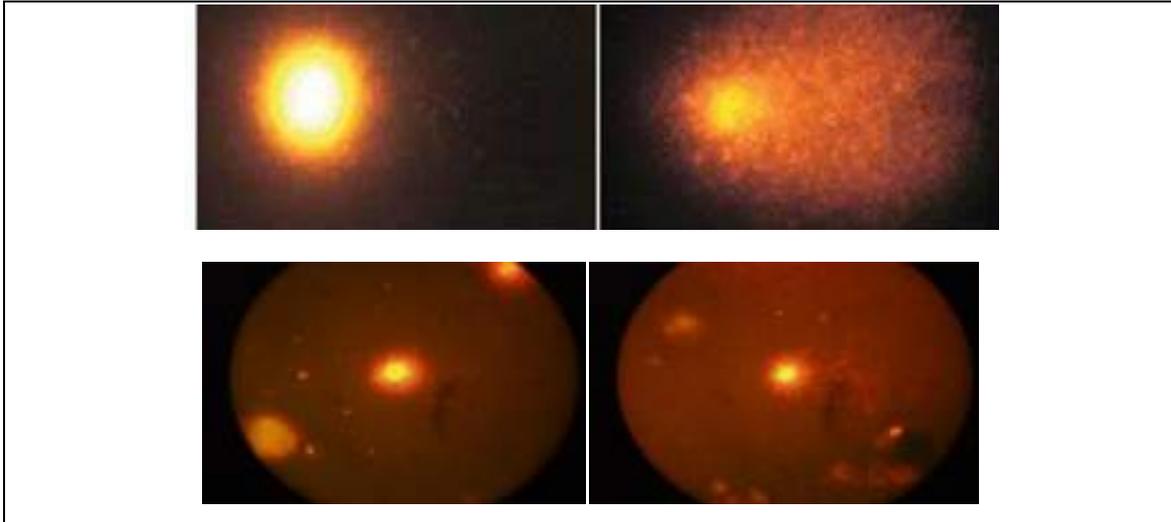


Figura 38: Imágenes de nucleoides de ADN imágenes logradas a través de las fotografías.

Luego se detecta y corrige suavemente la inclinación de algunas imágenes para que las luces y sombras generen un histograma ecualizado (Fig. 39).

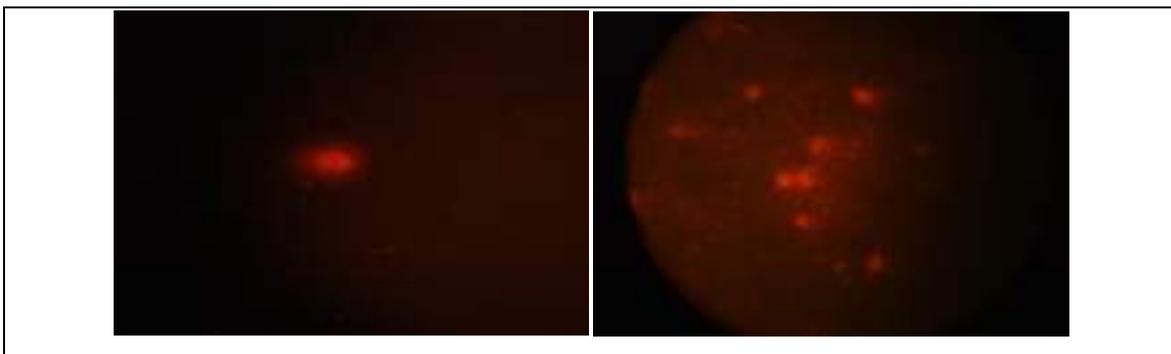


Figura 39: Imágenes de nucleoides de ADN con preproceso obtenemos imágenes ecualizadas.

Para ésto se utilizan diferentes técnicas como cambios de color a través de filtros y en escala de grises, determinación de centro del nucleoide, determinación de coordenadas y posición a través de funciones como seno, coseno, ismember, streal entre otras (Fig.40) y del armado del vector a través de algoritmos matemáticos y computacionales. Funciones que permiten tener imágenes adecuadas para realizar la detección y medición de los descriptores de las imágenes y su posterior clasificación en las diferentes Clases (Fig. 41).

Procesamiento digital de imágenes citogenéticas para su clasificación.

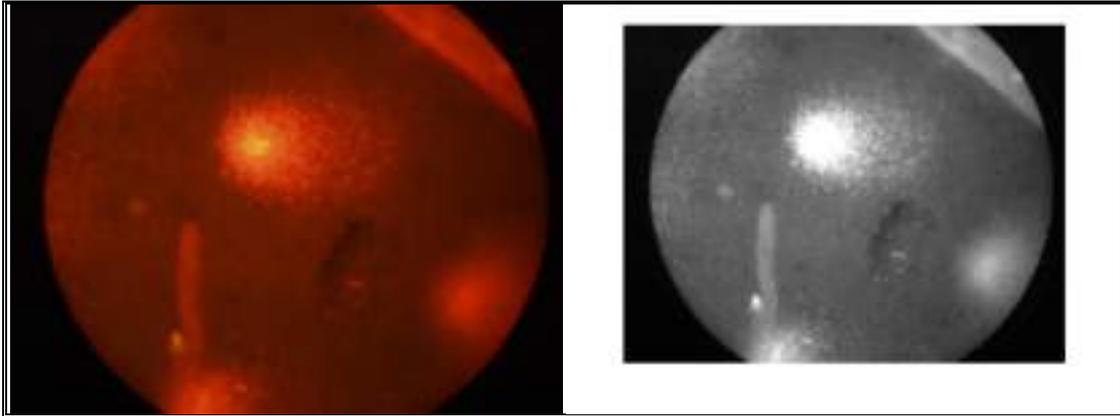


Figura 40: Se observan las imágenes de nucleoides de ADN antes y después de la transformación a escala de grises.

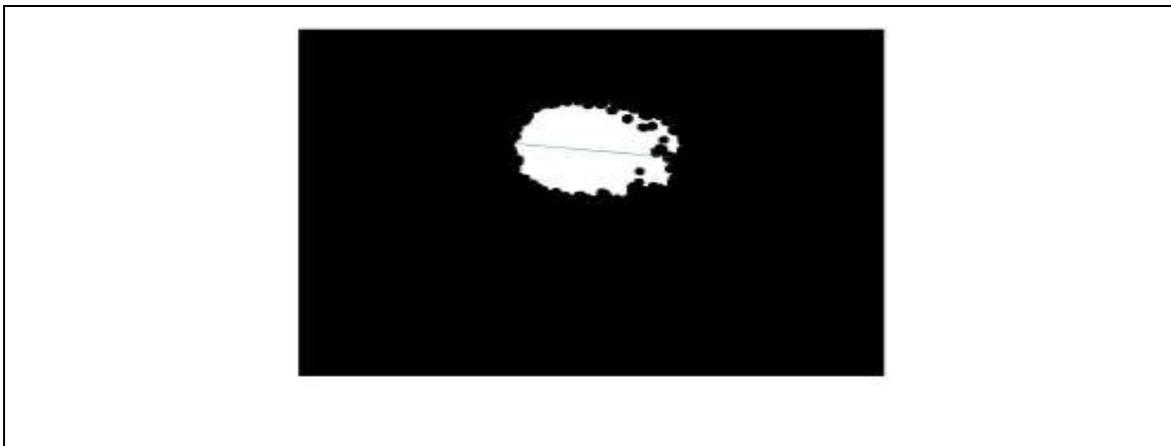


Figura 41: Imágenes de cabeza y cola del nucleoide en blanco y negro (binario).

El algoritmo desarrollado está basado en el procesamiento de imágenes y visión artificial, donde cobra una especial relevancia la segmentación, para medir y diferenciar las partes como proceso previo a la clasificación.

Procesamiento digital de imágenes citogenéticas para su clasificación.

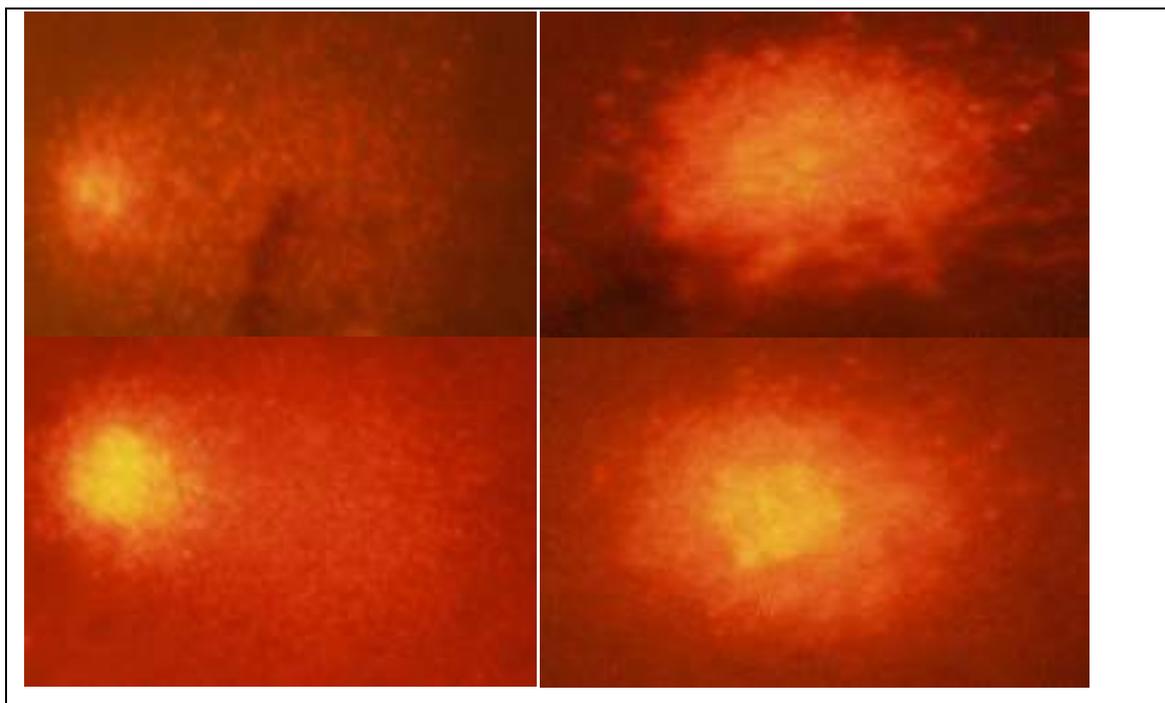


Figura 42 Imágenes del material biológico, puede verse muy diferente y sin embargo pertenecer a una misma clase.

Teniendo en cuenta que la imagen que se analiza es material biológico (el límite de pertenencia a las distintas clases no siempre es claro) (Fig.42), se opta por la adaptación de algoritmos neurodifusos, para que se pueda lograr la membresía a determinada clase a través de reglas que permitan flexibilidad, pues patrones muy específicos llevan a minimizar el error en esa especificidad pero aumentan el porcentaje de error en la generalidad. En el caso de estudio, al tener un set de datos con pocas imágenes es fácil caer en la especificidad. A la vez se debió encontrar el punto de defuzzificación que permite determinar hasta donde se considera la cabeza del cometa, donde cola y qué porcentaje de la imagen pertenece al fondo. Para eso se utilizan funciones como se ha visto que determinan las regiones y umbrales. El algoritmo clasificador (prototipo) desarrollado en MATLAB® y OCTAVE permite que ingrese una muestra y a través de funciones como **Imadjust** y otras que determinan las coordenadas del objeto o producen transformaciones, realizan la clasificación casi sin error, sin embargo no se puede decir que realmente hace un entrenamiento, sino que la fase de pruebas se dio más bien al determinar el umbral activo y allí se otorga la suficiente flexibilidad para luego generalizar.

Procesamiento digital de imágenes citogenéticas para su clasificación.

Se observa como ejemplo, una muestra que representa un anómalo (clase 4) (Fig. 43).

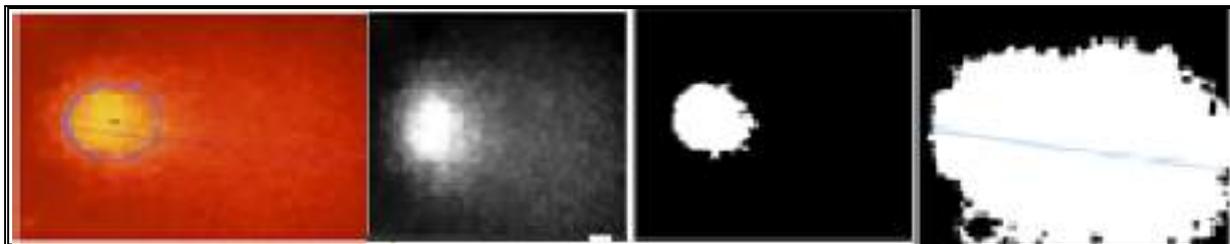


Figura 43 Clasificación de un registro fotográfico de una muestra perteneciente a un anómalo. Realiza el mismo procedimiento y lo clasifica como clase 4

Se diseña la red teniendo en cuenta estas reglas de aprendizaje difuso, de amplia difusión no sólo en desarrollos científicos sino también en la industria, pero dentro del modelo como se ve en el algoritmo, no se utilizan redes pre entrenadas del tipo Inception-v3,, Inception-v5 o Convnet en capas, sino que se desarrolla el algoritmo específicamente para el set de datos del caso de estudio.

Estos híbridos de computación inteligente permiten el manejo de decisiones en zonas de incertidumbre, donde la lingüística cobra interés y los límites no tienen bordes definidos.

Finalmente, en el procesamiento digital para la clasificación de patrones de imágenes, no solo se hizo inspección ocular sobre las muestras de las imágenes sino también se hallaron Índices cuantitativos que avalan esa evaluación visual. El estudio y la aplicación de las redes neuronales regidas por la heurística, definen la estructura de red y las entradas basadas en sus características morfológicas.

En base a lo expuesto, se demuestra en este trabajo final de maestría la aplicabilidad de estos algoritmos matemáticos y computacionales presentados a través de modelos o algoritmos adaptados para resolver casos puntuales como en la investigación de este caso de estudio, para tomar decisiones en clasificación de muestras microscópicas citogenéticas.

Procesamiento digital de imágenes citogenéticas para su clasificación.

El diseño de la herramienta prototipo con conceptos de inteligencia artificial, machine learning y su implementación a través de funciones simples, permiten trabajar con las imágenes de citogenética en Ensayo cometa.

Con respecto a lo que se estaba realizando (clasificación mediante inspección visual), esta variante permite obtener resultados efectivos de clasificación, sin tener que estar pendiente de la programación de los algoritmos cuando cambia la muestra y obtener estadísticas que se pueden ajustar y rectificar si es necesario.

A lo largo de este trabajo final obtuvimos los siguientes resultados:

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:

$$E_i = \frac{\text{numero de clasificaciones incorrectas}}{\text{total de datos del grupo test}} \quad (12)$$

El error final de la validación es:

$$\text{Error} = \frac{1}{K} \sum_{i=1}^k E_i \quad (13)$$

Se obtiene el siguiente error promedio: **Error 7.4%** conformado de la siguiente manera en cuanto a la precisión, Tabla (3):

Tabla 3 Cuadro de errores de precisión en cada clase.

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

Al medir la precisión de las clasificaciones y de acuerdo a las muestras presentadas, se obtuvieron los siguientes cuadros de precisión y exhaustividad (Tablas 4 y 5).

Procesamiento digital de imágenes citogenéticas para su clasificación.

Tabla 4 Cuadro de Precisión

Clase	Precisión
Clase 0	100,0%
Clase 1	91,8%
Clase 2	85,7%
Clase 3	42,0%
Clase 4	93,8%

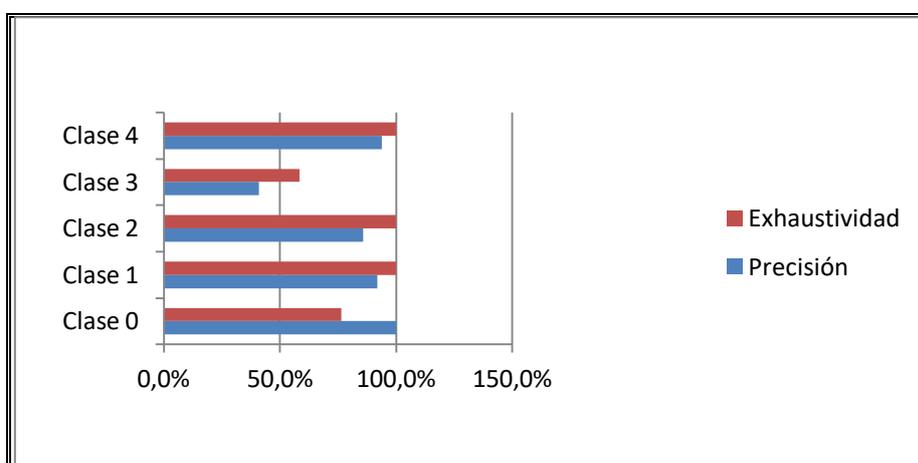


Figura 44: Gráfico de precisión por clases

Tabla 5 cuadro de exhaustividad

Clase	Exhaustividad
Clase 0	76,5%
Clase 1	100,0%
Clase 2	100,0%
Clase 3	58,3%
Clase 4	100,0%

Procesamiento digital de imágenes citogenéticas para su clasificación.

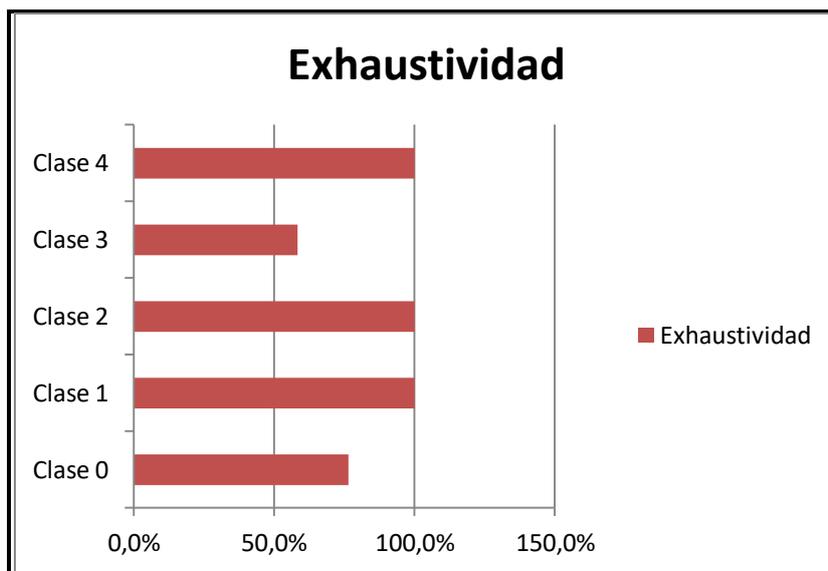


Figura 45: Grafico de exhaustividad por clases.

No obstante a través de la funcionalidad bien definida y un análisis adecuado de descriptores de características, junto con algoritmos matemáticos, se logran los resultados esperados, de clasificación con algoritmos de relativa sencillez (Fig. 44 y 45). La matriz de confusión queda representada de esta forma con los casos de éxito (Tabla 6: Fig. 46):

Tabla 6 Matriz de confusión

Matriz de confusión					
	Clase 0	Clase 1	Clase 2	Clase 3	Clase 4
Clase 0	13	0	0	0	0
Clase 1	4	45	0	0	0
Clase 2	0	0	18	0	3
Clase 3	0	0	0	30	2
Clase 4	0	0	0	0	7

Procesamiento digital de imágenes citogenéticas para su clasificación.

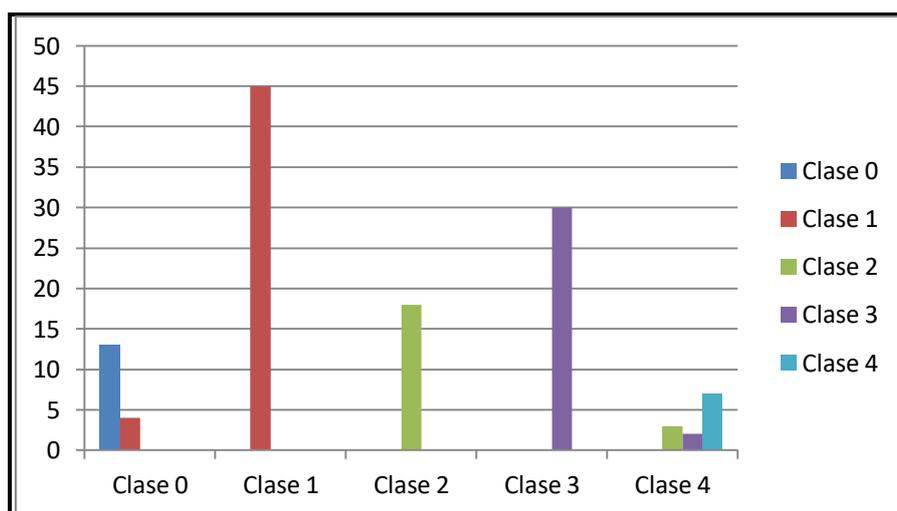


Figura 46: Grafico de matriz de confusión.

Para obtener el Accuracy que indica que tan bien entrenado está el modelo, se utilizó la siguiente ecuación:

$$\frac{TP+TN}{(TP+TN+FP+FN)} \quad (14)$$

Dando como resultado un Accuracy de 0.92 lo que indica que tanto en la prueba como en el entrenamiento el error va disminuyendo, indicador de que no hay sobreentrenamiento *overfitting* o sea en este caso el modelo fue bien entrenado y que se comportara de manera esperada con los nuevos datos.

Procesamiento digital de imágenes citogenéticas para su clasificación.

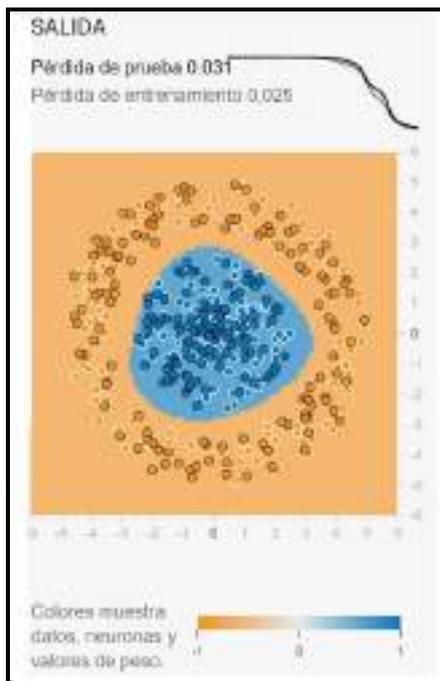


Figura 47: Se observa el comportamiento del error de prueba y del error de entrenamiento con curvas bastantes similares en las que el error decrece.

Como se observa en la figura 47 las curvas de entrenamiento y error decrecen. lo que indica que no hay overfitting.

También se verá cómo funciona el prototipo a través del índice de sensibilidad y precisión por clase.

$$\text{Recall } \frac{TP}{(TP+FN)} \tag{15}$$

$$\text{Precision } \frac{TP}{(TP+FP)} \tag{16}$$

Tabla 7: Recall y precisión

	Clase 0	Clase 1	Clase 2	Clase 3	Clase 4	Recall
Clase 0	13	0	0	0	0	1
Clase 1	4	45	0	0	0	0.91
Clase 2	0	0	18	0	3	0.86
Clase 3	0	0	0	30	2	0.93
Clase 4	0	0	0	0	7	1
Precision	0,76	1	1	1	0.58	0.9

El índice de precisión indica cuantos ítems reconocidos son relevantes y el Recall cuantos índices relevantes han sido seleccionados.

Procesamiento digital de imágenes citogenéticas para su clasificación.

Para encontrar un balance entre estas métricas se utiliza la media armónica o *F measure* (*F1 score*) cuya ecuación se aprecia a continuación:

$$F_{\text{measure}} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (17)$$

$$F_{\text{measure}} = 0,96$$

Como evaluación final se observa el Accuracy =0,92 y F measure =0,96.

Ambos indicadores describen un modelo equilibrado, que funciona y podrá reconocer nuevos datos clasificándolos correctamente.

Limitaciones:

En el procesamiento de imágenes se usaron modelos matemáticos implementados con algoritmos en Matlab®, y se ejecutaron con imágenes reales (preparadas a través de preprocesamiento, donde se elimina el ruido y se utiliza segmentación), que permiten identificar los objetos de estudio, medirlos y localizarlos para su clasificación (definiendo ciertas características visuales y haciendo uso de aprendizaje automático).

Se logra clasificar con especificidad a las especies presentadas y a la vez generalizar a otras especies como los bivalvos a través de umbrales activos, no provoca overfitting, pero tiene como limitación el escaso número de muestras, por lo que no se puede realizar el entrenamiento en modelos preentrenados de Deep Learning.

Tampoco se involucraron algoritmos genéticos pues no fue necesario para lograr una clasificación adecuada.

Discusión:

Se pudo avanzar con la presentación de una propuesta, que a través del uso de diferentes metodologías y la incorporación de la propiedad de transversalidad de las T.I.C. (Tecnología de la información y comunicación), facilita el desempeño de los recursos humanos en el laboratorio, logrando además automatizar procesos con precisión, bajo costo de equipamiento y breve tiempo de desarrollo. Teniendo un laboratorio equipado con microscopios

y filtros específicos, la incorporación de un Smartphone, no resta tecnología, sino que la suma, como accesorio que ofrece ventajas en la gestión, adquisición y almacenamiento de los registros fotográficos.

La aplicación del prototipo logra una autclasificación de nucleoides de células para las especies en estudio. Se han realizado pruebas con imágenes de diferentes especies, *Steindachnerina brevipinna* (sabalito), *Piaractus mesopotamicus* (pacú) y en moluscos pertenecientes a las especies *Corbícula fluminea* y *Limnoperna fortunei*.

El diseño de un algoritmo neurodifuso logra diferenciar la cabeza del cometa, la cola y el fondo, distinguiéndolos en el material biológico, aun en los casos en que la cola está difusa y se confunde con el fondo. Se puede analizar para medirlas y clasificarlas correctamente.

El diseño usando Inteligencia computacional nos permite estandarizar la entrada de datos a través de los algoritmos matemáticos que las ponderan, reducir la matriz de entrada y optimizar los procesos de entrenamiento, lo que acelera la convergencia del aprendizaje.

5.2 Propuesta para integrar equipamiento complementario

Algunos cambios o mejoras en el equipamiento del laboratorio beneficiarían y permitirían lograr una implementación exitosa. Actualmente, para analizar las células mediante el EC se usa un microscopio de epifluorescencia y para tomar las fotografías necesarias, se utiliza una cámara digital.

- Un Microscopio óptico marca Olympus CX31, donde se toman las microfotografías con una cámara Olympus C-5000 ZOOM, mediante un tubo adaptado en la parte superior del microscopio para encastrar la cámara.
- Un microscopio marca Motic BA310E, que funciona como microscopio óptico y también cuenta con los filtros necesarios para fluorescencia.

Como alternativa a las cámaras digitales descritas anteriormente, se proponen los Smartphone (Fig. 47) preparados para IA que puedan correr aplicaciones desarrolladas y compiladas en Matlab® por ejemplo.

Procesamiento digital de imágenes citogenéticas para su clasificación.

Este tipo de procesadores de Smartphone para IA permite que las API (interfaz de programación de aplicaciones) y los frameworks se puedan ejecutar en el mismo celular a través de sus NPU (unidad de procesamiento neuronal).

Huawei por ejemplo tiene arquitectura de cómputo móvil HIAI, una herramienta para que otros puedan desarrollar aplicaciones con su tecnología IA móvil para reconocimiento del lenguaje, análisis de audio, traducción contextual, interpretación de objetos mediante realidad aumentada, interpretación de fotos, clasificación y optimización, autenticación de usuarios mediante reconocimiento de patrones y sobre todo muy bajo costo en el mercado.

Según los vectores recolectados por los motores de IA se tendrán diferentes resultados. Como los desarrollos de IA requieren potencia para llevar adelante los cálculos, se habla de NPU, GPU (gráficos) y CPU.

La NPU se encarga de ajustar estos datos a realidades, de aplicar zoom digital o usar su sistema de detección de movimiento mejorado para corregir cualquier error. A este celular podemos incorporar algunas lentes específicas para microscopía que nos permitirán trabajar con mayor detalle si se necesita.

En definitiva, como se aprecia en la Fig. 47 la sugerencia es que el prototipo presentado se transforme en una aplicación para celular a través de la programación móvil con lo cual se simplifica el proceso de detección, guardado, clasificación y luego el conteo de células pertenecientes a cada clase. Con este desarrollo se cumpliría con la solicitud de mejora del laboratorio a través de la incorporación de Smartphones.

El celular reemplazaría en una primera instancia, a muy bajo costo, a la cámara que se usa actualmente adosando mejoras como guardado (lo que permite tener un histórico de muestras) e introducir mejoras como aplicar gestión de conocimiento, detección, clasificación y conteo a través de la ejecución del prototipo en sus núcleos preparados para tal fin. Esos núcleos o NPU lo que hacen es que las API y los Frameworks de las aplicaciones que se quieran correr lo puedan hacer en modo local sin necesitar procesadores externos como lo suelen hacer las aplicaciones que no tienen este tipo de tecnología. También queda abierta la posibilidad de anexar filtros específicos para trabajar

Procesamiento digital de imágenes citogenéticas para su clasificación.

con microscopías. Esas lentes solamente se anexan al Smartphone y se puede obtener un rendimiento y calidad de fotografía muy buenos. No solo para el uso en el laboratorio sino también en la cátedra de citogenética general, representando una verdadera innovación en la carrera.

El hecho de contar con este tipo de dispositivo, reduce notablemente los costos de incorporación de tecnología al laboratorio y aumenta las prestaciones. Con la implementación de este tipo de Smartphone se pueden desarrollar próximamente aplicaciones que ejecuten el prototipo aquí presentado, para clasificación, como se observa en el esquema presentado en la Fig. 48.

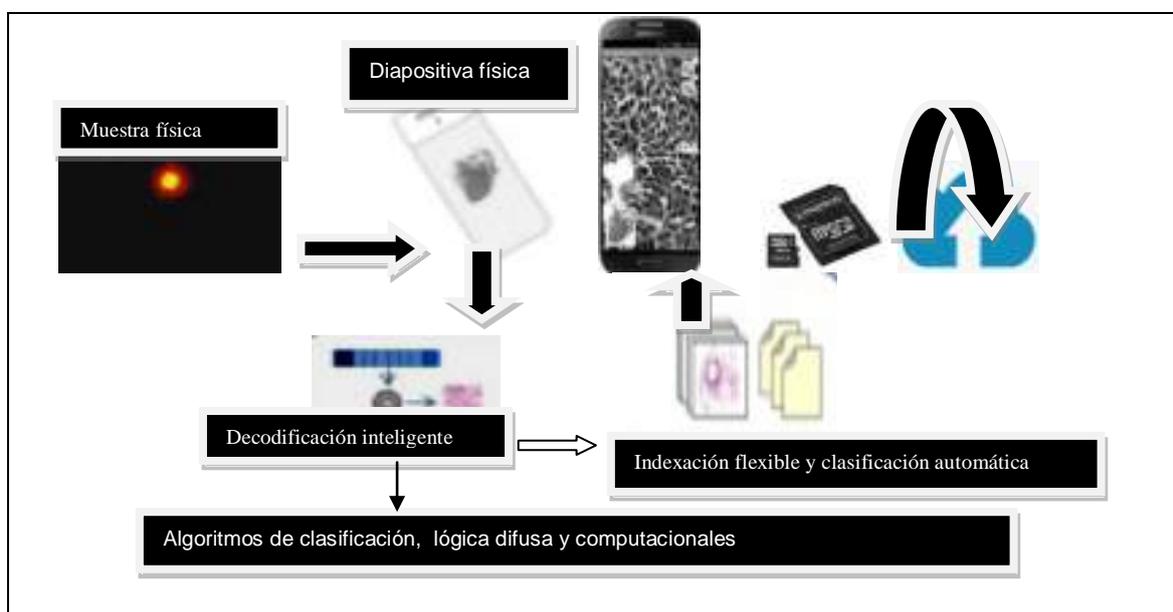


Figura 48: Ejemplo de desarrollo de la aplicación para el Móvil desde la muestra física usando el prototipo de de clasificación y llegando a la clasificación inteligente de las muestras. Fuente propia.

6

Conclusiones

Procesamiento digital de imágenes citogenéticas para su clasificación.

El desarrollo y la incorporación de las nuevas tecnologías a estudios citogenéticos permiten otra apertura al mundo científico y técnico, sobre todo en lo que a reconocimiento de patrones de imágenes y visión por computadora se refiere.

Esta tesis de maestría no solo aporta soluciones reales al laboratorio de citogenética en su línea de investigación sino también en el ámbito académico donde se convertirá en una herramienta muy útil en las cátedras que incluyen el aprendizaje de la técnica de ensayo cometa.

Además en el campo profesional aportan a la formación de recursos humanos especializados en técnicas de inteligencia artificial y en su aplicación en el procesamiento de imágenes.

Con respecto a los objetivos propuestos, se obtuvo la autoclasificación de las imágenes obtenidas en las 5 clases usando redes neuronales, algoritmos neurodifusos y funciones que permiten realizar automáticamente esta clasificación con muy poco margen de error.

También se logró el desarrollo del procedimiento prototípico que realiza la clasificación de las imágenes, a través del reconocimiento de patrones de imágenes citogenéticas obtenidas por la fotografía. Se realizó la normalización o pre-proceso y se generó además, como valor agregado, la creación del banco de datos con los archivos de las imágenes obtenidas de las muestras al aplicar el ensayo cometa.

La validación del prototipo se alcanzó a través del caso de estudio, con las imágenes obtenidas en el Laboratorio de Citogenética General y Monitoreo Ambiental del Instituto de Biología Subtropical.

Como aportes genuinos de este trabajo, enfatizamos la reducción en los tiempos del especialista experto, que mejora su eficiencia. El registro digital de las imágenes también brinda la posibilidad de generar una base de datos estándar de tal manera que permite pensar en futuros estudios en los que se analicen los datos históricos. En comparación con algunos software desarrollados como "Komet 7" o "Capslab" (mencionados con anterioridad), al ser éste un software desarrollado específicamente con el método de

Procesamiento digital de imágenes citogenéticas para su clasificación.

observación visual, mejora los tiempos y es sumamente útil en el laboratorio por su simpleza. De todas formas, se debe realizar la toma de la muestra fotográfica como inicio del proceso, es más simple pues no necesita ubicarse en las coordenadas preparadas para cada toma, que exigen softwares disponibles en el mercado como Komet7 y su clasificación responde a las métricas utilizadas en el laboratorio. Se podrían lograr proyecciones y otros tipos de diagnósticos a fin de prevenir enfermedades endémicas, maximizando la funcionalidad de las señales de alerta temprana. Además, a través de herramientas que permitan tener un panorama general de manejo de variables como indicadores claves por ejemplo, en cuadros de mandos, o bien a través de gestión del conocimiento o ciencia de datos.

También, aporta un perfil educativo y gratuito del desarrollo producido en esta investigación, que contribuye al desarrollo y preparación de alumnos que aprenden a clasificar las células de las distintas especies en ensayo cometa en la carrera de Genética de la UNaM, siendo la primera facultad en incorporar esta herramienta a la formación.

Luego del estudio de diferentes metodologías y de realizar las evaluaciones pertinentes, primero de característica empírica y luego llevando a la práctica varios algoritmos, se concluye que las redes neuronales y los algoritmos neuro fuzzy son una excelente opción en el manejo de imágenes citogenéticas. Con esta técnica se logra una optimización efectiva en la clasificación de diferentes tipos de células. Siendo las características biológicas y sus diferencias naturales, uno de los problemas que generaba limitaciones importantes, dado que muchos modelos solo permitían clasificar *ADN* de una sola especie. En este estudio se preparó al modelo para clasificar el *ADN* de diferentes especies de peces y bivalvos con muy buenos resultados.

En la actualidad existe una amplia variedad de arquitecturas de redes Neuronales convolucionales disponibles en forma gratuita y sin restricciones de uso, que pueden alcanzar un rendimiento razonable en tareas de reconocimiento visual. Sumado a esto, las grandes empresas como Google® o Microsoft® que reconocen la importancia de la ciencia de datos en nuestras

Procesamiento digital de imágenes citogenéticas para su clasificación.

actividades cotidianas ofrecen hardware en Cloud que nos permite usar GPUS o TPUS de manera gratuita sin ningún otro requisito para seguir investigando.

De la misma manera, se tienen a disposición extensas librerías de alto nivel como KERAS open source, que sirve como capa de abstracción sobre Tensor Flow para desarrollos de todo tipo. Python, numpy, numba etc. permiten entrenar modelos con relativa sencillez. También Google® pone a disposición repositorios de imágenes y datos de todo tipo para el entrenamiento y validación de algoritmos de clasificación y procesamiento.

Este prototipo y la experiencia que trae consigo el desarrollo de diferentes metodologías, algoritmos, fusiones y aplicaciones en un hardware novedoso con improntas creadas sin limitaciones, permite ofrecer en la región ciencia de datos, inteligencia artificial y Visión por computadora.

Nuestra región se encuentra en este momento con pocos exponentes en estas ramas y sobre todo en cuanto al reconocimiento de patrones de imágenes se trata. De ahí la importancia y el aporte de esta tesis tanto en el área de la BIOTECNOLOGIA como así también en GENETICA.

Estos primeros pasos alientan para continuar desarrollando herramientas que permitan destacar, la formación de recursos humanos preparados para la innovación tecnológica, desde la Maestría en TIC de la UNaM. También pone a disposición de la sociedad el conocimiento o (*know how*) para nuevos desarrollos en diferentes campos, Industrial, Genético, Biotecnológico, Domótica y para las aplicaciones de los proyectos de “Smart City” que propone la provincia en el desarrollo e implementación de las “Economías del conocimiento”.

Como valor agregado, con la difusión de estos proyectos a través de Workshop de Ciencias y Congresos, se establecen vínculos con otras Universidades que tienen desarrollos en el área de ciencia de Datos o Inteligencia Artificial y están interesados en trabajar de manera colaborativa expandiendo nuestros límites físicos de Misiones y el NEA para traspasar fronteras e integrar conocimientos.

Bibliografía:

- [1] Flavia Antonela Leveroni, Jacqueline Diana Caffetti, María Cristina Pastori, «*Genotoxic response of blood, gill and liver cells of Piaractus mesopotamicus after an acute exposure to a glyphosate-based herbicide*» *Caryologia*, 70:1, 21-28, DOI: 10.1080/00087114.2016.1254454
- [2] S. Wang, D. M. Yang, R. Rong, X. Zhan, y G. Xiao, «*Pathology Image Analysis Using Segmentation Deep Learning Algorithms*», *Am. J. Pathol.*, vol. 189, n.º 9, pp. 1686-1698, sep. 2019.
- [3] A. Krizhevsky, I. Sutskever, y G. E. Hinton, «*ImageNet classification with deep convolutional neural networks*», *Commun. ACM*, vol. 60, n.º 6, pp. 84-90, 2017.
- [4] H. Takagi y I. Hayashi, «*NN-driven fuzzy reasoning*», *Int. J. Approx. Reason.*, vol. 5, n.º 3, pp. 191-212, 1991.
- [5] 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.
- [6] D. Sebastião Comas, G.J. Meschino, «*Segmentación de Imágenes mediante Reconocimiento de Patrones*», *Esc. y Work. Argentino en Ciencias las Imagenes*, 2014.
- [7] A. Bookstein, «*Fuzzy Requests: An Approach to Weighted Boolean Searches*», *J. Am. Soc. Inform. Sci.*, vol. 31, pp. 240-247, 1980.
- [8] Wang Tiesheng y Rao Zhenxing, «*Fuzzy neural network algorithm and its application*», en *2011 International Conference on Electric Technology and Civil Engineering (ICETCE)*, 2011, pp. 6001-6004.
- [9] Xing, Fuyong, Xie, Yuanpu Su, Hai Liu, Fujun Yang, Lin «*Deep Learning in Microscopy Image Analysis: A Survey*», *IEEE Transactions on Neural Networks and Learning Systems*. vol 29 1109/TNNLS.2017.2766168
- [10] M.-Y. Cheng, H.-C. Tsai, C.-H. Ko, y W.-T. Chang, «*Evolutionary Fuzzy Neural Inference System for Decision Making in Geotechnical Engineering*», *J. Comput. Civ. Eng.*, vol. 22, n.º 4, pp. 272-280, jul. 2008.
- [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. S. Landau y L. Pantanowitz, «*Artificial intelligence in cytopathology: a review of the literature and overview of commercial landscape*», *J. Am. Soc. Cytopathol.*, vol. 8, n.º 4, pp. 230-241, 2019.
- [13] C. Giraldo_Omaira, «*La gestión del conocimiento en las organizaciones y las regiones : una revisión de la literatura the Knowledge Management in the organizatios and the regions* », n.º 1, 2018.
- [14] The MathWorks «*Introducing deep learning with matlab*» Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. 2018.
- [15] Colin Walker , «*Ecotoxicology Effects of Pollutants on the Natural Environment* » edit Taylor & Francis Group, an Informa business No claim to original U.S. Government works Version Date: 20140513 International Standard Book Number-13: 978-1-4822-4700-8 (eBook - PDF)

- [16] R. Van der Oost, J. Beyer, y N. P. E. Vermeulen, «*Fish bioaccumulation and biomarkers in environmental risk assessment: A review*», *Environmental Toxicology and Pharmacology*, vol. 13, n.º 2. pp. 57-149, feb-2003.
- [17] O. Ostling y K. J. Johanson, «*Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells*», *Biochem. Biophys. Res. Commun.*, vol. 123, n.º 1, pp. 291-298, ago. 1984.
- [18] N. Singh, MT. McCoy, R. Tice EL Schneider «*A simple technique for quantitation of low levels of DNA damage in individual cells*». *Exp Cell Res.* 1988;175(1):184-191. doi:10.1016/0014-4827(88)90265-0
- [19] J. D. Caffetti, M S. Mantovani, M.C. Pastori and A S. Fenocchio «*First genotoxicity study of Paraná river water from Argentina using cells from the clam Corbicula fluminea (Veneroida Corbiculidae) and Chinesehamster (Cricetulus griseus Rodentia, Cricetidae) K1 cells in the comet assay*» *Genetics and Molecular Biology*, 31, 2, 561-565 (2008) *Sociedade Brasileira de Genética, Brazil.*
- [20] M. MN Authman, MS Zaki, EA Khallaf, H. Abbas «*Use of Fish as Bio-indicator of the Effects of Heavy Metals Pollution*». *Hydrobiology Department, National Research Centre, 33 EL Bohouth St. (Former EL Tahrir St.), Dokki, Giza, Egypt, Faculty of Science, Minufiya University, Shebeen Alkoom, Egypt Authman et al., J Aquac Res Development 2015, 6:4 DOI: 10.4172/2155-9546.1000328*
- [21] A. P. Boyle, J. Guinney, G. E. Crawford, y T. S. Furey, «*F-Seq: A feature density estimator for high-throughput sequence tags*», *Bioinformatics*, vol. 24, n.º 21, pp. 2537-2538, nov. 2008.
- [22] A. M.Gonzalez, J. L.Gonzalez C. B. Loeza, A. Espinosa, Pacheco-Pantoja «*Introducción a Técnicas de Segmentación de Células del Ensayo Cometa*» *Ingeniería Revista Académica de la Facultad de Ingeniería Universidad Autónoma de Yucatán*.vo 22 Num 3(2018) .ISSN: 2448-8364.
- [23] F. Neri y V. Tirronen, «*Scale factor local search in differential evolution*», *Memetic Comput.* jun. 2009, vol. 1, n.º 2, pp. 153-171.
- [24] 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.
- [25] LeCun, L Bottou, Y Bengio, y P Haffner. «*Aprendizaje basado en gradientes aplicado al reconocimiento de documentos*». *Actas del IEEE*, (1998). 86 (11), 2278-2324.
- [26] Krizhevsky, A., Sutskever, I. y Hinton, GE «*Clasificación de Imagenet con redes neuronales convolucionales profundas*». En *Avances en sistemas de procesamiento de información neuronal*(2012). (pp. 1097-1105).
- [27] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. «*Profundizando en convoluciones*». En *Actas de la Conferencia IEEE (2015) sobre Visión por Computadora y Reconocimiento de Patrones* (pp. 1-9).
- [28] K. C. Chang, Y. W. Chiang, C. H. Yang, y J. W. Liou, «*Atomic force microscopy in biology and biomedicine*», *Tzu Chi Medical Journal*. Buddhist Compassion Relief Tzu Chi Foundation. Published by Elsevier Taiwan LLC. All rights reserved. July 2012

- [29] S. Robertson, H. Azizpour, K. Smith, y J. Hartman, «*Digital image analysis in breast pathology—from image processing techniques to artificial intelligence*», *Transl. Res.*, vol. 194, pp. 19-35, 2018.
- [30] 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.
- [31] J. C. F. Silva, R. M. Teixeira, F. F. Silva, S. H. Brommonschenkel, y E. P. B. Fontes, «*Machine learning approaches and their current application in plant molecular biology: A systematic review*», *Plant Sci.*, vol. 284, pp. 37-47, 2019.
- [32] M. K. K. Niazi, A. V Parwani, y M. N. Gurcan, «*Digital pathology and artificial intelligence*», *Lancet Oncol.*, vol. 20, n.º 5, pp. e253-e261, may 2019.
- [33] V. M. Garcia Luna, «*Introducción al procesamiento digital de señales*», *Elai-Upm*, vol. 1, p. 5, 2001.
- [34] M. Mart, «*Técnicas Clásicas de Segmentación de Imagen*», *Ecología*, n.º 1, pp. 1-23, 2004.
- [35] K. J. Lee y H. P. Soyer, «*Smartphones, artificial intelligence and digital histopathology take on basal cell carcinoma diagnosis*», *Br. J. Dermatol.*, p. bjd.18374, ago. 2019.
- [36] M.Gamarra, F.Bertel, J.Velasquez «*Herramienta de Software para el Aprendizaje de Sistemas Difusos en un Curso de Control Digital*». Form. Univ. La Serena 2016, vol.9, n.4 pp.33-40. ISSN 0718-5006.
- [37] H. Takagi, N. Suzuki, T. Koda, y Y. Kojima, «*Neural networks designed on approximate reasoning architecture and their applications*», *IEEE Trans. Neural Networks*, vol. 3, n.º 5, pp. 752-760, 1992.
- [38] F. Yin, J.Jun L. J. Wang, C. Guo, Dalian li gong da xue., Chinese University of Hong Kong., y IEEE Circuits and Systems Society., «*Advances in neural networks*» ISSN 2004 : International Symposium on Neural Networks, Dalian, China, August 19-21, 2004 : proceedings. Springer, 2004.
- [39] Horzyk A., Tadeusiewicz R. «*Self-Optimizing Neural Networks*» (Redes neuronales auto-optimizadas). En: Yin FL., Wang J., Guo C. (eds) *Avances en redes neuronales - ISBN 978-3-540-22841-7 Lecture Notes in Computer Science*, vol 3173. Springer, Berlín, HeidelbergA. icz, , pp. 150-155.
- [40] Rodriguez,AC «¿Reticularismo o neuronismo?: Diferente percepción de la misma circunstancia» Archivos de neurociencia(MexicoDF).vol 10 n1 (version impresa ISSN0187-4705 2005].
- [41] N. Kalchbrenner, E. Grefenstette, y P. Blunsom, «A convolutional neural network for modelling sentences», en *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, 2014, vol. 1, pp. 655-665.
- [42] S. Apu y S. Apu, «Hypertrophy», en *ECG for Medical Diagnosis*, vol. 10, n.º 1, Jaypee Brothers Medical Publishers (P) Ltd., 2018, pp. 63-63.
- [43] Thamer M.Jamel, B.M.K., «Implementation of a sigmoid activation function for neural network using FPGA». 13th Scientific I Published in the 13th Scientific Conference of Al-Ma'moon University College -18 April 2012 – Baghdad , Iraq .
- [44] Ramírez Q. Juan A. y Chacón M. Mario I. «Redes neuronales artificiales

- para el procesamiento de imágenes, una revisión de la última década » RIEE&C, Revista de ingeniería electrónica y computación, Vol. 9 No. 1, Julio 2011. ISSN 1870 - 9532
- [45] Du, Zhenlong, Li, Xiaoli «Laplacian filtering effect on digital image tuning via the decomposed eigen-filter» Computers and Electrical Engineering vol 78 (version impresa) ISSN 00457906.
- [46] Rojas, T., Sanz, W., y Arteaga F.(2008). «Sistema de visión por computadora para la detección de objetos esféricos a través de la transformada de Hough». Revista Ingeniería UC, 15 (1), 77-87.
- [47] Perez-Teruel, Karina; Leyva-Vazquez, Maikel; Espinilla, Macarena y Estrada-Senti, Vivian. « Computación con palabras en la toma de decisiones mediante mapas cognitivos difusos». Rev cuba cienc informat 2014, vol.8, n.2, pp.19-34. ISSN 2227-1899.
- [48] Sutton , Richard S, Barto, Andrew G «Reinforcement Learning»(Machine learning.Series Q325.6.S88 1998.Edicion Kindle.ISBN 0-262-19398-1.
- [49] Goodfellow, Ian, Bengio, Yoshua y Courville, Aaron. (2016, diciembre). «Aprendizaje profundo» MIT Press ISBN-13: 978-0262035613
- [50] Vera Laceiras M.S, Caffetti Y. “Reconocimiento de Patrones de imágenes digitales obtenidas mediante microscopio y parametrizadas según la técnica de Micronúcleos y la Técnica de Ensayo Cometa empleada por el Laboratorio de Citogenética General y Monitoreo Ambiental UNaM-IBS CONICET para la detección de daños celulares.” Ctes. 2018 Red UNCI - UNNE - ISBN 978-987-3619-27-4 Editorial Universidad Nacional del Nordeste (UNNE) 2018. Red de Universidades con Carreras en Informática (RedUNCI). <http://sedici.unlp.edu.ar/handle/10915/692053>
- [51] Vera Laceiras M. Silvia y Caffetti Yanina; “Reconocimiento de Patrones de imágenes digitales obtenidas mediante microscopio y parametrizadas según la técnica de Micronúcleos y la Técnica de Ensayo Cometa empleada por el Laboratorio de Citogenética General y Monitoreo Ambiental UNaM-IBS CONICET para la detección de daños celulares.” Encarnación-Paraguay 2019 SAETA Universitaria académica y científica del centro de documentación de la Univ. Autónoma de EncarnaciónISSN: 2414 2506. <https://www.unae.edu.py/tv/images/Memoria-del-I-Congreso-Internacional-de-ciencia-diseno-tecnologia-UNAE-2019.pdf3>
- [52] Vera Laceiras M. Silvia; “Procesamiento digital de imágenes citogenéticas para su clasificación según la técnica ensayo cometa para la detección de daños en el ADN”. Calafate 2020 (Red UNCI) - artículo 12770 perteneciente al área “Agentes y Sistemas inteligentes”. <https://wicc2020.unpa.edu.ar/> WICC 2020 (Workshop de Ciencias de Computación desarrollado en forma virtual por la pandemia)