

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA
BASADO EN REDES NEURONALES
ARTIFICIALES PARA LA CARACTERIZACIÓN DE
CÉLULAS MALIGNAS EN FLUIDO PLEURAL

JUAN FRANCISCO REMOLINA CAVIEDES
ADRIÁN FERNANDO RODRÍGUEZ VEGA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA
BUCARAMANGA
2007

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA
BASADO EN REDES NEURONALES ARTIFICIALES
PARA LA CARACTERIZACIÓN DE CÉLULAS MALIGNAS
EN FLUIDO PLEURAL

JUAN FRANCISCO REMOLINA CAVIEDES
ADRIÁN FERNANDO RODRÍGUEZ VEGA

Trabajo de Grado para optar el título de Ingeniero de Sistemas

Directora:
OLGA MERCEDES ÁLVAREZ OJEDA
MD. Especialista en Patología

Codirector:
VÍCTOR EDUARDO MARTÍNEZ ABAUNZA
Magíster en Ingeniería: Área Informática y Ciencias de la
Computación

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA
BUCARAMANGA
2007

A las víctimas del conflicto colombiano.
A ésta patria que sabe a dolor.
A quien hoy no está conmigo
pero quien siempre anheló
patear estrellas en la tarde
y tomar café juntos.
A Yanira, por su disciplinada
capacidad de soportarme
aunque muchas veces
sea terca.
A Edwin, no solo por haber
compartido las mismas
entrañas
sino por hacer
de la distancia
un constante
abrazo fraterno.
A mis amig@s,
por ser bien
poc@s.
A mis compañer@s,
por ser bastantes.

Juan F.

A mi madre, porque éste logro
es sólo suyo.

A mi familia, por el apoyo
brindado desde que inicié
éste sueño.

A mis amigos, los que siempre han
estado conmigo.

Adrián F.

AGRADECIMIENTOS

A la Dra Olga, por su valioso acompañamiento.

A Víctor, por su paciencia, compañerismo y dedicación.

CONTENIDO

1. INTRODUCCIÓN	24
2. PRESENTACIÓN DEL PROYECTO	26
2.1. OBJETIVOS	26
2.1.1. OBJETIVO GENERAL	26
2.1.2. OBJETIVOS ESPECÍFICOS	26
2.2. JUSTIFICACIÓN	26
2.2.1. DEFINICIÓN DEL PROBLEMA	26
2.2.2. IMPACTO ESPERADO	28
2.2.3. VIABILIDAD	28
3. MARCO TEÓRICO	29
3.1. TRATAMIENTO DIGITAL DE IMÁGENES	29
3.1.1. MODELOS DE COLOR	30
3.1.2. MODELO RGB	30
3.1.3. MODELO DE COLOR HSL	31
3.2. ETAPAS DEL TRATAMIENTO DIGITAL DE IMÁGENES	33
3.2.1. ADQUISICIÓN	34
3.2.2. PREPROCESAMIENTO	35
3.2.3. SEGMENTACIÓN	39
3.2.4. DESCRIPCIÓN	45
3.2.5. INTERPRETACIÓN Y RECONOCIMIENTO	48
3.3. MORFOLOGÍA MATEMÁTICA	54
3.3.1. TRANSFORMACIONES MORFOLÓGICAS	54
3.4. CITOPATOLOGÍA DE LA PLEURA	60
3.4.1. GENERALIDADES	60
3.4.2. ANATOMOFISIOLOGÍA DE LA PLEURA	61
3.4.3. CITOPATOLOGÍA DE LOS DERRAMES PLEURA- LES	64
3.4.4. CITOPATOLOGÍA DE LOS DERRAMES PLEURA- LES CON CÉLULAS MALIGNAS	74
4. ADQUISICIÓN	80
4.1. RELACIÓN PÍXEL-MICRA	82

5. PREPROCESAMIENTO	84
5.1. FILTRADO	85
5.1.1. FILTRO CIRCULAR	85
5.1.2. FILTRO GAUSS	85
5.1.3. FILTRO ROTACIONAL	86
5.2. AJUSTE DE IMAGEN	86
5.3. CONVERSIÓN AL MODELO IHSL	87
6. SEGMENTACIÓN	89
6.1. CRECIMIENTO DE REGIONES	89
6.2. DISTANCIA VECTORIAL	91
6.2.1. PASOS PARA LA SEGMENTACIÓN POR DISTAN- CIA VECTORIAL	94
6.2.2. RESULTADOS OBTENIDOS CON SEGMENTACIÓN POR DISTANCIA VECTORIAL	95
7. DESCRIPCIÓN	97
7.1. DESCRIPCIÓN DE CITOPLASMA	97
7.1.1. MEDIDAS MORFOLÓGICAS	97
7.2. DESCRIPCIÓN DEL NÚCLEO	99
7.2.1. MEDIDAS MORFOLÓGICAS	99
7.2.2. MEDIDAS DE TEXTURA	100
7.3. MODELO ENTIDAD-RELACIÓN	101
8. INTERPRETACIÓN Y RECONOCIMIENTO	103
8.1. CARACTERÍSTICAS CELULARES	103
8.1.1. GRÁFICAS DE DISPERSIÓN	104
8.1.2. INTERPRETACIÓN DE CARACTERÍSTICAS CE- LULARES SEGÚN GRÁFICOS DE DISPERSIÓN . . .	105
8.2. RED NEURONAL ARTIFICIAL DE CLASIFICACIÓN . . .	106
8.2.1. ARQUITECTURA Y ALGORITMO DE ENTRENA- MIENTO	106
9. MODELO COMPUTACIONAL	108
10.RESULTADOS	111
10.1. VALIDEZ DE LA PRUEBA DIAGNÓSTICA	111
10.1.1. SENSIBILIDAD	111

10.1.2. ESPECIFICIDAD	111
10.2. SEGURIDAD DE LA PRUEBA DIAGNÓSTICA	112
10.2.1. VALOR PREDICTIVO POSITIVO	112
10.2.2. VALOR PREDICTIVO NEGATIVO	112
10.3. INFLUENCIA DE LA PREVALENCIA	113
10.4. RAZONES DE PROBABILIDAD	113
10.4.1. COCIENTE DE PROBABILIDADES POSITIVO . . .	113
10.4.2. COCIENTE DE PROBABILIDADES NEGATIVO . .	114
10.5. RESULTADOS Y VALIDACIÓN DEL MODELO	114
11. CONCLUSIONES Y RECOMENDACIONES	118
BIBLIOGRAFÍA	119
ANEXOS	121
A. FUNCIONAMIENTO DEL MODELO COMPUTACIONAL	121

GLOSARIO

Acino: que tiene forma o estructura semejante a un racimo de uvas.

Acúmulos: aumento de la cantidad de una sustancia que está contenida en una víscera.

Adenocarcinoma: forma de carcinoma que se origina en el tejido glandular.

Adquisición: etapa del procesamiento digital de imágenes que permite obtener una imagen digital de un objeto, a partir de una escena.

Albúmina: es la proteína plasmática más abundante (60%), utilizada para el transporte de proteínas, ácidos grasos, hormonas, fármacos, etc., se sintetiza en el hígado.

Anfófilo: que puede teñirse con colores ácidos o básicos.

Anisonucleosis: variación en el radio nuclear, núcleos deformes.

Ascitis: acumulación de líquido ascítico en la cavidad peritoneal. se produce cuando hay una inflamación del peritoneo o bien cuando aumenta la presión en el sistema de la vena porta.

Autovalor: factor de escala por el que ha sido multiplicado un autovector.

Autovectores: vectores no nulos que, cuando son transformados por el operador, dan lugar a un múltiplo escalar de sí mismos, con lo que no cambian su dirección.

Binarización: procedimiento para convertir una imagen que se encuentra en el formato de color verdadero, o en escala de grises, a un formato de dos colores (blanco y negro).

Carcinoma escamoso: término descriptivo para tumor, cuyas células tienen apariencia escamosa.

Carcinoma: cualquier tipo de cáncer que se forma en las células epiteliales, maligno por definición.

Celulas mesoteliales: componentes benignos de citología pleural.

Cianófilo: de color azul o verde.

Citología de líquido pleural: el examen de fluido pleural es una prueba que se realiza para detectar la presencia de células cancerosas, véase toracentesis.

Citopatología: rama de la medicina que se dedica al estudio de los cambios celulares. Coagulopatía: enfermedad que consiste en un trastorno del sistema de la coagulación que funciona deficientemente.

Coagulopatía: enfermedad que consiste en un trastorno del sistema de la coagulación que funciona deficientemente.

Croma: también llamado tono o matiz, es el color en sí mismo, atributo que permite diferenciar a un color de otro.

Cromatina: sustancia contenida en unos filamentos que son susceptibles de teñirse y que está esparcida por todo el núcleo.

Cromocentros: falsos nucleolos o grumos de cromatina, zonas donde la cromatina está más condensada.

Eosinofilia: propiedad que tienen algunas células de teñirse con la eosina.

Filtrado: etapa del procesamiento digital de imágenes que permite mejorar las características de una imagen digital.

Firma de núcleo: función unidimensional del contorno nuclear.

Hematies: glóbulos rojos, que están presentes en la sangre y transportan oxígeno al resto de las células del cuerpo.

Hipercromasia: abundancia de cromatina en una célula. en imágenes digitales se manifiesta con bajos niveles de color.

Hipercromatismo: véase hipercromasia.

Hiperplásico: aumentado de tamaño.

Hipoproteinemia: disminución de la concentración sérica de proteínas ($\ll 6,5g/dl$).

Histiocito: véase macrófago.

Histograma: gráfico utilizado para la representación de la cada uno de los niveles de color.

Idiopático: se dice de cualquier proceso o síndrome con varias causas posibles conocidas, pero que, en el caso concreto que se estudia, se desconoce la causa específica, porque no se puede demostrar ninguna de las posibles causas. diagnóstico de exclusión.

Ldh: enzima catalizadora que se encuentra en muchos tejidos del cuerpo, pero es mayor su presencia en el corazón, hígado, riñones, músculos, glóbulos rojos, en el cerebro y en los pulmones.

Linfa: líquido corporal que recorre los vasos linfáticos.

Linfoma de hodking: véase linfoma.

Linfoma: enfermedad cancerosa que se origina en el sistema linfático.

Líquido pleural exudado: tipo de derrame ocasionado por diversas patologías que afecta la pleura.

Líquido pleural trasudado: indica que la enfermedad no reside en la pleura.

Macrófagos: células presentes en el derrame pleural, de núcleo arrañado y citoplasma de morfología agujereada.

Mediastino: región media entre los dos sacos pleurales, que se extiende desde el orificio superior del tórax hasta el diafragma y desde el esternón y los cartílagos costales por delante hasta la superficie anterior de las doce vértebras dorsales por detrás.

Mesotelioma: tumor primario de la pleura, poco frecuente, derivado de las células mesoteliales que puede ser benigno o maligno

Metastásico: véase metástasis. Metástasis: propagación a órganos distantes del tumor primario maligno o cáncer que ocurre generalmente por vía sanguínea o linfática.

Neoplasia: sinónimo de tumor, en el sentido de cualquier proceso que curse con proliferación celular excesiva, se aplica generalmente para tumores malignos.

Nivel de color: valor asociado con la cantidad de luz detectada por un dispositivo de adquisición. depende del número de bits empleado, para 8 bits se encuentra entre el intervalo $[0, 255]$.

Papilar: que tiene forma de verruga o papila.

Patología pleural: enfermedad de la pleura.

Perinuclear: situado alrededor del núcleo.

Pleomórfico: que adquiere varias formas. Preprocesamiento: véase filtrado.

Presión oncótica: presión ejercida por las proteínas dentro de los vasos sanguíneos.

Radio Núcleo/Citoplasma: relación entre el tamaño del núcleo y el citoplasma.

Reconocimiento: etapa del procesamiento digital de imágenes que permite la clasificación de los objetos presentes en una escena.

Red neuronal: conjunto de operadores interconectados en forma de red, que son utilizados para resolver problemas de optimización y clasificación, mediante funciones de activación.

Sarcoma: cáncer del tejido conectivo (hueso, cartílago, músculo, vasos sanguíneos).

Segmentación: etapa del procesamiento digital de imágenes que permite la separación de los objetos presentes en una imagen.

Técnicas inmunohistoquímicas: técnicas que utilizan anticuerpos específicos, previamente marcados mediante enlaces químicos con una sustancia que se puede transformar en visible.

Toracocentesis: técnica sencilla, en la cual se administra anestesia local y luego se introduce una aguja dentro del espacio pleural para extraer el líquido.

Transformada de fourier: función matemática que permite obtener las componentes de frecuencia de una señal.

Umbralización: técnica de segmentación que consiste en binarizar una imagen, tomando como referencia un nivel del histograma.

Vacuolización: proceso de formación de vacuolas que tiene lugar en los procesos de degeneración celular.

Vénulas: pequeños vasos sanguíneos que llevan sangre procedente de los plexos capilares.

Índice de tablas

1.	Signos macroscópicos y analíticos de los DP	66
2.	Distancias vectoriales Gradiente y RGB en células malignas. . .	92
3.	Distancias vectoriales Gradiente y RGB en células benignas. . .	93
4.	Número y tipo de imágenes celulares caracterizadas.	103
5.	Salida de la RNA de clasificación	106
6.	Documento de descripción de los casos de uso.	109
7.	Clasificación celular realizada por el modelo computacional. . .	114
8.	Número de Falsos Positivos y Negativos.	114

Índice de figuras

1.	Espacio de color del modelo RGB	30
2.	Escalas del Modelo RGB	30
3.	Espacio de color del modelo HSL	31
4.	Escalas del Modelo HSL	31
5.	Etapas Fundamentales del Procesamiento de Imágenes	33
6.	Producción de Imagen Digital	35
7.	Ejemplos filtrado en dominio espacial	37
8.	Imágenes en negativo	37
9.	Histogramas Bimodal y Multimodal	40
10.	Segmentación por Umbralización	41
11.	Segmentación por Crecimiento de Regiones	42
12.	Crecimiento de región a partir de semilla	44
13.	Funcionamiento gráfico del algoritmo SPLIT	45
14.	Representación de firmas	46
15.	Estructura que representa un PE de una red	50
16.	Modelo No Lineal de una RNA	51
17.	Funciones de Activación más comunes	52
18.	Perceptrón con dos entradas	52
19.	Funciones AND y OR	53
20.	Función XOR	53
21.	Arquitectura de una RNA Perceptrón Multicapa	54
22.	Formas básicas de un elementos estructurante	55
23.	Erosión mediante Elemento Estructurante Y	55
24.	Erosión de una imagen	56
25.	Dilatación mediante Elemento Estructurante Y	56
26.	Dilatación de una imagen	57
27.	Apertura mediante Elemento Estructurante Y	57
28.	Apertura de una imagen	58
29.	Cierre mediante Elemento Estructurante Y	58
30.	Cierre de una imagen	59
31.	Espacio Pleural	60
32.	Derrame Pleural	61
33.	Trasudados	63
34.	Exudado Inflamatorio	64
35.	Toracocentesis	65

36.	Células Mesoteliales	67
37.	Células Mesoteliales Degeneradas	69
38.	Células Mesoteliales Reactivas	70
39.	Células Mesoteliales Hiperplásicas	71
40.	Macrófagos	72
41.	Hematíes	72
42.	Linfocitos	73
43.	Células de adenocarcinoma	75
44.	Células Vacuoladas	76
45.	Carcinoma de Pulmón	77
46.	Carcinoma de mama	78
47.	Carcinoma de Ovario	79
48.	Equipo de adquisición.	80
49.	Imágen digital con diagnóstico de malignidad.	81
50.	Imágen digital con diagnóstico negativo.	81
51.	Cámara de Neubauer	82
52.	Imágenes células malignas	83
53.	Imagen con sus componentes de color RGB	84
54.	Aplicación filtro circular.	85
55.	Aplicación filtro gauss.	86
56.	Aplicación filtro rotacional.	86
57.	Aplicación de ajuste.	87
58.	Imagen con sus componentes HSL.	88
59.	Crecimiento de regiones. Selección de semilla	90
60.	Crecimiento de regiones. Segmentación de núcleo.	90
61.	Crecimiento de regiones. Segmentación de citoplasma.	90
62.	Imagen a segmentar por método de Distancia Vectorial.	95
63.	Segmentación por método Distancia Vectorial en RGB	95
64.	Segmentación por método Distancia Vectorial en IHSL	95
65.	Descripción de Citoplasma: área y perímetro	98
66.	Descripción de Citoplasma: Eje mayor y menor	98
67.	Distancia centroide citoplásmico y nuclear.	99
68.	Descripción de Núcleo: área y perímetro.	100
69.	Modelo Entidad - Relación	102
70.	Gráficas de dispersión medidas morfológicas y plano saturación nuclear.	104
71.	Gráficas de dispersión Núcleo - Citoplasma.	105
72.	Arquitectura de la Red Neuronal Artificial.	107

73.	Convergencia del Algoritmo de Entrenamiento.	107
74.	Diagrama de Casos de Uso	108
75.	Gráfica de validación examen citológico.	116
76.	Gráfico de comparación prueba citológica y modelo propuesto.	117
77.	Inicio de la Herramienta Software.	121
78.	Ventana principal de la herramienta software.	122
79.	Opción abrir imagen.	122
80.	Opción seleccionar célula.	123
81.	Célula seleccionada.	123
82.	Ventana para Tratamiento de imagen celular.	124
83.	Imagen celular filtrada y ajustada.	125
84.	Segmentación de un núcleo.	126
85.	Segmentación de un citoplasma.	127
86.	Descripción de parámetros nucleares.	128
87.	Descripción de parámetros Citoplasma y Núcleo.	128
88.	Resultado de la Clasificación Celular.	129

RESÚMEN

TÍTULO:

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA BASADO EN REDES NEURONALES ARTIFICIALES PARA LA CARACTERIZACIÓN DE CÉLULAS MALIGNAS EN FLUIDO PLEURAL.*

AUTORES:

JUAN FRANCISCO REMOLINA CAVIEDES
ADRIÁN FERNANDO RODRÍGUEZ VEGA.**

El presente proyecto continúa el trabajo realizado por el Grupo de Investigación en Ingeniería Biomédica (GIIB), en la línea de tratamiento de imágenes médicas, orientada a la construcción de aplicaciones futuras que permitan contribuir al diagnóstico de ciertas patologías que se apoyan en la lectura de Citología de Fluido Pleural, de tal manera que se reduzca el error ocasionado por subvaloración o sobrevaloración de las características cuantitativas y cualitativas de las células observadas. La investigación ha sido enfocada en la caracterización y detección de lesiones glandulares dado que éstas han mostrado un aumento sensible en las tasas de incidencia [Oropesa and Soto, 2002].

*Trabajo de Grado

**Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas.
Directora: Olga Mercedes Álvarez Ojeda.
Co-Director: Víctor Eduardo Martínez Abaunza.

El desarrollo informático permite la identificación de células afectadas por lesiones neoplásicas aumentando la precisión diagnóstica que actualmente, mediante estudio citológico, presenta un porcentaje importante de falsos positivos, conduciendo a tratamientos innecesarios; así como falsos negativos que a pesar de su baja frecuencia conduce a diagnósticos tardíos peligrosos para la vida del paciente.

El desarrollo de la aplicación software está soportado por la metodología clásica del procesamiento digital de imágenes y el proceso unificado de desarrollo de software. Se recomienda continuar con la investigación que permita determinar sitios primarios de neoplasia mediante el análisis de gran número de imágenes celulares., el uso de arquitectura en paralelo y el desarrollo de modelos computacionales basados en paradigmas evolutivos para la construcción de algoritmos más robustos.

Palabras Clave: *Tratamiento Digital de Imágenes, Morfología Matemática, Redes Neuronales Artificiales, Citología Pleural, Células Mesoteliales, Adenocarcinoma.*

ABSTRACT

TITLE:

DESIGN AND IMPLEMENTATION OF A SYSTEM BASED ON ARTIFICIAL NEURAL NETWORKS FOR CHARACTERIZATION OF MALIGNANT CELLS IN PLEURAL FLUID.***

AUTHORS:

JUAN FRANCISCO REMOLINA CAVIEDES
ADRIÁN FERNANDO RODRÍGUEZ VEGA.****

This project continues the work done by GIIB (Grupo de Investigación en Ingeniería Biomédica), in digital medical image processing. Focused on building of applications that will decrease the error caused by undervaluation or overvaluation of quantitative and qualitative cells's characteristics in effusions pleural. Research has been developed in characterization and recognition of glandular lesions, due to this has been shown a sensitive growing on incidence rates.

Informatic tools advancing allows neoplasical lesion cells identification more accurately.

*** Degree work

**** Physical-Mechanical Engineering Faculty. Computer Science and Systems Engineering School.

Director: Olga Mercedes Álvarez Ojeda.

Co-Director: Víctor Eduardo Martínez Abaunza.

Software development process was supported between classical methodology for digital image processing and rational unified process. It's recommended that research continuance is on founding primary sites by analysis with a large cells number, distributed architecture for parallel algorithms and computational models on evolutionary paradigms.

Keywords: *Digital Images Processing, Mathematical Morphology, Artificial Neural Networks, Pleura Cytology, Atypical Glandular Cells, Mesothelial Cells, Adenocarcinoma.*

1. INTRODUCCIÓN

La patología pleural ha evidenciado un preocupante incremento a nivel mundial, sobretodo la originada por procesos neumónicos dentro de los cuales presenta mayor frecuencia el derrame de pleura, generalmente asociado con fallos respiratorios agudos, infecciones, o tumores neoplásicos; éstos últimos de especial interés.

A pesar que el adenocarcinoma es el tipo más común de células neoplásicas en líquido pleural, también están presentes en derrames de pleura células tales como mesoteliales atípicas, reactivas, macrófagos, linfocitos, etc., que en muchos casos, debido a semejanzas morfológicas, condicionan el dictamen médico. El examen citológico es el principal método de estudio, remplazando procedimientos invasivos con similares grados de precisión diagnóstica con tasas de sensibilidad muy bajas que oscilan entre 50 % y 80 % con especificidad del 97 % [Pertuze and Aparicio, 1997].

Éstas tasas porcentuales abren la posibilidad de diagnosticar lesiones neoplásicas inexistentes o, más preocupante aún, descartar aquellos casos neoplásicos que escapan de una detección oportuna.

En un estado normal las estructuras clave son: núcleo, un patrón de cromatina uniforme, posición dentro de la célula, borde, forma, membrana nuclear prominente y lisa. Los procesos patológicos que condicionan la aparición de un derrame pleural conducen a una serie de cambios reactivos que pueden ser interpretados por error como signos de malignidad: binucleación, multinucleación, aumento en la relación núcleo/citoplasma. Los cambios debidos a cáncer están relacionados con alteraciones similares de éstas estructuras: aumento en el tamaño nuclear muy importante, elevada relación núcleo/citoplasma, irregularidad en el contorno nuclear y alteraciones en la textura de la cromatina o nucléolos muy llamativos.

El presente trabajo funge de complemento al implementado por Delgado et al. (2006) cuyo objetivo principal fue la determinación de criterios morfológi-

cos para la clasificación de células mesoteliales como no tumorales malignas en líquido pleural, a través de la herramienta informática MESOFT.

Durante la construcción de la herramienta se utilizó como metodología el Proceso Unificado de Desarrollo Software. Las imágenes fueron tomadas mediante una cámara fotográfica digital acoplada a un microscopio y conectada a un computador.

La investigación está orientada a la construcción de aplicaciones futuras que permitan contribuir al diagnóstico de ciertas patologías, que mediante la lectura de Citología de Fluido Pleural, permitan reducir el error ocasionado por subvaloración o sobrevaloración de las características cuantitativas y cualitativas de las células mesoteliales observadas.

El aporte de la Inteligencia Artificial al tratamiento de imágenes digitales mediante el uso de Redes Neuronales Artificiales y demás, ha abierto un campo de posibilidades insospechado que apenas comienza a tomar fuerza en el ámbito local. Trabajos desarrollados por el GIIB, de la Universidad Industrial de Santander, han impactado positivamente en el sector salud, con amplio reconocimiento nacional e internacional; motivando nuevos actores y procesos de investigación, pero también demostrando el importante valor de las nuevas tecnologías en la solución de problemáticas sociales.

2. PRESENTACIÓN DEL PROYECTO

2.1. OBJETIVOS

2.1.1. OBJETIVO GENERAL

Desarrollar un software mediante un modelo computacional que respalde la identificación de células malignas en líquido pleural.

2.1.2. OBJETIVOS ESPECÍFICOS

- Establecer los criterios cualificables y cuantificables de los parámetros correspondientes a células malignas en fluido pleural.
- Desarrollar algoritmos para la identificación de parámetros en células de líquido pleural que permitan clasificarlas como células malignas.
- Construir un software que permita clasificar las células de acuerdo con los parámetros extraídos de un extendido citológico de líquido pleural.

2.2. JUSTIFICACIÓN

2.2.1. DEFINICIÓN DEL PROBLEMA

Es alarmante el incremento de la patología pleural a nivel mundial¹, sobre todo la originada por procesos neumónicos traumáticos tal como la ocupación pleural, ya sea por la acumulación de aire (neumotórax) o líquido (derrame pleural) en la cavidad pleural [Chávez et al., 2005], siendo ésta última la de mayor frecuencia; el 62 % de pacientes ingresados en una unidad médica tenían derrame pleural, de los cuales solo un 41 % fueron detectados a su ingreso [Mattison et al., 1997].

Ésta anormalidad puede estar asociada con enfermedades tales como insuficiencia cardíaca, cirrosis, diálisis peritoneal, malnutrición, hipoalbuminemia,

neumonía, tuberculosis y enfermedades del tejido conectivo; así como la presencia de tumores malignos.

Un 75 % de los derrames pleurales neoplásicos son secundarios a cáncer pulmonar, cáncer mamario y linfomas. Alrededor de un 50 % de pacientes con cáncer bronquial y un 40 % a 50 % de pacientes con cáncer de mama diseminado presentan un derrame pleural en el curso de su evolución. Otros tumores primarios también suelen dar metástasis pleural, como el caso de los carcinomas ováricos, cáncer cervicouterino, melanomas, sarcomas, cáncer del tubo gastrointestinal, etc.

El adenocarcinoma es el tipo más común de células neoplásicas encontradas en líquido pleural, proviniendo en su gran mayoría de neoplasias de pulmón, mama y ovario. Sin embargo es importante tener en cuenta otros tipos celulares presentes en derrames de pleura, tales como células mesoteliales atípicas, reactivas, macrófagos, linfocitos, etc.; que en muchos casos, debido a similitudes morfológicas, condicionan indebidamente el dictamen médico; por ejemplo, desde el punto de vista citológico e histológico, suele ser difícil diferenciar un adenocarcinoma metastásico de un mesotelioma, necesitamos técnicas de laboratorio avanzadas.

Muchos estudios han demostrado que la sensibilidad del examen citológico oscila entre 50 % y 80 % [Pertuze and Aparicio, 1997], con especificidad del 97 % [Maghfoor et al., 2000]; además, que el rendimiento del examen patológico y la presencia de resultados positivos dependen de la experiencia del patólogo [Delgado et al., 2006].

Comprobándose que dos muestras para citología son suficientes, dada la escasa rentabilidad aportada por una tercera muestra [Rodríguez, 1993].

Dentro de las técnicas de diagnóstico en el derrame pleural, se encuentra también la biopsia por aguja cuya sensibilidad está cercana al 50-60 %, relativamente menor en comparación con la citología pleural, debido a que los exudados de origen neoplásico tienen células malignas por descamación desde metástasis pleurales o por salida desde la sangre al líquido pleural de elementos tumorales sin invasión pleural [Fariña and Rodríguez, 1996], reafirmando lo dicho por [Johnston, 1985], cuando considera el citodiagnóstico como un método más eficaz que la biopsia pleural. Con todo y ello, es aún

preocupante el porcentaje importante de falsos negativos debido a errores de muestreo y de interpretación. Éstos mismos errores no solo pueden conducir a la sospecha de lesiones malignas inexistentes sino en el peor de los casos al diagnóstico tardío de tumores no benignos sin la intervención oportuna y el tratamiento necesario que puedan mejorar la calidad de vida del paciente. Con el presente proyecto se espera disminuir éste tipo de irregularidades, construyendo las bases que permitan establecer un diagnóstico más acertado.

2.2.2. IMPACTO ESPERADO

Se espera reducir el error ocasionado por la subvaloración o sobrevaloración de las características cuantitativas y cualitativas de las células mesoteliales atípicas y reactivas observadas en la citología de fluido pleural, contribuyendo a la exactitud diagnóstica de ciertas patologías relacionadas con malignidad del tipo adenocarcinoma, permitiendo la oportuna intervención y el tratamiento adecuado que mejore la calidad de vida del paciente.

Además, se pretende plantear las bases de futuras investigaciones que den continuidad al estudio de la caracterización de células malignas del tipo adenocarcinoma permitiendo determinar sitios primarios de neoplasia.

A nivel investigativo, el modelo busca sentar los cimientos para futuros proyectos que relacionen conocimientos de ingeniería y de medicina, no solo con el fin de aportar soluciones a problemáticas de connotación social, sino también afianzar el papel de la Ingeniería de Sistemas como ciencia aplicada a la Medicina mediante el uso y el diseño de nuevos modelos computacionales que aporten positivamente al desarrollo del conocimiento científico.

2.2.3. VIABILIDAD

Existen ventajas como las encontradas en el Departamento de Patología de la Escuela de Medicina y el Grupo de Investigación en Ingeniería Biomédica (GIIB) de la Escuela de Ingeniería de Sistemas e Informática, que además de contar con el recurso humano y tecnológico, facilitan la investigación científica en búsqueda de herramientas tecnológicas que permitan dar respuesta a interrogantes y necesidades propias del afán académico que no solo involucra sino que también determina el curso de su respectivo devenir profesional.

3. MARCO TEÓRICO

3.1. TRATAMIENTO DIGITAL DE IMÁGENES

El tratamiento digital de imágenes es bastante útil cuando se quiere obtener mejoras de la información gráfica para hacer más fácil la interpretación humana, procesar datos de la información escénica captada y percibida por dispositivos de sistemas de procesamiento de imágenes.

Una imagen puede ser definida como una función bidimensional $I(x, y)$, donde x e y son coordenadas en el plano, y la amplitud de I en algún par de coordenadas (x, y) es llamada intensidad o nivel de gris en ese punto. Cuando x e y , y los valores de la amplitud son todos finitos, cantidades discretas, se le llama a la imagen imagen digital. En el diseño y análisis de sistemas de tratamiento digital de imágenes es conveniente y necesario caracterizar matemáticamente la imagen a ser procesada.

Los sistemas adquieren los datos píxel a píxel, la imagen es guardada como un conjunto de píxeles. El proceso de digitalización transforma datos análogos continuos en datos digitales discretos. Un número de valores finitos enteros es usado para representar todos los valores de los datos originales y continuos.

La digitalización deja a los datos con un formato adecuado para su manipulación por parte del computador. Estos pueden ser transformados en cualquier función matemática elegida. Cada sistema de procesamiento de imágenes tiene su propio software asociado, cada uno desarrollando una tarea diferente. La secuencia de tareas aplicadas a una imagen se denomina tratamiento de imágenes, la cual puede variar dependiendo de las metas que se tengan en vista.

3.1.1. MODELOS DE COLOR

En esta sección se trata sobre el modelo RGB (Red, Green, Blue) y el modelo HSL (Hue, Saturation, Luminance). El primero es orientado a Hardware y el segundo orientado hacia la manipulación de color.

3.1.2. MODELO RGB

El color aparece en sus componentes espectrales primarias de rojo, verde y azul. Basado en un sistema de coordenadas cartesianas. En la Figura 1 se observa que el subespacio de color de interés es un cubo. En este modelo los colores son definidos por todos los vectores que se extienden desde el origen hasta cualquier punto del cubo.

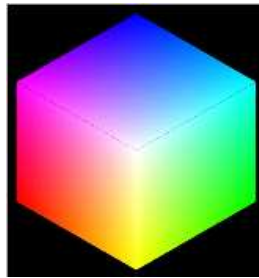


Figura 1: Espacio de color del modelo RGB

Las imágenes de este modelo consisten en tres planos independientes, uno por cada color primario, los cuales se superponen para formar una imagen en color. La Figura 2 muestra las variaciones que se presentan en los colores primarios del modelo RGB. Cuando la proporción de colores es igual en sus tres componentes se obtiene la escala de grises.



Figura 2: Escalas del Modelo RGB

3.1.3. MODELO DE COLOR HSL

En este modelo, el color se representa como dos conos unidos por su base (Figura 3). La componente H representa el matiz del color (rojo, amarillo, verde, cian, azul, magenta) y está definida como un ángulo; La componente S o saturación es la cantidad del matiz que ha sido diluido en blanco. Mientras que la componente L es la escala de grises o Luminancia. En este modelo las componentes del matiz y saturación están íntimamente relacionadas con la forma en que los seres humanos perciben el color.

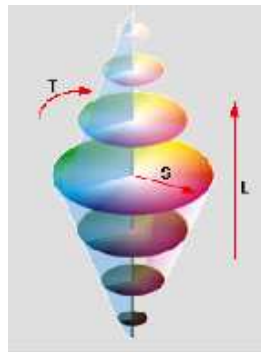


Figura 3: Espacio de color del modelo HSL

En la Figura 4a se observa una escala de matices en la que se distinguen tonos naranjas, amarillos, verdes, azules, etc. En la escala de saturaciones para un estímulo amarillo (Figura 4b) se observa cómo varía el tono en función de la saturación. En la escala de luminosidades (Figura 4c), el matiz emite más o menos luz de acuerdo con la modificación que se haga en los valores.



(a) Matices



(b) Saturación para el amarillo



(c) Luminancia para el amarillo

Figura 4: Escalas del Modelo HSL

Conversión del Modelo RGB a HSL La representación del espacio RGB en términos de coordenadas polares 3D es usada frecuentemente en el análisis de imágenes. Aunque presenta algunas fallas, ofrece una representación más intuitiva, y revela características de la imagen que no son visibles con ese modelo. La transformación entre estos dos sistemas es una simple operación de coordenadas [González and Woods, 1996].

Las siguientes funciones proporcionan los valores HSL en el intervalo $[0,1]$. Se debe cumplir también que las componentes RGB se encuentren normalizadas en el mismo intervalo. La función para el matiz es:

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R - G) + (R - G)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right\} \quad (1)$$

Para la saturación se tiene:

$$S = 1 - \frac{3}{R + G + B} [\min\langle R, G, B \rangle] \quad (2)$$

Finalmente, para la luminancia:

$$L = \frac{1}{3} \langle R + G + B \rangle \quad (3)$$

Sin embargo, en ésta transformación se encuentra que:

- La saturación no necesariamente tiene un bajo valor para píxeles acromáticos.
- La información cromática y de brillo no son independientes. La función de luminancia es usada para normalizar la saturación, así que los valores de saturación dependen en gran medida de la función brillo.

Las anteriores desventajas deben solucionarse entonces, dentro de una nueva alternativa se tiene que las cantidades máx-mín asociadas a la saturación permiten una mejora del espacio HLS. Sin embargo, esta semi-norma no puede utilizarse para el brillo o luminancia, así que para ésta se utiliza la ecuación recomendada por el estándar de televisión de alta definición. Obteniéndose un sistema HLS mejorado o Improve Hue Luminance Saturation System (IHLS) [Hanbury and Serra, 2003]. El cual está fundamentado en las siguientes ecuaciones:

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R - G) + (R - G)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right\} \quad (4)$$

$$H_{IHS} = \left\{ \begin{array}{ll} 360^\circ - H' & \text{Si } B > G \\ H' & \text{en otro caso} \end{array} \right\} \quad (5)$$

$$S_{IHS} = \text{máx}(R, G, B) - \text{mín}(R, G, B) \quad (6)$$

$$L_{IHS} = 0,213R + 0,715G + 0,0072B \quad (7)$$

Estas nuevas ecuaciones permiten separar la información del color de la información lumínica.

3.2. ETAPAS DEL TRATAMIENTO DIGITAL DE IMÁGENES

Las etapas [González and Woods, 1996] del tratamiento digital de imágenes son básicamente: Adquisición, preprocesamiento, segmentación, descripción e interpretación.

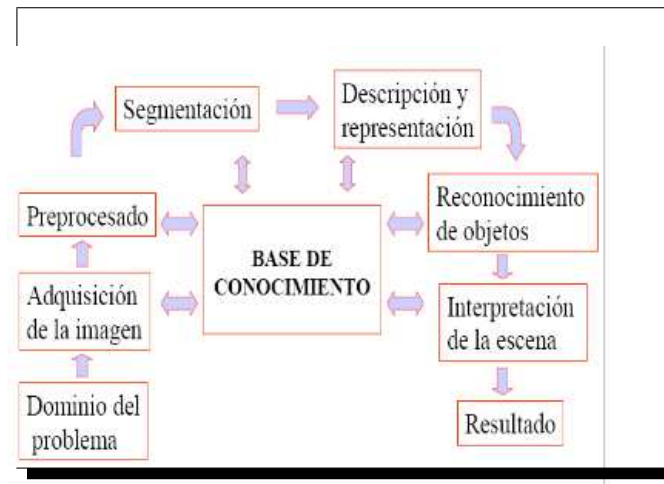


Figura 5: Etapas fundamentales del procesamiento de imágenes. Fuente: *Ibíd.*, p. 1-22.

Adquisición: Etapa en la cual se obtienen las imágenes mediante dispositivos adaptados para captar imágenes junto con otros dispositivos que se encargan de digitalizar las imágenes.

Preprocesamiento: En esta etapa se llevan a cabo una serie de operaciones que tienen como propósito mejorar la imagen para así facilitar su posterior uso.

Segmentación: Se encarga de separar los objetos de una imagen (objetos y fondos) con el fin de estudiar cada uno o los objetos que sean necesarios para cierto fin.

Descripción: Lo que se busca en esta etapa es asignarle propiedades a los objetos que contribuyan con el análisis de la imagen.

Interpretación: En esta etapa se clasifican los objetos mediante patrones de reconocimiento (como algoritmos).

3.2.1. ADQUISICIÓN

Los tipos de imágenes de las cuales estamos interesados son generadas por combinaciones de una fuente de iluminación y la reflexión o absorción de energía de esa fuente por los elementos de esa fuente que está siendo reflejada. Se necesita un dispositivo físico sensible a cierta determinada banda del espectro electromagnético (bandas de rayos X, ultravioleta, visible o infrarrojo), que cree una señal eléctrica de salida proporcional al nivel de energía detectado, además se necesita de un digitalizador para transformar la señal de salida del dispositivo físico a tipo digital. La energía entrante es transformada en un voltaje por la combinación de la entrada de corriente eléctrica y un material sensorial que responde al tipo particular de energía siendo detectada. La salida de voltaje en forma de onda es la responsable de los sensores, y una cantidad digital es obtenida por cada sensor.

Como ejemplo en la Figura 6 la primera función realizada por el sistema de proyección de imagen es recoger la energía de entrada y enfocarla sobre un plano. Si la iluminación es luz, el frente delantero del sistema de proyección de imagen es un lente, el cual proyecta la escena vista sobre el plano focal

de la lente, el sensor, el cual coincide con el plano focal, produce salidas proporcionales a la integral de la luz recibida en cada sensor. La salida es una imagen digital como se muestra en la figura.

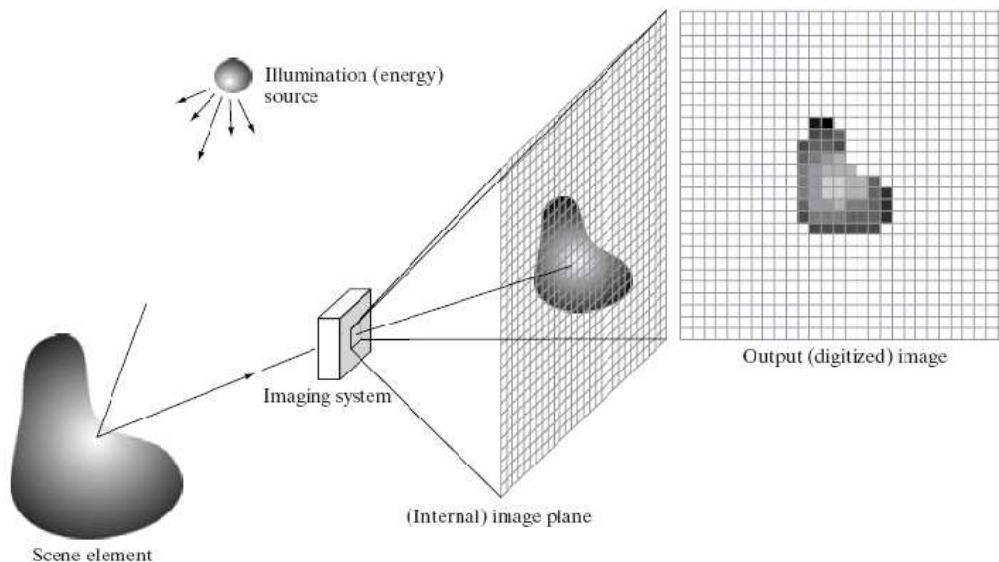


Figura 6: Producción de Imagen Digital

3.2.2. PREPROCESAMIENTO

El objetivo es obtener una imagen tal que, después de procesar la imagen original, el resultado sea más conveniente para algún uso específico.

Existen métodos para llevar a cabo el preprocesamiento como el de mejora en el dominio espacial y el método de dominio de la frecuencia.

Dominio espacial: Se refiere a la manipulación directa de píxeles en una imagen.

Dominio de la frecuencia: Se basa en la modificación de la transformada de Fourier de una imagen.

Cuando una imagen se procesa para la interpretación visual, el espectador es el último juez de los resultados particulares de un método.

Métodos en el Dominio Espacial: El término dominio espacial se refiere al agregado de píxeles que componen una imagen. Además los métodos del dominio espacial se basan en el uso de valores de niveles de gris de los píxeles para lograr la mejora de la imagen. Los métodos del dominio espacial son los procedimientos que funcionan directamente en estos píxeles.

Los procedimientos de los métodos del dominio espacial serán denotados por la expresión:

$$g(x, y) = T[f(x, y)] \quad (8)$$

Donde $f(x, y)$ es la imagen de entrada a ser procesada, $g(x, y)$ es la imagen procesada y T es un operador en f . Por lo tanto T desde el punto de vista matemático puede ser algún conjunto de transformaciones que se deben aplicar a una imagen para que el resultado sea una imagen con unas características visuales mucho mejores.

La idea general consiste en determinar g en un punto (x, y) a partir de los valores de f en un entorno predefinido de (x, y) . Una de las aproximaciones principales en este tipo de formulación se basa en el empleo de las denominadas máscaras (plantillas, ventanas o filtros) básicamente, una máscara es una pequeña distribución bidimensional donde los valores de los coeficientes determinan la naturaleza del proceso como la acentuación de los bordes.

Filtros paso - bajo: Son aquellos que reducen las componentes de alta frecuencia, debido a esto está relacionado con los bordes y detalles marcados de una imagen, el resultado de aplicar este tipo de filtro es hacer más borrosa la imagen.

Filtro paso - alto: Son aquellos que reducen las componentes de baja frecuencia, debido a ello resaltan los bordes y otros detalles.



(a) Imagen con ruido



(b) Imagen filtrada

Figura 7: Ejemplos filtrado en dominio espacial

Métodos basados en los píxeles:

- **Imágenes de negativos** Dada por la expresión:

$$g(x, y) = 255 - f(x, y) \quad (9)$$

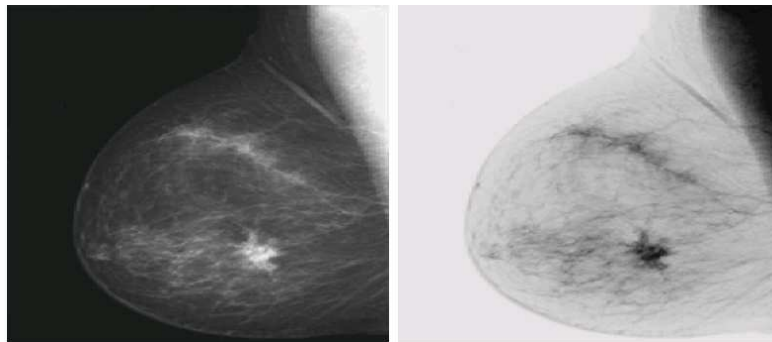


Figura 8: Imágenes en negativo

En algunas situaciones se calcula la imagen invertida de una imagen dada, ya que características que sean difíciles de percibir puedan estar más claras en la imagen invertida.

- **Mejora del contraste** Se puede presentar cierta falta de contraste en una imagen debido a poca iluminación, por dispositivos con un rango

de iluminación bajo, etc. Lo que se quiere es modificar los valores de los píxeles así obteniendo en el rango dinámico de valores, un aumento en los valores de niveles de gris de la imagen. Se deben seleccionar dos valores de nivel de gris, uno por la parte alta de los blancos y otro por la parte baja de los negros. Siendo así aquellos píxeles que estén por encima de la parte alta de los blancos sean más claros y los que están por la parte baja de los negros sean más oscuros.

- **Comprensión del rango dinámico** Se presenta debido a dispositivos detectores que tengan un rango dinámico de niveles de gris mucho mayor en comparación a los niveles de gris de la pantalla de representación de la imagen. Se reduce el rango de valores de nivel de gris para así poder representar las imágenes al rango de la pantalla de representación.
- **Realce de rasgos de intensidad** Teniendo una imagen, se realiza una región de niveles de gris que pertenecen a cierto rango de valores de la imagen.

Métodos basados en el Histograma: Una imagen en escala de grises posee un histograma cuyo rango de valores de nivel de gris está comprendido entre 0 y 255.

- **Ecualización del histograma** Consiste en distribuir los niveles de gris de una imagen, es decir una variación de una forma a otra forma, esta distribución es uniforme queriendo así que exista el mismo número de píxeles para cada nivel de gris.
- **Mejora Local** Este método es para casos en que la distribución de niveles de gris en la imagen es tal que determinadas transformaciones de tipo global pueden ser adecuadas para unas zonas de la imagen como no pueden ser adecuadas para otras.

Sustracción de Imágenes: Este método puede ser útil cuando al momento de capturar una imagen no se garantiza que todas las imágenes se tomen con la misma intensidad de luz. La sustracción puede corregir posibles variaciones que puedan ocurrir en el transcurso de las tomas.

Dominio de la Frecuencia: En el cual sus técnicas se encuentran basadas en el teorema de la convolución. Sea $g(x, y)$ una imagen formada por la convolución de una imagen $f(x, y)$ y un operador lineal invariante $h(x, y)$ denotados por la expresión:

$$g(x, y) = h(x, y) * f(x, y) \quad (10)$$

Entonces, por el teorema de la convolución se cumple la siguiente relación en el dominio de la frecuencia:

$$G(u, v) = H(u, v) * F(u, v) \quad (11)$$

Donde G , H y F son respectivamente las transformadas de Fourier de g , h y f .

3.2.3. SEGMENTACIÓN

La segmentación se encarga de descomponer una imagen en las partes que la conforman, es decir los objetos (regiones con características similares) de interés y el fondo, utilizando algoritmos que permiten identificar las características para poder distinguir los objetos del fondo. Para facilitar la segmentación las imágenes deben estar constituidas por zonas de características (Característica más básica es la amplitud) homogéneas (niveles de gris, textura, contornos, etc.), generalmente estas zonas son los objetos de la imagen.

Los algoritmos de segmentación en imágenes binarias se basan en algunas propiedades de los niveles de gris como la discontinuidad, la similitud y la continuidad.

La segmentación puede parecer un proceso sencillo pero cuando se lleva a cabo es posible que aparezcan ciertos problemas que la complican impidiendo obtener buenos resultados, estos problemas son tales como dificultad en la adquisición ya que se puede encontrar ruido en la imagen que distorsiona las características en distintas regiones, la heterogeneidad debido a que los objetos no puedan presentar una iluminación uniforme, en este caso es la textura la que los diferencia de otros objetos, las formas de los objetos ya que a veces se presentan bordes no definidos y ramificaciones, como se encuentran unidos los objetos la segmentación se hace muy difícil.

Umbralización: La umbralización es una de las técnicas más usuales de la segmentación, utiliza como característica principal la luminancia de cada píxel, es bastante útil si distintos objetos se corresponden con niveles de gris diferentes, si el objeto presenta un pequeño rango de niveles de gris frente a un fondo uniforme se puede establecer un nivel de gris intermedio llamado Umbral para separar objeto y fondo. Para separar por medio de la umbralización es útil recurrir al histograma. El histograma presenta los picos que indican la presencia de zonas homogéneas y valles que establecen los umbrales de separación.

Formulación:

$$U = T[x, y, p(x, y), f(x, y)] \quad (12)$$

Donde $f(x, y)$ es el nivel del punto (x, y) , y $p(x, y)$ representa alguna propiedad local de este punto. Una imagen umbralizada $g(x, y)$ se define como:

$$g(x, y) = \begin{cases} 1 & \text{Si } f(x, y) > U \\ 0 & \text{Si } f(x, y) \leq U \end{cases} \quad (13)$$

Ejemplo con dos niveles de gris dominantes tenemos:

Todos los píxeles $> T(\text{umbral})$, forman un objeto
 Todos los píxeles $< T(\text{umbral})$, forman otro objeto

Umbralización multinivel: Se presenta en imágenes que tienen varios niveles de gris dominantes.

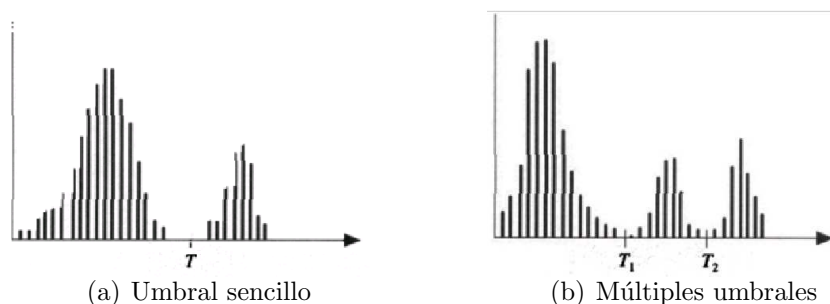


Figura 9: Histogramas Bimodal y Multimodal

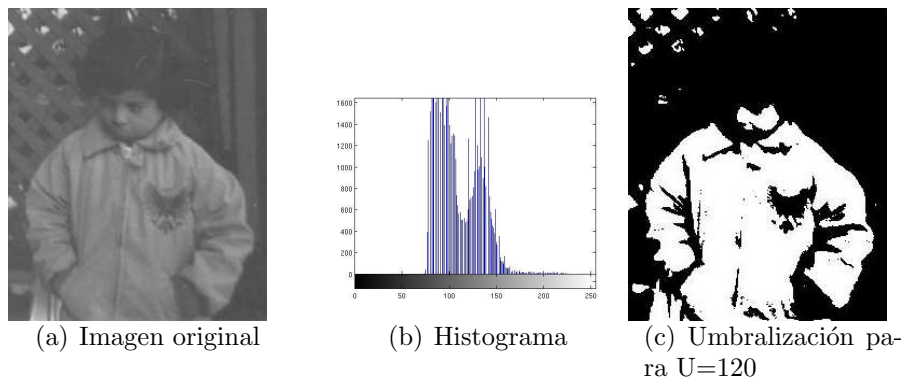


Figura 10: Segmentación por Umbralización

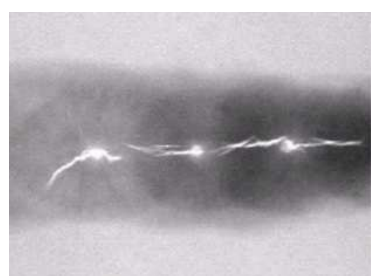
Segmentación por Métodos basados en Regiones:

Crecimiento de Regiones Este método parte de unos puntos o semillas, en la misma región incluye píxeles vecinos con características similares, el proceso culmina hasta que todos los píxeles sean asignados a una de las semillas.

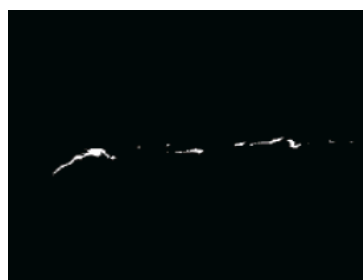
Suponiendo que particionamos una imagen en regiones, se obtienen las regiones $R_1, R_2, R_3, \dots, R_n$ las cuales son disjuntas y su unión debe ser la imagen completa, cada región tiene que ser conexa.

Procedimiento: Se eligen aleatoriamente las semillas (píxeles), a partir de las semillas se van agregando nuevos píxeles a las regiones usando un mecanismo para el crecimiento.

El crecimiento de la región depende del umbral \mathbf{T} que se ha seleccionado. Si el umbral es pequeño lo más probable es que se originen varias regiones pequeñas, en cambio si el umbral es grande, hay fusión de regiones y estas son mucho más grandes.



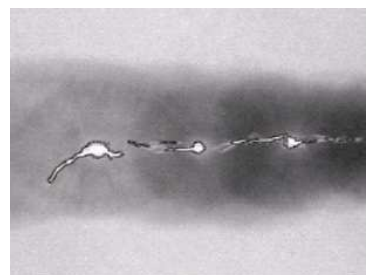
(a) Imagen original



(b) Selección de semillas



(c) Resultado del crecimiento



(d) Borde de la región segmentada sobre la imagen original

Figura 11: Segmentación por Crecimiento de Regiones

Ejemplo a partir de la siguiente subimagen:

1	2	4	5	8
2	0	5	4	8
2	1	7	6	5
1	2	4	6	8
2	2	8	7	4

Si escogemos $T=3$ obtenemos las siguientes regiones:

A	A	B	B	B
A	A	B	B	B
A	A	B	B	B
A	A	B	B	B
A	A	B	B	B

Por otro lado si $T=8$ se obtiene una sola región como se ve a continuación.

A	A	A	A	A
A	A	A	A	A
A	A	A	A	A
A	A	A	A	A
A	A	A	A	A

Uno de los problemas más frecuentes consiste en hallar el umbral para las regiones que se desean segmentar.

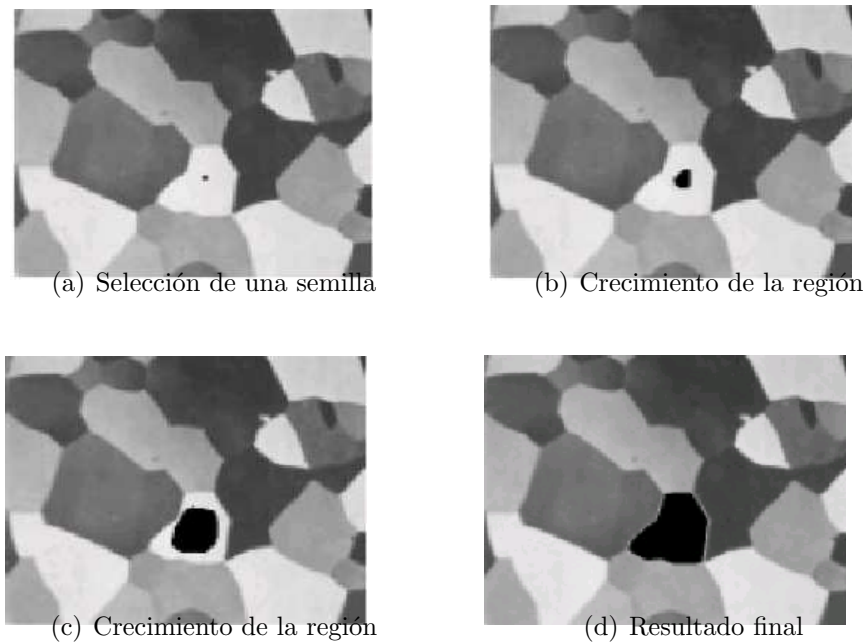


Figura 12: Crecimiento de región a partir de semilla

Segmentación por División y Fusión de regiones Como su nombre lo indica, este método trata de dos procesos característicos, de división y de fusión. También es conocido como el algoritmo de Split Merge.

El algoritmo consiste en hacer divisiones sucesivas partiendo de la imagen original, después de cada división se obtienen cuatro subimágenes, si alguna subimagen no es homogénea, se procede a dividir la subimagen en otras cuatro subimágenes hasta conseguir que sean homogéneas como se muestra en la Figura 13. Esta parte del algoritmo es conocida como el SPLIT.

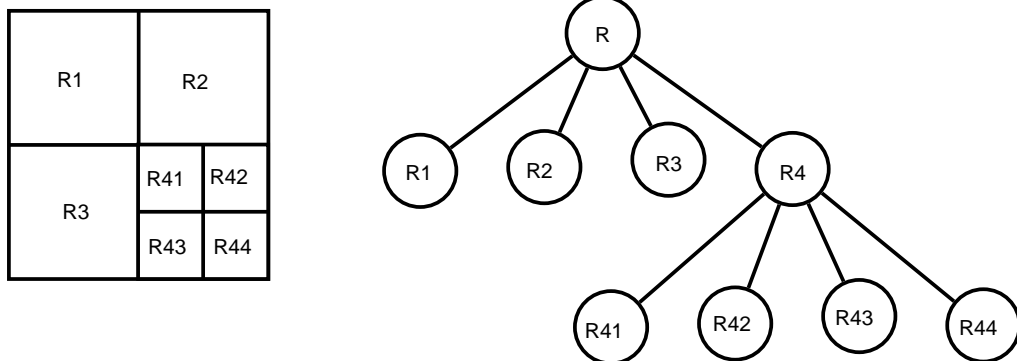


Figura 13: Funcionamiento gráfico del algoritmo SPLIT. Fuente: [González and Woods, 1996]

Luego se analizan todas las parejas de regiones vecinas y se fusionan solo sino no incumplen con el criterio de homogeneidad, el algoritmo finaliza cuando ya no se pueden fusionar más bloques. A esta parte se le conoce como el MERGE.

3.2.4. DESCRIPCIÓN

Posterior a la segmentación llega la etapa en la cual se explica en forma detallada las características que nos interesan de los objetos obtenidos durante la segmentación. Para describir las características externas de los objetos como sus contornos y formas se utilizan los descriptores de forma y para los píxeles que comprenden el interior de los objetos se utiliza los descriptores de textura. También aparecen los descriptores estadísticos que pueden ser usados para contornos como para regiones.

Descriptores de forma:

- **Circularidad** es la relación que existe entre el cuadrado del perímetro del objeto de interés y su área multiplicada por cuatro veces el número π [Sampallo, 2003].

$$C = \frac{P^2}{4\pi A} \quad (14)$$

De ésta manera un objeto completamente circular tendría $C = 1$.

- **Firma** Representación unidimensional de un contorno. Es la distancia que existe entre el centroide y cada punto del borde del objeto, también se puede representar como una función del ángulo [González and Woods, 1996, p.529-530].

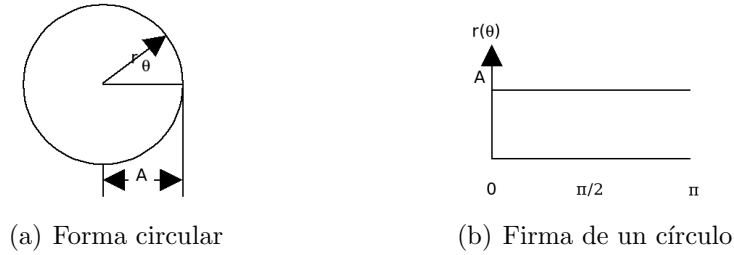


Figura 14: Representación de firmas

De acuerdo con lo anterior, la firma puede ser expresada como:

$$r = \sqrt{(c_{ix} - b_{ix})^2 + (c_{iy} - b_{iy})^2} \quad (15)$$

Siendo c la posición del centroide y b la posición de un píxel del borde.

Descriptores de Textura:

Con los descriptores de textura se busca la caracterización de la información que está contenida en la textura de una imagen con el fin de encontrar un valor que cuantifique las variaciones en los niveles de gris de la imagen. Queriendo que este descriptor sea invariante al tamaño, forma, posición, orientación y al nivel promedio del objeto.

Entropía Es uno de los descriptores utilizados para la representación de texturas, definida por la ecuación 16:

$$entropia = - \sum_{i=0}^{255} \rho * \log(\rho) \quad (16)$$

Siendo ρ el valor del histograma para el nivel $i \in [0, 255]$. La Entropía es baja cuando los elementos son cercanos a 0 ó 1 (por ejemplo, cuando la imagen es uniforme) [Presutti, 2004].

Matriz de Co-ocurrencia Cada elemento matricial representa la probabilidad de ocurrencia de dos valores en escala de gris, separados por una distancia entre pares de píxeles, y un ángulo θ en una dirección determinada. Siempre que se calcula la matriz de co-ocurrencia también se calculan los descriptores de textura señalando la dispersión de los elementos matriciales con respecto a la diagonal principal. Los descriptores son: contraste, correlación, entropía, energía, homogeneidad y varianza.

- **Contraste** definido por:

$$contraste = \sum_{i,j=0}^{N-1} |i - j|^2 \rho(i, j) \quad (17)$$

$\rho(i, j)$ viene siendo la probabilidad de co-ocurrencia de los valores i, j para cierta distancia y N es el tamaño de la matriz.

El contraste es cero para una textura homogénea y un valor alto en caso contrario.

- **Correlación:**

$$correlacion = \sum_{i,j=0}^{N-1} \frac{(i - \mu_i)(j - \mu_j)\rho(i, j)}{\sigma_i\sigma_j} \quad (18)$$

El rango de este descriptor se encuentra en el intervalo $[-1, 1]$. El cual indica el estado en que se encuentra un píxel respecto a toda la región.

- **Energía** es una medida de homogeneidad y obtiene su valor máximo en regiones homogéneas; definida por:

$$energia = \sum_{i,j=0}^{N-1} \rho(i, j)^2 \quad (19)$$

Se encuentra en el intervalo $[0, 1]$.

- **Homogeneidad** toma valores altos cuando la matriz se encuentra a lo largo de la diagonal, es decir, cuando la imagen es localmente homogénea. Posee un comportamiento similar a la energía.

$$homogeneidad = \sum_{i,j=0}^{N-1} \frac{\rho(i, j)}{1 + |i - j|} \quad (20)$$

Descriptores estadísticos: Algunos modelos de la estadística descriptiva como la varianza (Ec. 21), el promedio (Ec. 22), la desviación estándar o la mediana representan algunas características de los objetos.

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum X_i^2}{n} - \bar{X} \quad (21)$$

$$\bar{x} = \frac{\sum_{i=1}^n a_i}{n} = \frac{a_1 + \dots + a_n}{n} \quad (22)$$

3.2.5. INTERPRETACIÓN Y RECONOCIMIENTO

Esta etapa se encarga de etiquetar los diferentes tipos de objetos bajo los resultados obtenidos en la etapa de descripción por medio del reconocimiento de formas e identificación de las características obtenidas, también de identificar y entender cada patrón que contenga información importante para llevar a cabo una investigación de procesamiento de imágenes digitales y así llegar con un alto grado de certeza a la toma de decisiones.

La base fundamental del reconocimiento es la clasificación de patrones mediante la información obtenida en la extracción de características de cada patrón. Los cuales se clasifican según los grupos de medidas y observaciones.

Sistema de reconocimiento de patrones

Un sistema de reconocimiento de patrones está compuesto por un sensor que se encarga de recolectar cada medida para ser clasificada, luego mediante la extracción de características se transforma la información en valores numéricos y finalmente un sistema basándose en las características extraídas clasifica la medición.

Para la clasificación se puede utilizar uno de los dos tipos de procedimientos que existen, los cuales son:

Clasificación Estadística (teoría de decisión): El reconocimiento de patrones se basa en las características estadísticas de cada patrón, suponiendo que han sido generados por un sistema probabilístico.

Clasificación Estructural: Se basa en las relaciones estructurales de las características, utilizando un grupo de aprendizaje del cual ya se conoce la clasificación de la información a priori y es utilizado para el entrenamiento del sistema. Por otro lado sin tener un grupo de aprendizaje la clasificación se basa en cálculos estadísticos.

Redes Neuronales Artificiales (RNA)

Las redes neuronales artificiales ofrecen modelos para la solución de problemas de tratamiento digital de imágenes, aplicando control y modelado de procesos, percepción, diagnóstico médico, clasificación y reconocimiento de patrones. Se le da el nombre de Red Neuronal Artificial a modelos que toman ciertas características del cerebro como base para nuevos modelos de procesamiento implementados como un procesado hardware o software.

La aproximación de redes neuronales, además de su utilidad para encontrar soluciones factibles a problemas, se puede usar para obtener información interna acerca de la naturaleza del problema.

Las RNAs son conjuntos de procesadores paralelos conectados entre si con forma de grafo dirigido, el cual se encuentra organizado de tal modo que la estructura de la red sea la adecuada para el problema que se está considerando.

Cada elemento de procesamiento de la red se representa con un nodo y las conexiones entre los nodos se representan mediante arcos. La dirección del flujo se indica mediante flechas en los arcos.

Elemento general de procesamiento (PEs)

PEs es un elemento de procesamiento que tiene muchas entradas pero una sola salida que se puede aplicar a otros PEs de la red.

Un PE_i recibe una entrada en la forma x_j proveniente del PE_j (siendo x_j la salida del PE_j). Cada conexión con el PE_i tiene asociada una magnitud denominada peso w_{ij} , que es procedente del j ésimo nodo y que llega al i ésimo nodo.

La entrada neta de la i -ésima unidad se define como:

$$Neta_i = \sum X_j W_{ij} \quad (23)$$

Donde el índice j recorre todas las conexiones que posea el PE .

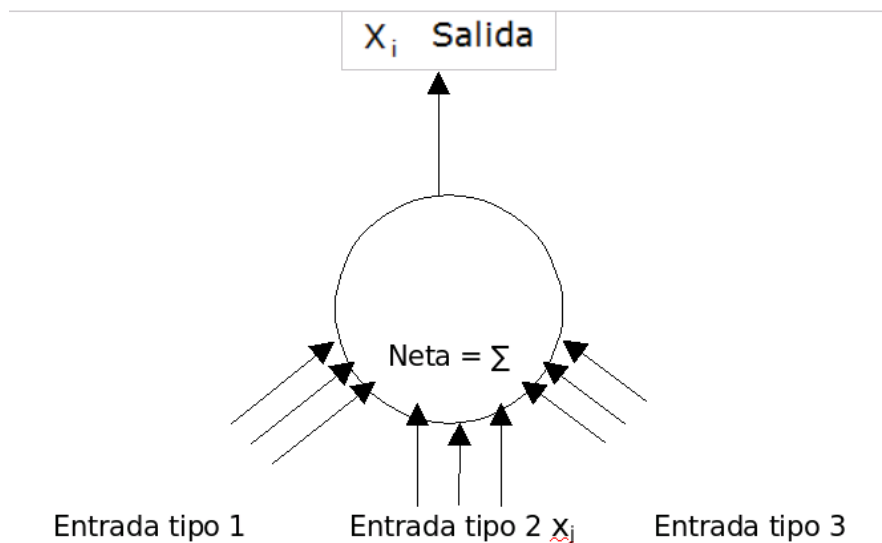


Figura 15: Estructura que representa un PE de una red

Proceso de entrenamiento de una RNA

El proceso de entrenamiento básicamente consiste en ajustar los pesos W para lograr la salida que se desea. El modelo de una neurona contiene tres elementos básicos [Haykin, 1999, p.32] que se muestran a continuación:

- Un conjunto de sinapsis o conexiones, cada una de las cuales está caracterizada por su peso o fuerza de enlace. La señal de entrada x_i es multiplicada por el peso sináptico w_{ij} .
- Un sumador para las señales de entrada.
- Una función de activación que limita la amplitud de la salida de la neurona.

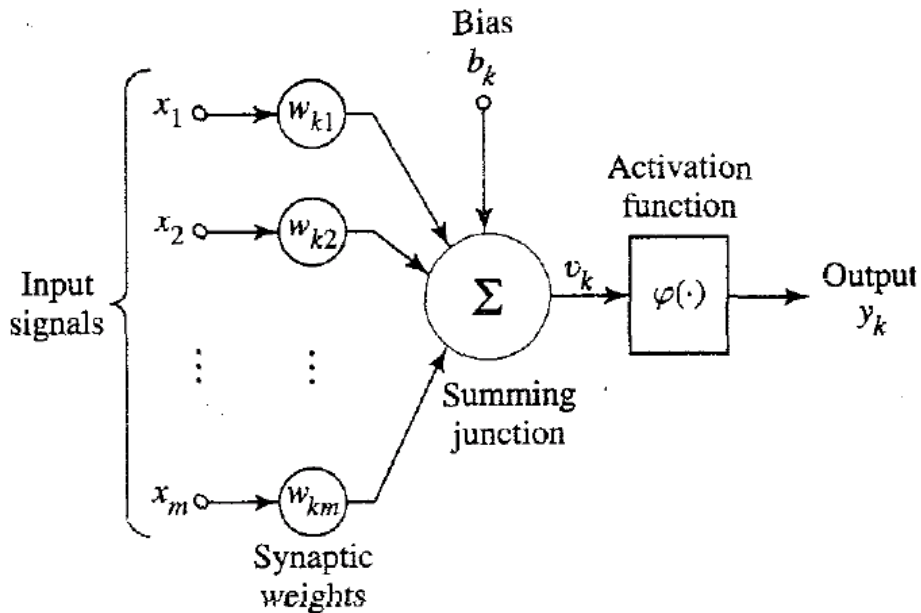


Figura 16: Modelo no lineal de una RNA. Fuente: [Haykin, 1999]

La función de activación designada por $\varphi(\cdot)$ define la salida de la neurona en términos del campo local inducido v . Se pueden identificar tres tipos básicos de funciones de activación:

- **Función Umbral.** Para esta función, presentada en la Figura 17.a se tiene:

$$\phi(v) = \begin{cases} 1 & \text{si } v \geq 0 \\ 0 & \text{si } v < 0 \end{cases} \quad (24)$$

- **Función lineal a trazos.** Para esta función, presentada en la Figura 17.b se tiene:

$$\phi(v) = \begin{cases} 1 & \text{si } v \geq +\frac{1}{2} \\ v + \frac{1}{2} & \text{si } +\frac{1}{2} > v > -\frac{1}{2} \\ 0 & \text{si } v \leq -\frac{1}{2} \end{cases} \quad (25)$$

- **Función sigmoideal.** Es la función de activación más común, presentada en la Figura 17.c Un ejemplo de ésta es la función logística:

$$\phi(v) = \frac{1}{1 + e^{-av}} \quad (26)$$

Donde a es el parámetro de inclinación de la función sigmoideal.

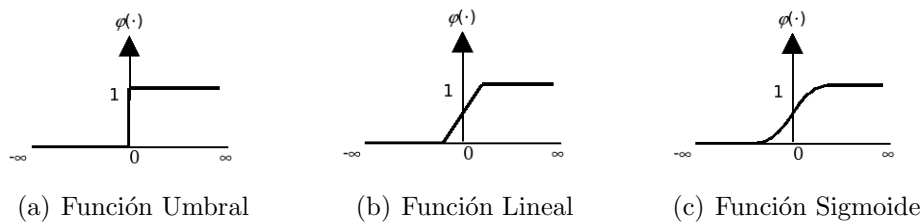


Figura 17: Funciones de Activación más comunes

El Perceptrón

El perceptrón es una neurona artificial, además de ser la unidad básica de inferencia en forma de discriminador lineal que hace parte de una Red Neuronal Artificial.

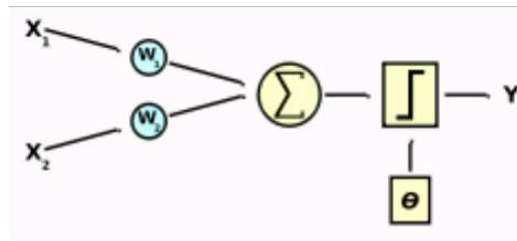


Figura 18: Perceptrón con dos entradas

El perceptrón puede clasificar datos siempre y cuando estos sean linealmente separables.

Las funciones **AND** y **OR** (Figura 19) son linealmente separables, por lo tanto pueden ser aprendida por un perceptrón.

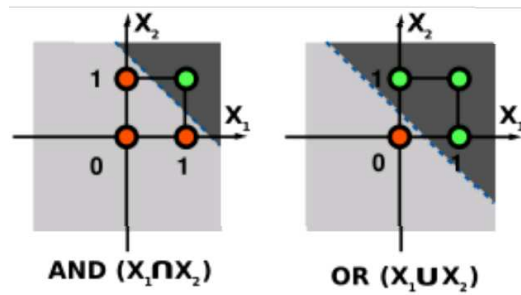


Figura 19: Funciones AND y OR

La función **XOR** (Figura 20) no puede ser aprendida por un solo perceptrón requiriendo por lo menos dos líneas para separar las clases (0,1).

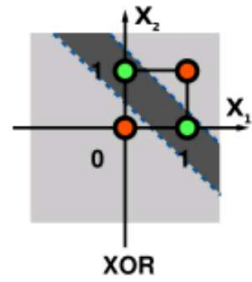


Figura 20: Función XOR

Para resolver el problema planteado por la compuerta XOR se ideó un modelo de red compuesto por varias capas de neuronas intermedias, además de la capa de entrada y la capa de salida, denominada Perceptrón Multicapa. La novedad consistió en que cada nodo estaba conformado por un perceptrón, de ahí que se denominara perceptrón multicapa [Martínez, 2007]. La Figura 21 ilustra la arquitectura de ésta Red Neuronal con dos capas ocultas.

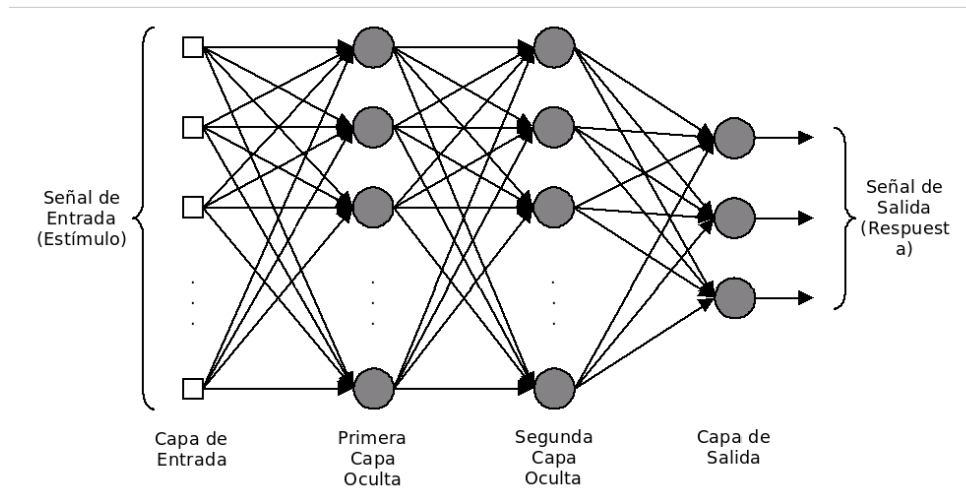


Figura 21: Arquitectura de una RNA Perceptrón Multicapa. Fuente: *Íbid.*, p. 47

3.3. MORFOLOGÍA MATEMÁTICA

La morfología matemática se basa en la teoría de conjuntos, posteriormente estudios la darían a conocer bajo la denominación de morfología matemática como una técnica no lineal de tratamiento de señales.

Dentro de la morfología matemática en el ámbito del tratamiento digital de imágenes existen varias aplicaciones como: Restauración de imágenes, análisis de texturas, aumento de contraste, segmentación, detección de bordes, compresión, etc.

3.3.1. TRANSFORMACIONES MORFOLÓGICAS

Tienen como objetivo extraer estructuras geométricas en los conjuntos sobre los que se opera mediante un elemento estructurante.

El tamaño y la forma del elemento estructurante se escogen de manera a priori de acuerdo a la morfología y a la forma de los objetos que se desean obtener del conjunto en el que se va a interactuar.



Figura 22: Formas básicas de un elementos estructurante

Erosión de una imagen A por un elemento estructurante Y se define como el conjunto de puntos o elemento x pertenecientes a A , de forma que cuando el elemento estructurante Y se traslada a ese punto, el elemento queda incluido en A .

$$A \ominus Y = \{x | (\hat{Y})_x \subseteq A\} \quad (27)$$

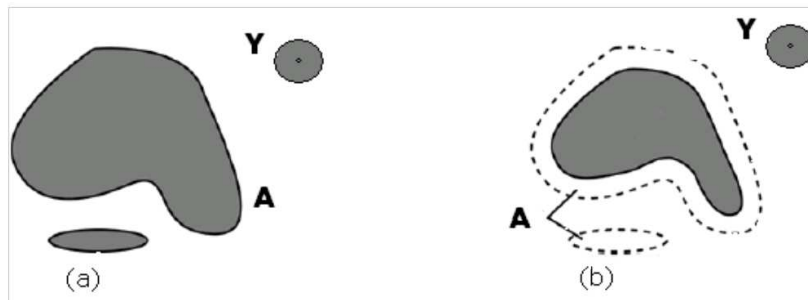


Figura 23: Erosión de A por el elemento estructurante Y . Los elementos conectados del conjunto A más pequeños que Y son eliminados



Figura 24: Erosión de una imagen por un elemento estructurante 3x3

Dilatación dada por la expresión:

$$A \oplus Y = \{x | (\hat{Y})_x \cap A \neq \phi\} \quad (28)$$

El resultado de una operación de dilatación en una imagen A aplicando un elemento estructurante Y es el aumento en la definición de la imagen.

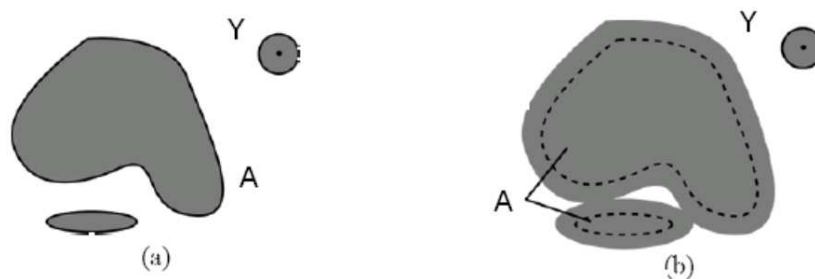


Figura 25: Dilatación de A por el elemento estructurante Y . El conjunto A aumenta su definición



Figura 26: Dilatación de una imagen por un elemento estructurante 3x3

Apertura de una imagen por un elemento estructurante se define como erosión de una imagen por un elemento estructurante seguida de la dilatación por el mismo elemento estructurante. La apertura permite separar objetos traslapados o unidos.

Está definida por [González and Woods, 1996, p.347]:

$$A \circ B = (A \ominus B) \oplus B \quad (29)$$

Otra fórmula matemática alterna es:

$$A \circ B = \cup\{(B)_z | (B)_z \subseteq A\} \quad (30)$$

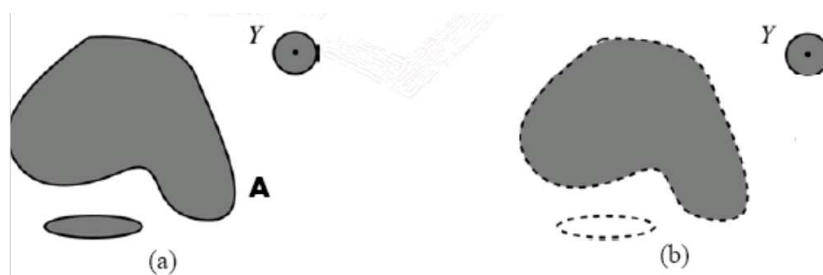


Figura 27: Se aplica apertura al conjunto A con un elemento estructurante Y. Al aplicar la erosión se elimina una estructura que en la operación de dilatación no se puede recuperar.



Figura 28: Apertura de una imagen por un elemento estructurante 3×3 . Se observa que los objetos claros son eliminados.

Cierre se define como la dilatación de una imagen A por un elemento estructurante Y , seguido de la erosión por el mismo elemento estructurante. Definido por la expresión:

$$A \cdot Y = (A \oplus Y) \ominus Y \quad (31)$$

Tal como la apertura tiende a suavizar el contorno de los objetos, pero a diferencia de ésta permite rellenar huecos dentro del objeto, siempre y cuando éstos sean más pequeños que el elemento estructurante [González and Woods, 1996, p.348].

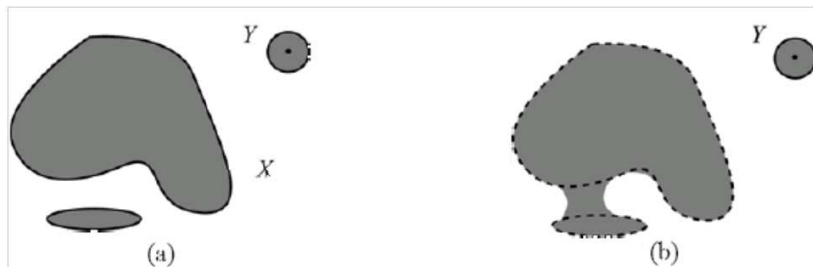


Figura 29: El cierre redondea las concavidades importantes.

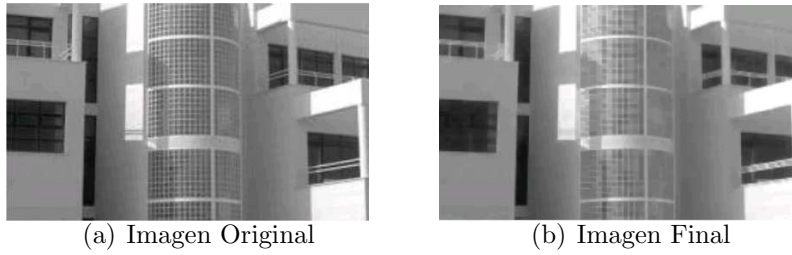


Figura 30: Cierre de una imagen por un elemento estructurante 3x3.

Top-Hat definida como:

$$\rho(A) = A - A \circ B \quad (32)$$

Es el resultado de la resta de la imagen con la apertura. Sirve para extraer las componentes positivas.

Bottom-Hat definida como:

$$\rho(A) = A \cdot B - A \quad (33)$$

Es el resultado de la resta del cierre con la imagen. Sirve para extraer las componentes negativas.

3.4. CITOPATOLOGÍA DE LA PLEURA

3.4.1. GENERALIDADES

La pleura es la capa del tejido mesenquimal que envuelve por separado a cada uno de los pulmones, la cara interna del tórax, la superior del diafragma y el mediastino (Figura 31). La zona de la pleura que está en contacto con los pulmones se denomina capa u hoja visceral y la que se apoya a la pared torácica, diafragma y mediastino se denomina capa u hoja parietal. Entre estas dos capas queda el espacio potencial o virtual de la cavidad pleural, lubricado por unos centímetros cúbicos de un líquido procedente del plasma sanguíneo, el volumen normal de líquido pleural contenido en esta cavidad es de 0,1 a 0,2 ml/kg de peso.

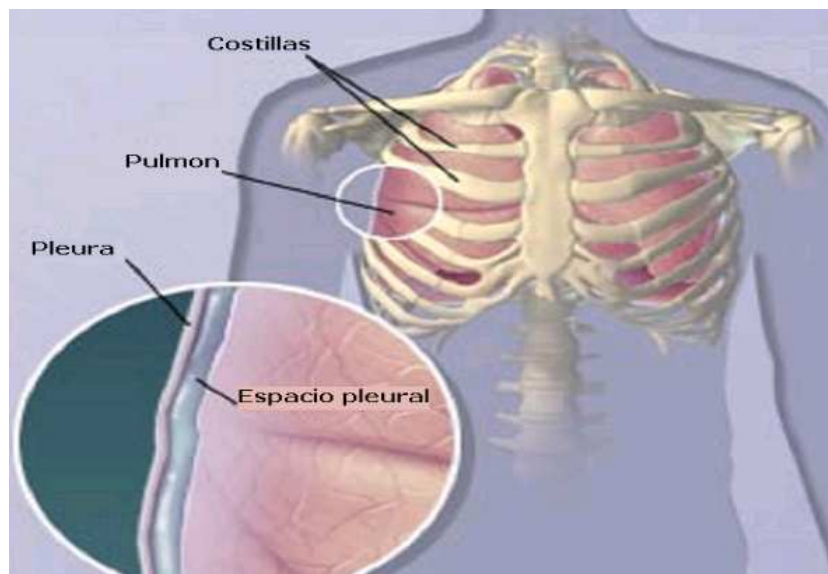


Figura 31: Espacio Pleural. Fuente: Web MD Corporation. En línea: <http://www.ghi.com/topic.aspx?page=537>

Si el exceso de líquido se acumula entre estas dos capas pleurales (Derrame Pleural, ver Figura 32), se compromete el espacio que tiene el pulmón para expandirse dando lugar a una insuficiencia respiratoria y también estimular

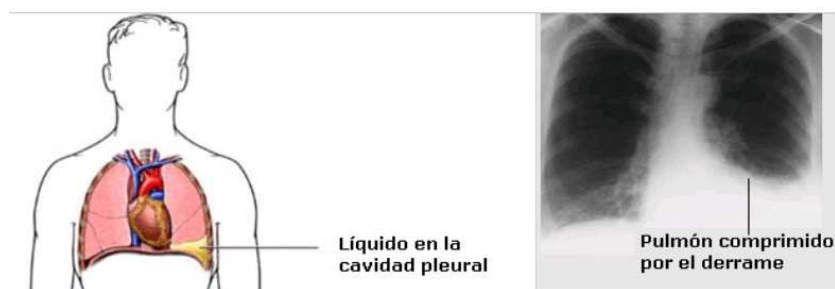


Figura 32: Derrame Pleural. Fuente: En línea: <http://www.sepeap.org/archivos/revisiones/neumologia/empiema2.htm>

las raíces nerviosas sensitivas originando dolor torácico.

El aumento de líquido pleural es ocasionado por alteraciones circulatorias, enfermedades inflamatorias debidas, la mayoría, a infecciones, o diseminación de células tumorales malignas.

Conocer la causa por la que fue ocupada la cavidad pleural, es muy importante, ya que la presencia de células malignas en esta cavidad, casi siempre indica que el proceso canceroso se encuentra muy avanzado y por lo general no se puede llevar un tratamiento quirúrgico para poder curar.

La Toracocéntesis es una técnica que permite extraer el líquido acumulado en el espacio pleural.

3.4.2. ANATOMOFISIOLOGÍA DE LA PLEURA

Cada capa pleural se encuentra cubierta en la superficie de la cavidad por una hilera de células mesoteliales aplanadas y de contorno poligonal. Por debajo de esta hilera de células se encuentra una pequeña cantidad de tejido conectivo con vasos sanguíneos, linfáticos y nervios.

La capa visceral no presenta fibras nerviosas sensoriales, mientras que la capa parietal tiene una gran cantidad de estas fibras.

Por lo general hay menos de 10 cc en cada pleura de un líquido transparente seroso en el que puede haber hasta 3g/dl de proteína y unas 4500 células

por mililitro, entre ellas se encuentran las células mesoteliales descamadas, monolitos, linfocitos y granulocitos que vienen de la sangre. Este líquido es reabsorbido por vía sanguínea, por otro lado las células y las proteínas son absorbidas por los linfáticos de las capas pleurales.

El fluido de los derrames pleurales está compuesto por los constituyentes habituales del líquido pleural, además se pueden presentar células malignas, procedentes del mesotelio o metastásicas, colesterol, gérmenes etc.

Situaciones anómalas como la presión excesiva en la pleura visceral, la presión oncótica disminuida, la obstrucción de los vasos linfáticos o alteraciones de la permeabilidad son las que producen un derrame pleural.

Existen dos tipos de derrames: Los Trasudados (Figura 33) y los Exudados (Figura 34), en ambos casos el exceso de líquido evita que los pulmones se expandan completamente y ocasionan dificultad para respirar.

Trasudados

Este tipo de derrame pleural se presenta cuando el fluido contiene menos de 3 g/dl de proteínas y un peso específico menor o igual 1,015 [Mattison et al., 1997, p.559].

Pueden observarse en insuficiencia cardíaca, cirrosis, diálisis peritoneal, enfermedad del hígado, enfermedad del riñón, desnutrición o hipoalbuminemia.

Esté líquido tiene un aspecto claro, amarillento, y se encuentran en él, pocas células mesoteliales, algunos leucocitos e histocitos.

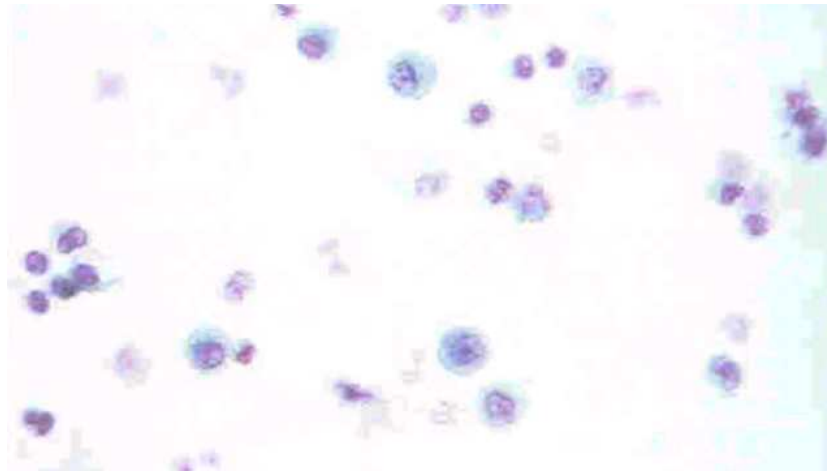


Figura 33: Trasudado. Fuente: [Fariña and Rodríguez, 1996, p.122]

Exudados

Se denomina de esta forma cuando la obstrucción de los vasos linfáticos o alteraciones de la permeabilidad capilar se presentan en una cantidad superior a 3 g/dl y el peso específico mayor de 1,015 [Mattison et al., 1997, p.559].

Estos se deben a múltiples patologías que logran modificar la permeabilidad de las paredes vasculares sanguíneas y/o linfáticas. A veces esta alteración es debida a la ruptura de la pared capilar o a la obstrucción de los capilares linfáticos [Fariña and Rodríguez, 1996].

Se pueden observar en tumores malignos, diversas enfermedades inflamatorias del pulmón o de la pared torácica que secundariamente afectan la capa pleural.

Se destacan las enfermedades inflamatorias, las bacterianas productoras de neumonía, la tuberculosis, las micosis diversas y enfermedades parasitarias pulmonares, el infarto pulmonar, el lupus diseminado, la artritis reumatoide y la enfermedad de weigner.

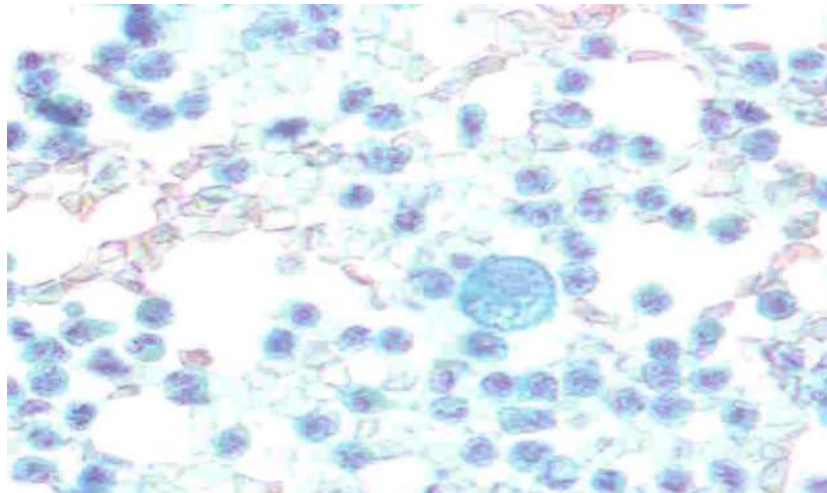


Figura 34: Exudado inflamatorio. Fuente: Íbid., p.122.

3.4.3. CITOPATOLOGÍA DE LOS DERRAMES PLEURALES

Existen dos tipos de técnicas para diagnosticar un derrame pleural: Citología de líquido pleural y la Biopsia pleural.

Citología de Líquido Pleural

Este examen detecta la presencia de células cancerosas y otros factores, al proceso se le denomina Toracocéntesis mencionado anteriormente, este examen se lleva a cabo insertando una aguja en la cavidad pleural para tomar una muestra del fluido. La prueba del líquido se examina bajo el microscopio para la detección de células anormales. Este examen es poco traumático para el enfermo y relativamente fácil de obtener. Del estudio citológico se puede diagnosticar del 80 al 90% de las diseminaciones tumorales a pleura.



Figura 35: Toracocentesis. Fuente: <http://www.msd.es/>

Biopsia Pleural

Cuando los resultados de la Toracocentesis no llegan a determinar la causa del derrame pleural o cuando se necesita una muestra de tejido de un tumor se realiza una biopsia pleural con aguja. Primero se anestesia la piel con el mismo procedimiento de la Toracocentesis. Luego mediante una aguja más grande se extrae un fragmento de tejido de la pleura que es enviado al laboratorio para determinar si existen signos de cáncer o de tuberculosis. La biopsia pleural es una prueba muy precisa que permite diagnosticar estas enfermedades en el 85 al 90 por ciento de los casos. Las complicaciones son similares a las de la toracocentesis.

Las características macroscópicas y analíticas de los derrames sugieren, más no aseguran, algunos diagnósticos (ver Tabla 1).

Elementos Formes No Neoplásicos

Los elementos formes no neoplásicos son las células mesoteliales, hematíes, polimorfonucleares, linfocitos, histiocitos, células plasmáticas, cristales de colesterol, cristales de Charcot-Leyden, cuerpos de asbesto y artefactos extraños como células hepáticas, cutáneas, de músculo, vegetales, etc.

<i>Color</i>	<i>Consistencia</i>	<i>Otro</i>	<i>Probable Causa</i>
Verdoso	Espeso	Olor malo	Empiema
Naranja-Verde	Acuoso		Ictericia
Rojizo-Rosado	Espeso	Coágulos	Hemorragia
Lácteo	Espeso		Leucemia
Pardo	Viscoso		Mesotelioma
Grumos	Espeso		Ca. Mama, timoma
Marrón	Espeso		Melanoma
Verde-Amarillo	Media	< 12 mg/ %ml de GLC	A. reumatoide
Amarillo-Verde	Brillante		C. colesterol
Gris	Media	$\geq 15mg\%ml$ de GLC Densidad: 1,020- 30	Tuberculosis
Lácteo	Denso	Proteínas: 3-4 g %ml Lípidos: 1-4 g %ml	Quilotórax
Amarillo-Claro	Ligero	Densidad < 1,015 Proteínas: 3g/dl	Trasudado

Tabla 1: Signos macroscópicos y analíticos de los DP.

Células Mesoteliales

En general su diámetro varía entre 18 y 40 micras, aunque en ciertas condiciones puede llegar a ser enormes y alcanzar un diámetro de 70 micras, lo que suele asociarse a la aparición de varios núcleos. Sin embargo existe la posibilidad en algunas entidades de ser muy pequeñas con un diámetro aproximado de 10 micras (Figura 36).

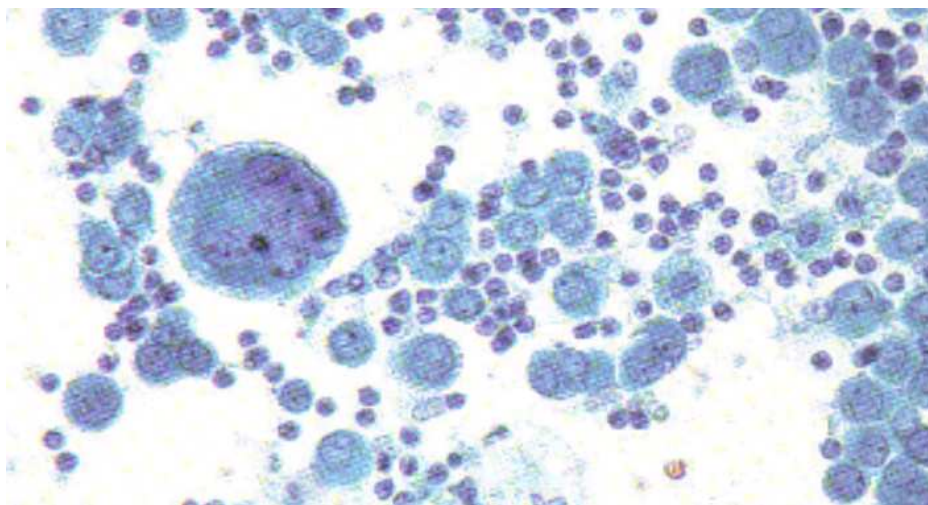


Figura 36: Células Mesoteliales. Fuente: [Fariña and Rodríguez, 1996]

En los tejidos, forman una capa única y lisa sobre superficies corporales, comparten rasgos de células escamosas pero tienen una característica única y distintiva consistente en la presencia de microvellosidades largas y delgadas que intervienen en la absorción; aunque éstas pueden perderse cuando llevan mucho tiempo en el derrame. La activación de la célula mesotelial puede iniciar toda la cascada de procesos implicados en la inflamación así como provocar cambios en la permeabilidad de la membrana pleural y es capaz de emigrar para iniciar procesos de reparación pleural y además, puede transformarse en macrófago.

Suelen aparecer aisladamente, pero a menudo se presentan como agrupaciones monoestratificadas de varias células. Los núcleos suelen ser centrales aunque pueden aparecer desplazados, son regulares, redondos u ovals, presentan una dispersión uniforme de cromatina, que también puede ser pálida y fina o gruesa y oscura. Los cromocentros son muy visibles y pueden encontrarse uno o más nucleolos pequeños y regulares. Los núcleos ocupan entre la tercera parte y la mitad de la superficie celular. La binucleación es frecuente y entre los núcleos suele haber una zona clara. La membrana nuclear es prominente y lisa.

Con la técnica de Papanicolaou, muestran un citoplasma de color verde, azulado o casi anaranjado y suele ser más grueso en torno al núcleo, donde se

demuestra la existencia de filamentos, en la periferia el citoplasma va adelgazándose. Algunas veces puede ser anfófilo. El citoplasma perinuclear es denso y de aspecto esmerilado debido a los filamentos intermedios concentrados en el centro de la célula, en la periferia de la célula hay a menudo una zona de aspecto menos denso que difumina los contornos celulares. La vacuolización puede alcanzar tales extremos que con frecuencia desplaza al núcleo, por lo general situado en el centro, hacia la periferia. En esos casos el núcleo aparece deformado por las vacuolas y muchas veces dificulta el diagnóstico diferencial con el adenocarcinoma. Las células mesoteliales pueden contener células o partículas fagocitadas.

Los procesos patológicos que condicionan la aparición de un derrame conducen a una serie de cambios reactivos interpretados por error como signos de malignidad. Aspectos como la binucleación y la multinucleación generalmente son respuesta a la agresión celular, algunos procesos reactivos son la proliferación de células mesoteliales y un aumento del número de mitosis que provoca un incremento de tamaño nuclear; un aumento en la relación núcleo: citoplasma, cromatina grosera y nucleolos prominentes pueden confundirse con un proceso neoplásico.

Las células mesoteliales reactivas generalmente aparecen sueltas aunque a veces crean grupos cohesivos o estructuras papilares que pueden confundirse con tumores malignos. La presencia de numerosos acúmulos de células mesoteliales de gran tamaño es muy rara en procesos benignos y debe sospecharse malignidad.

Las células mesoteliales en los derrames reciben a veces el adjetivo de degeneradas, atípicas, reactivas o hiperplásicas.

- **Degeneradas** Se considera que son benignas y suelen encontrarse en derrames de larga duración o con intensos signos inflamatorios. Presentan numerosas vacuolas en el citoplasma, las cuales pueden llegar a desplazar el núcleo y formar entre sí células en anillo de sello. Aunque la mayoría de las células benignas en anillo de sello son macrófagos. Algunos procesos como la quimioterapia producen células con citoplasmas y núcleos vacuolizados (Figura 37).

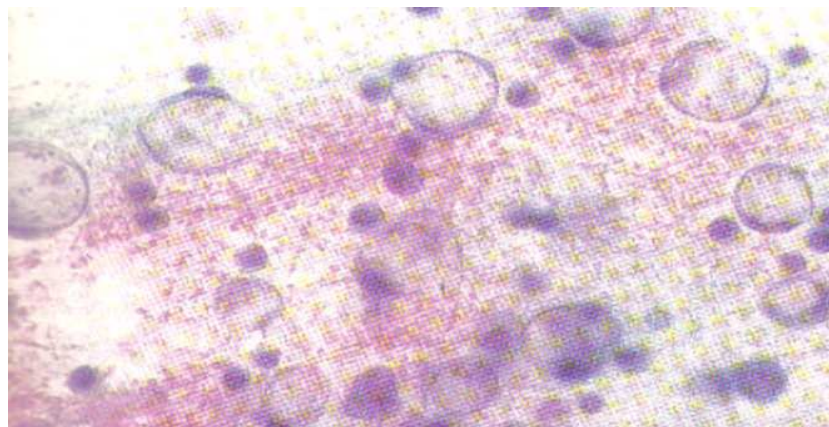


Figura 37: Células Mesoteliales Degeneradas. Fuente: [Fariña and Rodríguez, 1996]

- **Atípicas** Son células con caracteres sugerentes de malignidad. El diagnóstico de derrame con células atípicas casi siempre significa que sólo se han visto varias células de aspecto compatible con malignidad. Debe repetirse el estudio y se denomina derrame sospechoso de malignidad.
- **Reactivas** Abarca diversas variaciones morfológicas que puede adquirir el mesotelio en un derrame acompañado o ausente de células neoplásicas (Figura 38). La célula reactiva no es más grande que la normal, el citoplasma tiene mucha cantidad de filamentos en torno al núcleo, lo que hace más visible la delimitación entre la zona perinuclear y periférica. A menudo se visualizan vacuolas llenas de glucógeno.

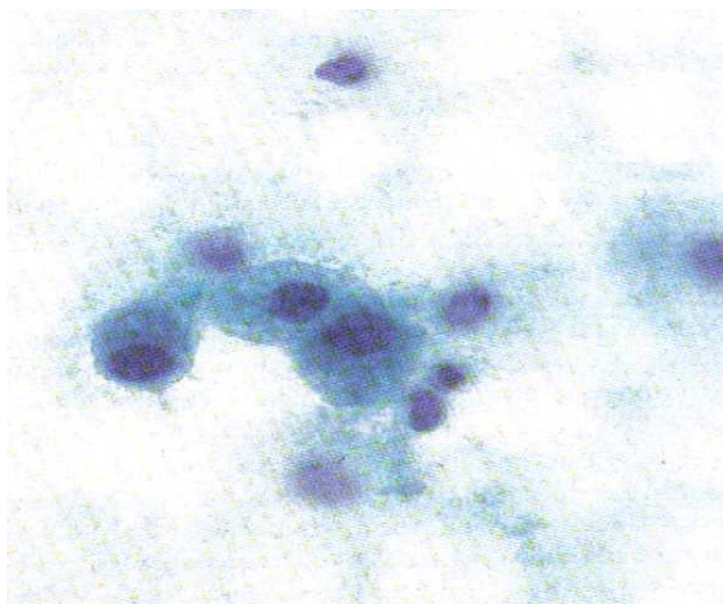


Figura 38: Células Mesoteliales Reactivas. Fuente: [Fariña and Rodríguez, 1996]

La vacuolización citoplasmática puede conducir a las imágenes de células degeneradas o en anillo de sello que se supone uno de los posibles finales de las células reactivas. Pero las células reactivas pueden sufrir mitosis, aunque típicas y por ello su núcleo tiene moderado grado de anisonucleosis. El borde celular puede aparecer liso por haber perdido las microvellosidades al llevar mucho tiempo en el líquido. Las células reactivas entre sí se articulan en dobletes o tripletes o en agrupaciones algo papilares, pero son más pequeñas que las conocidas como grupos proliferativos papilares hiperplásicos.

- **Hiperplásica** Se caracteriza por tener mayor tamaño que la reactiva pero menor que la neoplásica, con núcleo aumentado de tamaño, hipercromático y con nucleólos grandes y más de uno (Figura 39). El citoplasma se acompaña también de una hipercoloración, indicando que la reacción de la célula es del citoplasma y del núcleo.

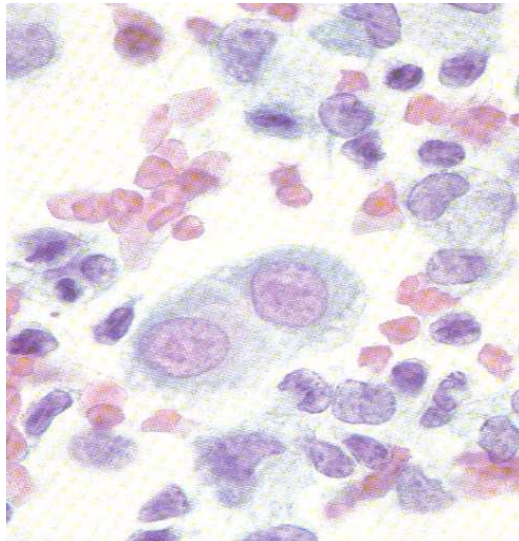


Figura 39: Células Mesoteliales Hiperplásicas. Fuente: Ibid.

En las células malignas el hiperchromatismo nuclear no va acompañado de modificaciones en la tinción citoplasmática, la célula hiperplásica a menudo presenta mitosis, pero son típicas, en la superficie celular tiene, muy frecuentemente, microvellosidades muy largas y enredadas formando vesículas entre ellas, en otras ocasiones hay zonas de la membrana en que se han perdido las microvellosidades.

Macrófagos

La cantidad que está en los derrames es muy variable, pueden presentarse en los derrames exudados por: infecciones, infarto pulmonar, hemotórax, materiales extraños en la cavidad pleural, diseminación neoplásica, histiocitosis, artritis reumatoide y lupus eritematoso diseminado (Figura 40).

El núcleo tiene forma arriñonada y el citoplasma es de morfología algo agujereada, la característica principal de los histiocitos es la fagocitosis*, por lo que se denominan macrófagos. En ellos puede encontrarse células muertas, restos celulares, partículas de carbón, talco, almidón.

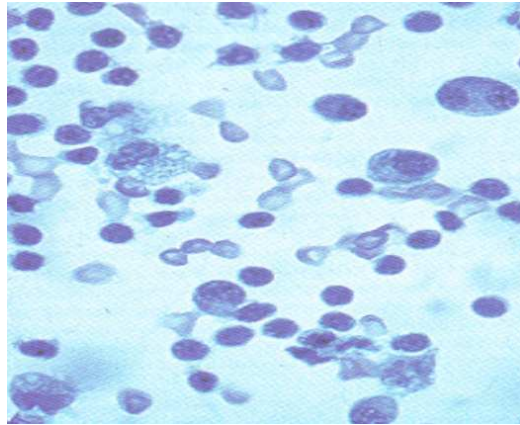


Figura 40: Macrófagos. Fuente: [Fariña and Rodríguez, 1996].

Hematíes

Su color apenas rosado o fuertemente rojizo depende de la relación entre la cantidad de la sangre y el volumen del exudado (Figura 41). Las causas de la presencia de eritrocitos en los derrames es muy variada, entre estas se encuentran: traumas, el mismo proceso de toracocentesis, neumonías de diversos orígenes, infarto pulmonar, tuberculosis, invasión neoplásica o derrames de larga evolución por insuficiencia cardíaca.

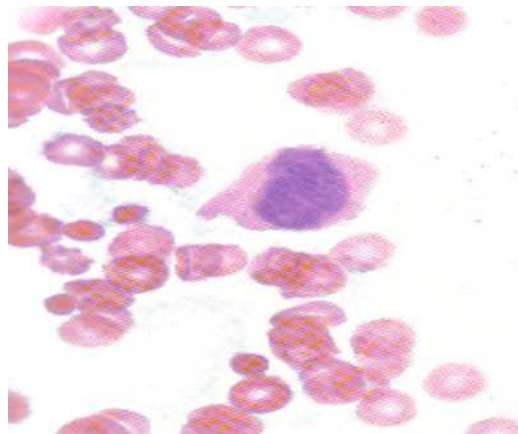


Figura 41: Hematíes en color rojo. Fuente: Íbid.

Con Papanicolaou se ven como discos de unas 7 micras de diámetro, de color

amarillo anaranjado, pero a menudo por la fijación se alisan transformándose en cáscaras vacías de variados colores verdosos.

Linfocitos

En casi todos los derrames pleurales se observan algunos linfocitos (Figura 42), los exudados con predominio de estos se deben a tuberculosis con invasión o no de la pleura, infecciones virales, carcinoma broncogénico u otras neoplasias.

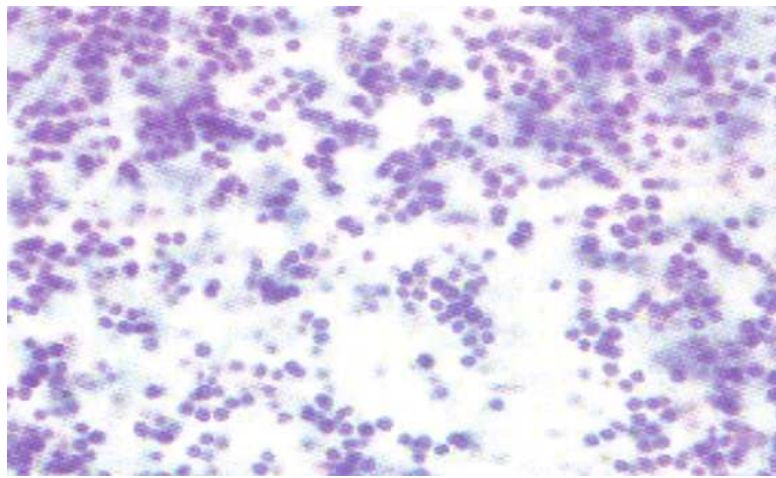


Figura 42: Linfocitos. Fuente: [Fariña and Rodríguez, 1996].

Con Papanicolaou la mayoría son células de citoplasma escaso o indistinguible, el núcleo es redondeado y ocupa toda la célula, su cromatina es densa y no se ven los nucleolos. Los linfocitos maduros son de un tamaño un poco más grande, citoplasma evidente, núcleo a veces hendido, de contorno más irregular e incluso nucleolo. En los derrames linfocitarios benignos los linfocitos no presentan mitosis ni necrosis.

Trastornos No Neoplásicos

Los derrames que tienen gran cantidad de neutrófilos se deben a infecciones bacterianas como la tuberculosis pulmonar en fase primaria. Ante la presencia de linfocitos con escasas células mesoteliales es difícil distinguir un

trastorno no neoplásico de un linfoma o leucemia.

Las efusiones de tipo eosinofílico, es decir con un 10 % de eosinófilos o más entre leucocitos, suelen ser idiopáticos pero pueden asociarse con traumatismos, infecciones e infartos.

En caso de artritis reumatoide y afección de la pleura, se visualizan células amorfas con histiocitos, células gigantes multinucleadas y pocas células mesoteliales.

Efusiones Malignas

Las células neoplásicas malignas no siempre son más grandes, pleomórficas e hipercromáticas que las mesoteliales. Algunos signos de naturaleza maligna son: aumento importante en el tamaño nuclear, elevada relación núcleo:citoplasma, irregularidad en el contorno nuclear, alteraciones en la textura de la cromatina o nucleolos muy llamativos.

Aunque la mayor parte de los tumores metastáticos en los derrames pleurales forman estructuras tridimensionales complejas como esferas de proliferación, papilas o acinos, algunos tipos de tumores (linfoma, melanoma, adenocarcinoma de origen gástrico o mamario) aparecen como células sueltas.

3.4.4. CITOPATOLOGÍA DE LOS DERRAMES PLEURALES CON CÉLULAS MALIGNAS

Los órganos en donde se originan habitualmente la mayoría de las Neoplasias son el pulmón en el hombre y la mama en la mujer. El diagnóstico maligno que se presenta con más frecuencia es el de Adenocarcinoma. Son más frecuentes los derrames en el sexo femenino.

El adenocarcinoma es el tipo más común de células neoplásicas encontradas en fluidos serosos, especialmente en derrames pleurales, y con experiencia es posible sugerir exactamente el sitio primario de la neoplasia, especialmente si se tiene información acerca del sexo y el trasfondo clínico del paciente.

Adenocarcinoma

Es un carcinoma que tiene su origen en células que constituyen el revestimiento interno de las glándulas de secreción externa. La mayoría de células adenocarcinoma en fluidos serosos tienen origen neoplásico en seno, ovario o pulmón (Figura 43).

Principales Características del Adenocarcinoma

- Contornos celulares suavizados.
- Grupos cohesivos compuestos de grandes células excéntricas.
- Núcleos con apariencia maligna.
- Nucleolos prominentes.
- Citoplasmas vacuolados.

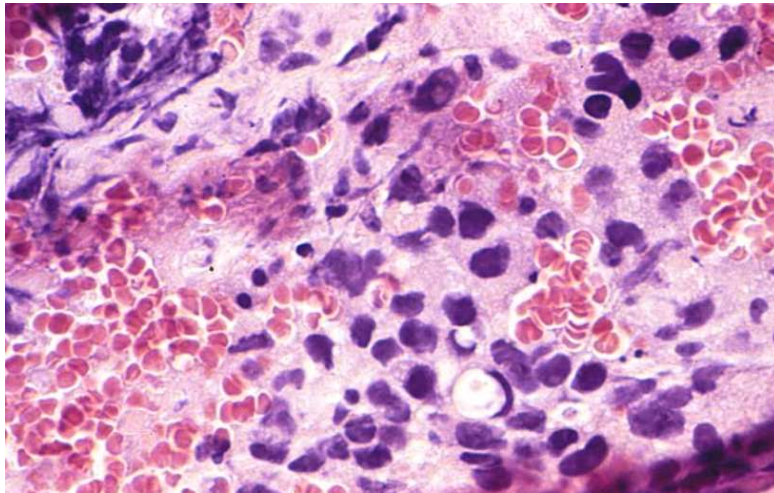


Figura 43: Células de adenocarcinoma. Fuente: [Fariña and Rodríguez, 1996].

El adenocarcinoma puede estar formado por una célula o puede formar racimos compuestos de pocas células o muchos fragmentos papilares o esferoides conformados por cientos de células.

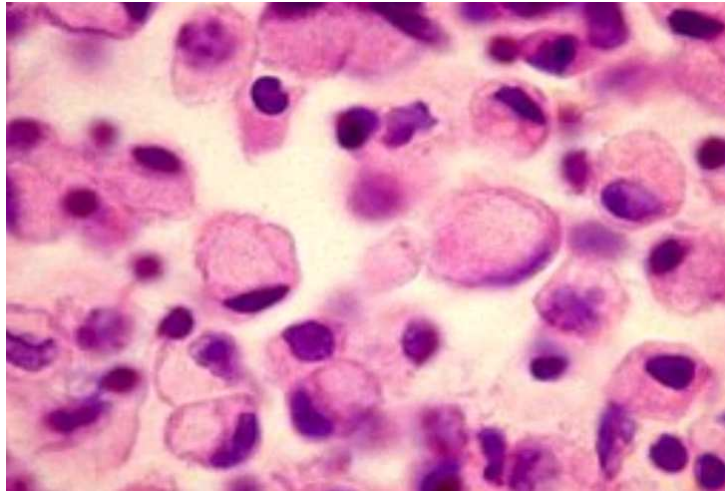


Figura 44: Células Vacuoladas. Fuente: [Fariña and Rodríguez, 1996].

Vacuolización de Células

La vacuolización citoplásmica es común en nidos de células y puede ser vista en adenocarcinomas de todos los sitios primarios, particularmente en carcinomas de ovario y pulmón [Mattison et al., 1997, p.591]. En células aisladas del adenocarcinoma gástrico es muy frecuente el tipo anillo de sello, con su núcleo desplazado por vacuolas de considerable tamaño (ver Figura 44).

Carcinoma de Pulmón

Cerca del 40% de los derrames pleurales contienen células cancerinas, estos derrames originados por carcinomas de pulmón están constituidos por grupos grandes de células vacuoladas y rara vez se observan tumores pulmonares de células pequeñas (Figura 45).

Grandes Células

El núcleo tiene nucleolos y cromatina gruesa y puede no haber aumento de la relación núcleo/citoplasma.

Células malignas que aparecen solas y que aparecen en clusters. Muchas células muestran varios degeneres de vacuolación. Muchos de los diferentes

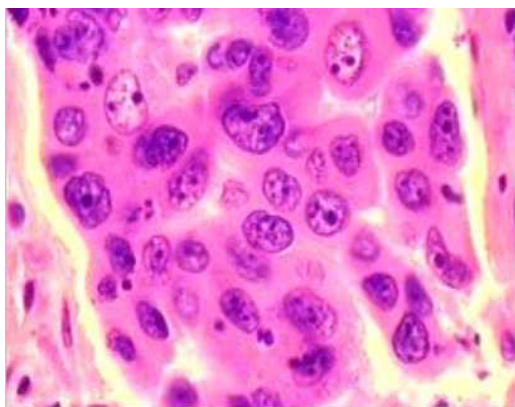


Figura 45: Carcinoma de Pulmón. Fuente: [Fariña and Rodríguez, 1996].

ejemplos de adenocarcinoma de pulmón en fluidos serosos provienen del carcinoma bronquio-alveolar.

Carcinoma de célula pequeña-pulmón

El tumor consiste de células pequeñas y bastante uniformes con semejanza a los linfocitos. Algunas son redondas. Tienen un borde de citoplasma tan delgado que casi no se ve.

Carcinoma de mama

En varios casos la primera manifestación de persistencia neoplasmática es una efusión pleural. Tal efusión puede no desarrollarse hasta una década o más después del tratamiento del neoplasma primario.

Contribuye a la citología de fluidos serosos por su alta incidencia y la frecuencia con la cual hace metástasis a membranas serosas, especialmente la pleura, causando derrame pleural [Mattison et al., 1997, p.603].

Células metastáticas de cáncer de seno se encuentran con más frecuencia en los derrames pleurales, aunque ocasionalmente son encontrados en líquidos pericardiales y ascíticos.

Los carcinomas de mama más frecuentes son los ductales infiltrantes, además de ser los más observados en líquido pleural.

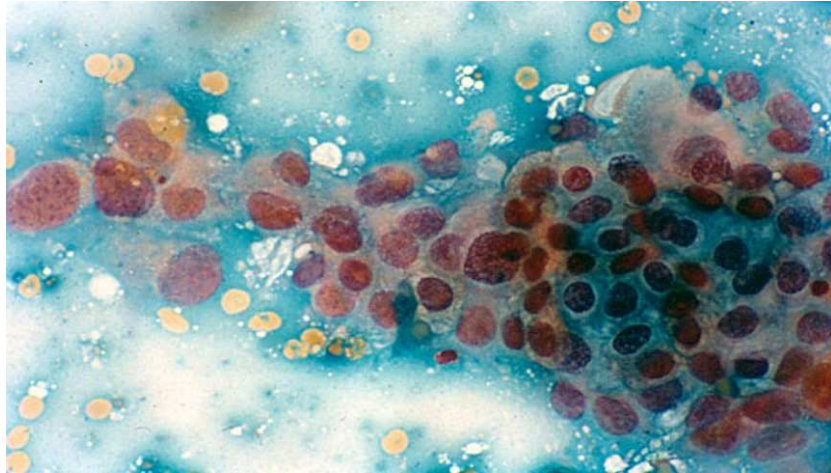


Figura 46: Carcinoma de mama. Fuente: Ibid.

Su presentación clásica tiende a ser compacta, densa, de bordes suavizados (proliferación de esferas). La cual puede representarse como una multiplicación de células libres (Figura 46).

Características de células de carcinoma de mama:

- Grandes esferas de células neoplásicas.
- Más frecuentes son los ductales infiltrantes.
- No se distinguen los caracteres celulares, sin embargo la periferia de la bola presenta los citoplasmas de las células y los núcleos hacia el interior.
- Aspecto cribiforme.
- Zonas vacías en medio de las esferas en donde se pueden encontrar células o espacios huecos redondeados.

Las células aisladas del derrame permiten estudiar los caracteres malignos celulares, además los citoplasmas pueden tener vacuolas y estas pueden contener mucina y glucógeno, núcleos en mitosis; las esferas pueden estar formadas

por células mesoteliales malignas o benignas o por células de adenocarcinomas de diversos orígenes.

Carcinoma de Ovario

Es frecuente causa de ascitis, en grado de menor causa efusión pleural, generalmente es fácil de identificar en fluidos serosos, no necesariamente como ovario, pero sí como adenocarcinoma.

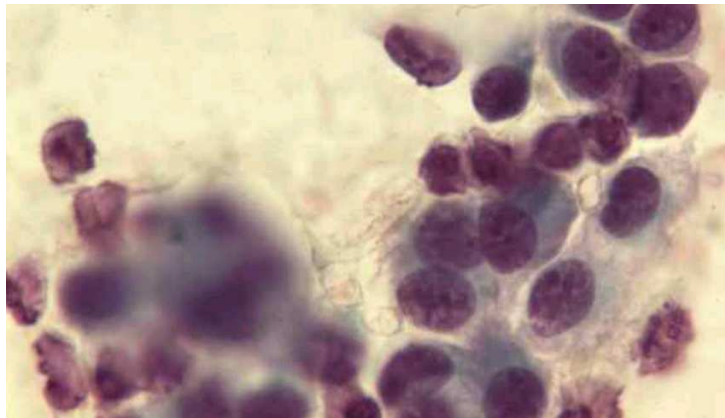


Figura 47: Carcinoma de Ovario. Fuente: Ibid.

Sus características son:

- Número alto de células neoplásicas.
- Las células presentan clusters papilares mezclados con numerosas células, muchas de estas vacuoladas.
- Los clusters papilares pueden contener cuerpos psammoma.
- Las células individuales muestran la característica de adenocarcinoma.

Son absolutamente grandes, uniformes, y frecuentemente hipervacuoladas. Tal vacuolación no necesariamente significa que el citoplasma contenga mucina. Esto puede ser debido al glucógeno o un cambio degenerativo.

4. ADQUISICIÓN

La adquisición de imágenes digitales se llevó a cabo en el departamento de patología de la Universidad Industrial de Santander. Allí fueron adquiridas láminas de citología para hacer la respectiva toma de imágenes. Dentro de los implementos utilizados para llevar a cabo la adquisición tenemos: cámara fotográfica digital Olympus C7070 wide zoom, microscopio trinocular Olympus CX31. Las imágenes fueron almacenadas en un computador portátil utilizando un software para archivarlas bajo el formato JPEG con resolución de 1600x1200 píxeles, para cada imagen se documentaba el caso según el diagnóstico patológico.



Figura 48: Equipo de adquisición.

Durante la toma se adquirieron un total de 47 imágenes digitales que presentan el diagnóstico de malignidad. Las imágenes de células benignas fueron

proporcionadas por el Grupo de Investigación de Ingeniería Biomédica de la UIS GIIB.

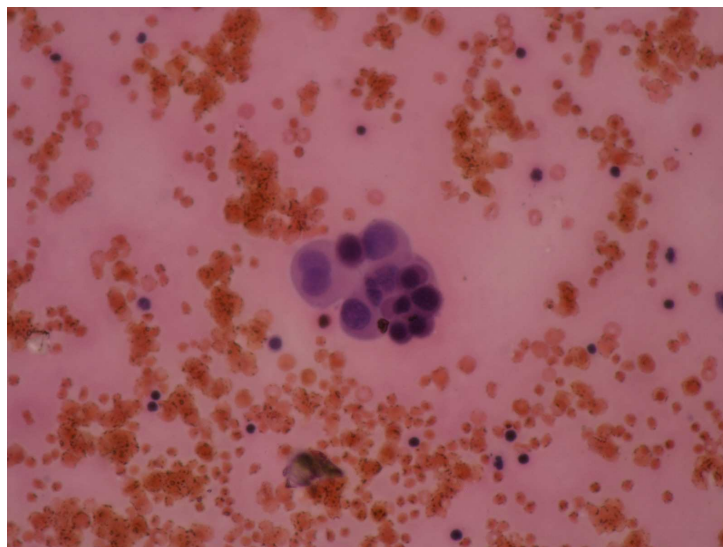


Figura 49: Imágen digital con diagnóstico de malignidad.

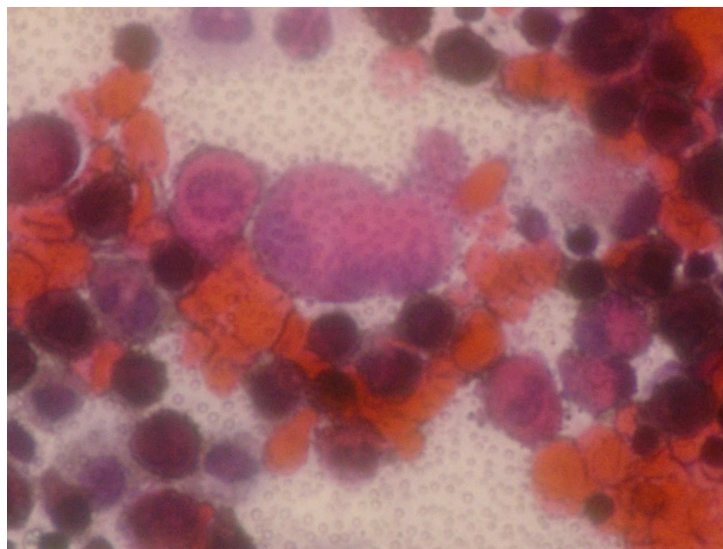


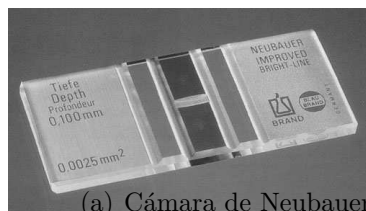
Figura 50: Imágen digital con diagnóstico negativo.

4.1. RELACIÓN PÍXEL-MICRA

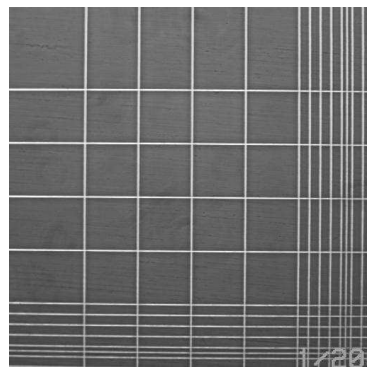
La imagen que se obtiene posee 1200 filas por 1600 columnas, pero no se contaba con una medida que permitiera conocer el tamaño del píxel, para calcular las dimensiones de las células. Con el fin de determinar la medida de un píxel, asignadas en micras (μm), se utilizó una cámara de Neubauer [Reina, 2003].

Se adquirieron fotografías de los cuadros interiores, puesto que la medida de sus lados es conocida. Para la cámara de Neubauer utilizada, el lado del cuadro interior más pequeño tiene una medida de $62.5 \mu m$ y mediante técnicas de detección de bordes y aproximaciones estadísticas se encontró que cada lado tenía 247 píxeles; si se utilizaba: un objetivo de 40x en el microscopio, un zoom óptico de 4x para la cámara fotográfica y un lente de acople Olympus C5060.

Estas condiciones de aumento se utilizaron para todas las imágenes adquiridas. A partir de las relaciones mencionadas, se obtuvo que cada píxel tenía un lado de $0.2530 \mu m$. En la Figura 51.a se muestra la cámara de Neubauer.



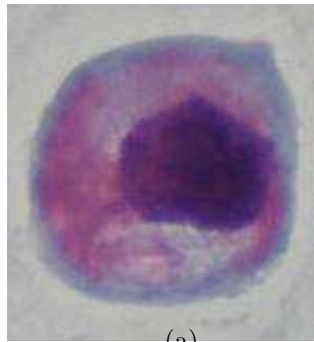
(a) Cámara de Neubauer



(b) Cuadros interiores

Figura 51: Cámara de Neubauer. Fuente: En línea <http://www.ub.es/biocel/wbc/tecnicas/contajecelular.htm>

El resultado de la adquisición fue satisfactorio, ya que no se presentó ningún inconveniente por resolución, contraste, etc. A continuación se muestran algunos resultados de la toma (Figura 52).



(a)



(b)

Figura 52: Resultados de toma satisfactoria.

5. PREPROCESAMIENTO

Obtenidas las imágenes digitales se procede a la segunda etapa del tratamiento digital de imágenes. En esta etapa de preprocesamiento lo que se quiere es mejorar la imagen siendo el núcleo y citoplasma objetos que puedan ser observados y trabajados posteriormente.

La imagen obtenida en la etapa anterior (Adquisición) está representada en el modelo de color RGB, para casi todas las imágenes se utilizó éste modelo, sólo en algunos casos especiales es utilizado el modelo HSL.

A continuación se muestra una imagen y sus componentes de color RGB.

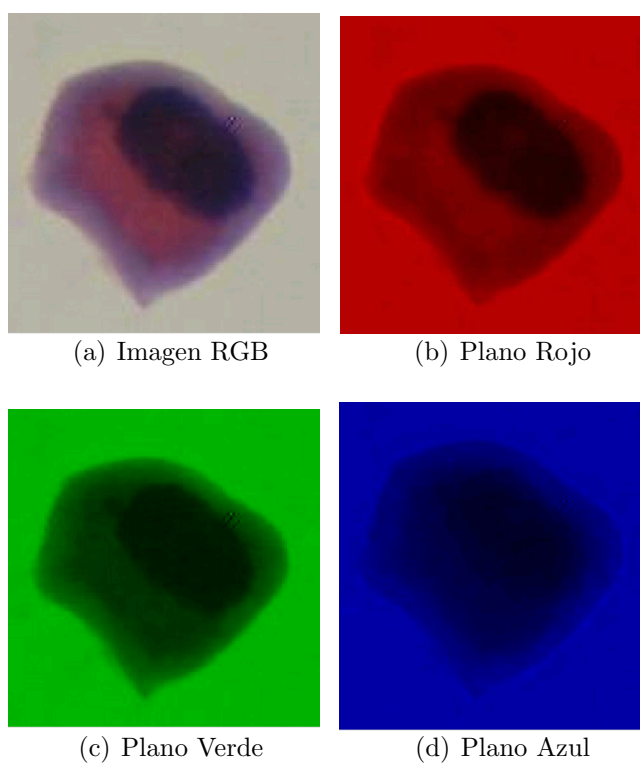


Figura 53: Imagen con sus componentes de color RGB.

5.1. FILTRADO

Para mejorar la imagen solo se utilizaron los filtros Circular, Gauss y Rotacional, ya que haciendo varias pruebas brindaron mejores resultados sobre las imágenes. Cada filtro cuenta con sus respectivos parámetros de configuración.

5.1.1. FILTRO CIRCULAR

En este tipo de filtro el resultado depende del tamaño del filtro que a su vez está determinado por la longitud del radio (Figura 54).

Parámetros: Tamaño del filtro.

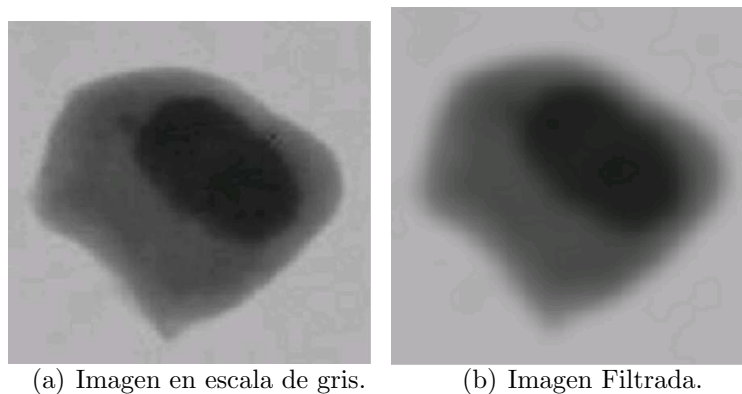


Figura 54: Aplicación filtro circular.

5.1.2. FILTRO GAUSS

En este tipo de filtro el resultado depende del tamaño de la matriz y de la desviación estándar utilizada que también es llamada sigma. Un filtro de tamaño tres y sigma 0.5 (Figura 55) se define de la siguiente manera:

$$\begin{bmatrix} 0,0113 & 0,0830 & 0,0113 \\ 0,0838 & 0,6193 & 0,0838 \\ 0,0113 & 0,0838 & 0,0113 \end{bmatrix} \quad (34)$$

Parámetros: Tamaño del filtro y Sigma del filtro.

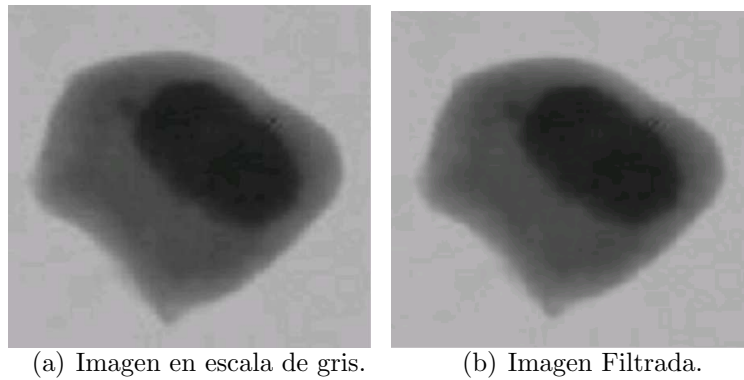


Figura 55: Aplicación filtro gauss.

5.1.3. FILTRO ROTACIONAL

Parámetros: Tamaño del filtro, Sigma del filtro, Longitud del filtro y Ángulo del filtro (Figura 56).

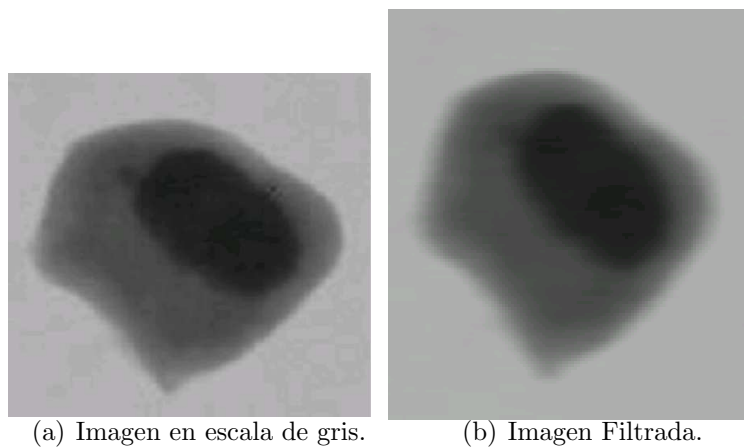


Figura 56: Aplicación filtro rotacional.

5.2. AJUSTE DE IMAGEN

Posterior al filtrado se mejora la imagen mediante un procedimiento encargado de ajustar los niveles de gris, alcanzando niveles más oscuros en regiones

ídem y más claros en regiones claras.

Como puede apreciarse en la Figura 57 la imagen ha mejorado notablemente, los detalles de núcleo y citoplasma son claros, fácilmente diferenciables.

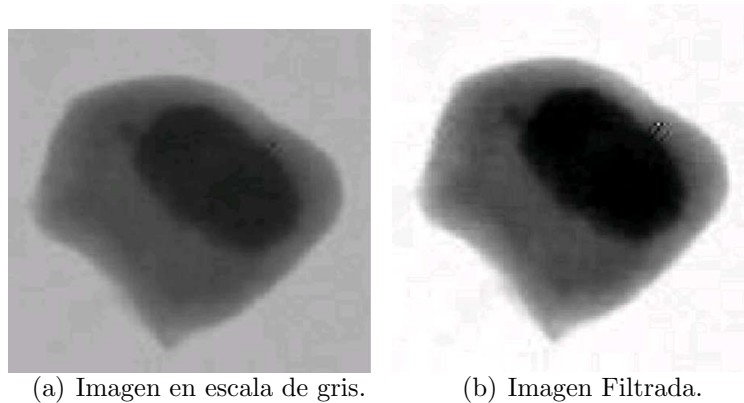


Figura 57: Aplicación de ajuste.

5.3. CONVERSIÓN AL MODELO IHSL

En el modelo RGB es imposible separar la información cromática de la lumínica, por tanto se hace difícil segmentar los objetos allí localizados.

Actualmente existen diversos modelos que permiten ésta separación, uno de ellos es el modelo IHSL. Los planos de color RGB se convierten en tres planos con diferente información respecto a la imagen. La transformación puede implementarse a través de la función `rgb2hsy.m` de Matlab ©.

El plano H (Hue en inglés), representa los matices distribuidos a lo largo del campo de la imagen. La Saturación, representa las variaciones que sufre el matiz en términos de la cantidad de color disuelto en un fondo blanco, pudiéndose observar cambios en la textura nuclear y citoplásmica que generan indicios a posibles variaciones neoplásicas.

El plano de la Luminancia representa el nivel de luz o escala de grises en la imagen, importante en la determinación del grado hiper Cromático en núcleo y citoplasma, debido a la relación directa entre éste y los bajos niveles de

luminosidad. La Figura 58 muestra una imagen convertida al modelo IHSL en cada uno de sus planos.

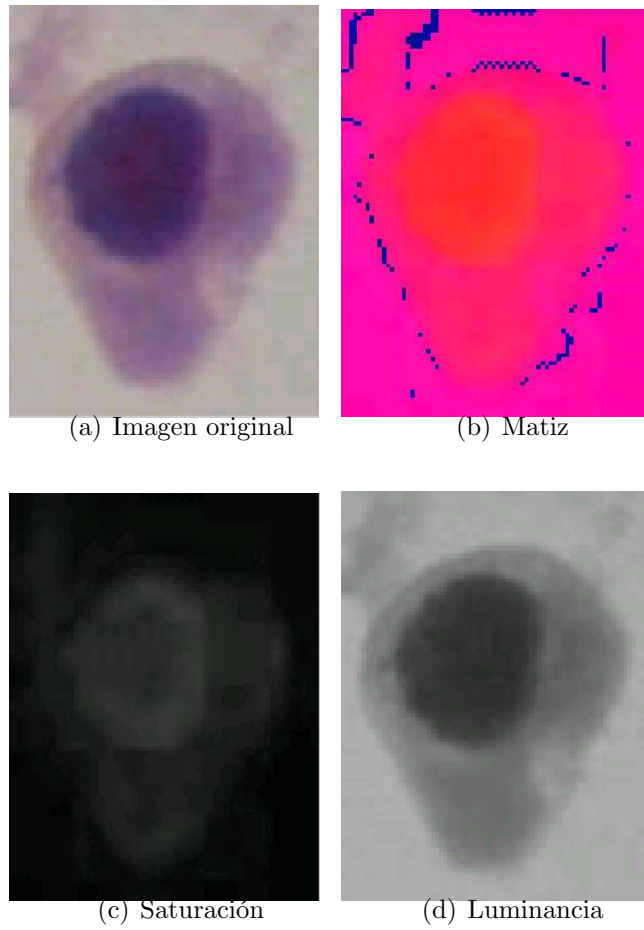


Figura 58: Imagen con sus componentes HSL.

6. SEGMENTACIÓN

Observando los resultados de la etapa anterior se decide utilizar la umbralización como método de segmentación. Los siguientes métodos fueron probados: **Umbralización por el algoritmo de Pun-Kapur** Los resultados no fueron los esperados.

Umbralización por el método iterativo Los resultados no fueron los esperados.

Umbralización por el método automático de matlab Los resultados no fueron los esperados.

Umbralización por el método triangular Los resultados fueron mejores en comparación con anteriores. Después de probar con distintos tipos de imágenes se observó que este método fallaba, aunque no fue descartado de lleno se decide continuar probando con otros algoritmos, hasta conseguir óptimos resultados con el método crecimiento de regiones por agregación de píxeles.

6.1. CRECIMIENTO DE REGIONES

Método con mejores resultados para la mayoría de núcleos y citoplasmas.

Consiste en seleccionar una semilla y a partir de ésta, con diferentes umbrales, agregar píxeles por cada umbral hasta hacer crecer totalmente la región. Observando que con la misma semilla es posible segmentar tanto núcleo como citoplasma.

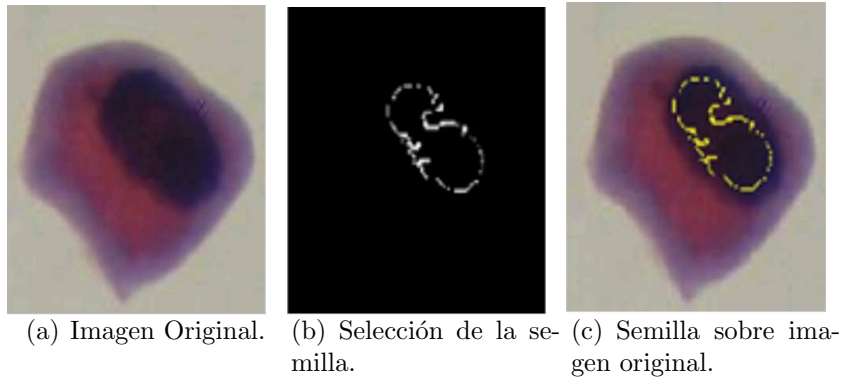


Figura 59: Crecimiento de regiones. Selección de semilla

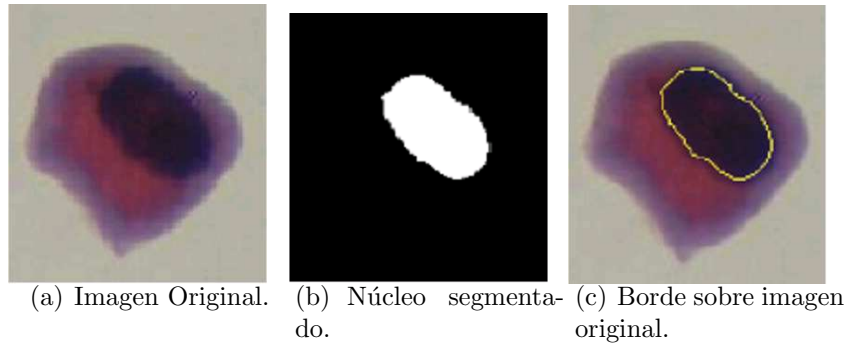


Figura 60: Crecimiento de regiones. Segmentación de núcleo.

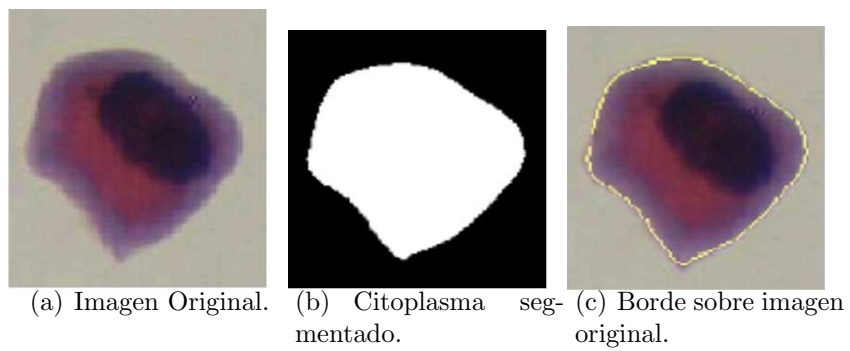


Figura 61: Crecimiento de regiones. Segmentación de citoplasma.

6.2. DISTANCIA VECTORIAL

Aunque los métodos implementados segmentaron adecuadamente, fallaron en casos de imágenes con adenocarcinoma donde se presentaban alteraciones relacionadas con presencia de nucleolos y/o vacuolización, incluso en células reactivas con similares atipicidades.

Implementándose un algoritmo que segmentara adecuadamente éste tipo de imágenes, tanto malignas como reactivas.

Cada píxel fue considerado como un vector R^3 en sistema RGB e IHSL. Se efectuaron pruebas de gradiente mediante el uso de la función Matlab *gradient* con el fin de encontrar variaciones en píxeles de zonas consideradas homogéneas y no homogéneas. Los resultados obtenidos experimentalmente, tanto en el sistema RGB como IHSL, no mostraron regularidad en la diferenciación de píxeles pertenecientes a regiones comunes dentro de la imagen, el mismo comportamiento se pudo determinar para píxeles de áreas disímiles, es decir, la diferencia entre magnitud gradiente para píxeles de áreas similares en muchas ocasiones era mayor que la calculada para píxeles de áreas disímiles, y viceversa, sobretodo en imágenes de células reactivas, no siendo el caso de aquellas con diagnóstico maligno, como puede observarse comparando las tablas 2 y 3.

En consecuencia, se realiza la misma prueba utilizando como parámetro de comparación la *Norma* del *vector-píxel* cuyas componentes son su sistema de color RGB o IHSL, observando que píxeles de regiones similares presentan valores pequeños en su diferencia de magnitud frente a aquellos localizados en regiones no compartidas, tanto para células malignas como reactivas. Esto permite diseñar un algoritmo de segmentación cuyo umbral de comparación es un valor límite de distancia.

El método consiste, básicamente, en definir un umbral de máxima distancia entre los *vectores-píxeles* a comparar. Se toma como principal *vector-píxel* el correspondiente al centroide de la semilla seleccionada, se compara con los *vectores-píxeles* vecinos directos del borde de imagen de la cual hace parte el principal *vector-píxel* y se integran a la imagen segmentada aquellos cuya distancia sea igual o menor al umbral establecido. El algoritmo se detiene cuando se agotan los *vectores-píxeles* habilitados por la condición, tal como

lo define Delgado et al., (2006, 70).

Tabla 2: Distancias vectoriales para Gradiente y sistema RGB en zonas comunes y no comunes de imágenes celulares con malignidad.

<i>Regiones comunes</i>		<i>Regiones no comunes</i>	
<i>Dist. Gradiente</i>	<i>Dist. RGB</i>	<i>Dist. Gradiente</i>	<i>Dist. RGB</i>
0,2584	0,0130	7,2860	0,3784
0,5517	0,0157	6,0553	0,2374
0,2058	0,0096	8,3292	0,4681
0,3860	0,0200	6,5614	0,3352
0,7919	0,0288	3,7206	0,1241
0.2711	0.0263	6.6936	0.2340
0	0,0068	4,8541	0,1143
0	0,0068	6,3058	0,1901
1,0006	0,0452	7,6218	0,2812
0,2334	0,0130	6,2879	0,2553
0,6266	0,0215	5,6403	0,1861
0	0,0136	4,1210	0,1834
0,0695	0,0096	4,9908	0,2231
0	0,0156	3,8538	0,1261
0,3705	0,0130	3,0242	0,0847
1,9860	0,0386	6,0792	0,2139
0,3262	0,0196	3,9720	0,1226
0,5400	0,0346	5,8100	0,2024
0,1778	0,0251	5,2647	0,1978
0,2334	0,0078	5,2634	0,1703
1,5980	0,0642	7,2249	0,2698
1,8285	0,0609	4,0040	0,1287
2,6922	0,1027	3,4523	0,1182
0,8221	0,0147	4,5518	0,1784
0,4995	0,0118	4,2948	0,1051
1,1271	0,0600	4,7623	0,1460
0,8844	0,0445	7,1133	0,4378
1,2637	0,0204	9,7956	0,4886
2,5053	0,0663	7,0465	0,2631
1,3413	0,0400	7,8206	0,4209

Tabla 3: Distancias vectoriales para Gradiente y sistema RGB en zonas comunes y no comunes de imágenes celulares no malignas.

<i>Regiones comunes</i>		<i>Regiones no comunes</i>	
<i>Dist. Gradiente</i>	<i>Dist. RGB</i>	<i>Dist. Gradiente</i>	<i>Dist. RGB</i>
0,8736	0,0326	0,6035	0,0795
1,5497	0,0254	0,8082	0,0993
2,1959	0,0319	1,4120	0,0801
1,7502	0,0561	2,7898	0,0703
0,5481	0,0257	3,1347	0,1390
1,2584	0,0374	0,7838	0,1126
2,4430	0,0654	2,3269	0,1110
1,8768	0,0370	1,5911	0,1041
1,0811	0,0316	0,6745	0,0969
1,5718	0,0232	1,2089	0,1271
0,5092	0,0180	2,1532	0,1849
0,5237	0,0200	1,1422	0,1154
1,0558	0,0346	1,3928	0,1167
0,8591	0,0370	0,9704	1,1838
0,2983	0,0417	0,1642	0,1377
0,3719	0,0370	0,8315	0,1840
0,7401	0,0254	1,0667	0,0980
1,8232	0,0248	0,3154	0,0839
2,0747	0,0277	0,8246	0,1406
0,4744	0,0239	2,3795	0,0915
3,2365	0,0411	2,8784	0,1166
1,2615	0,0419	0,6735	0,0961
2,4114	0,1014	0,1808	0,1382
1,4498	0,0200	0,4240	0,1122
0,2216	0,0845	0,1876	0,1542
1,6940	0,0288	1,5246	0,1111
0,8265	0,0166	2,9163	0,1683
0,7247	0,0378	1,9898	0,0953
0,4946	0,0662	3,5893	0,1623
1,4236	0,0171	3,2192	0,1090
1,8681	0,0411	0,3872	0,1500

Es importante señalar que a pesar de obtener óptimos resultados en imágenes

difíciles de segmentar por métodos anteriores, valdría la pena -en experiencias futuras- variar el presente algoritmo cambiando el *vector-píxel* correspondiente al centroide de la semilla seleccionada inicialmente, por un vector de componentes $\hat{R}\hat{G}\hat{B}$ o $\hat{H}\hat{S}\hat{L}$, según el sistema elegido, que luego de anexar nuevos píxeles variara en función del promedio, teniendo en cuenta la totalidad de píxeles que conforman la nueva imagen segmentada.

6.2.1. PASOS PARA LA SEGMENTACIÓN POR DISTANCIA VECTORIAL

Los pasos son los siguientes:

- Mejoramiento de la imagen mediante la aplicación de un filtro, ya sea Rotacional, Circular o Gaussiano.
- Ajuste de la imagen (ver sección 4.2).
- Selección de una zona como semilla mediante métodos de umbralización o con la ayuda del puntero del mouse.
- Definición del umbral o distancia vectorial.
- Selección del sistema o modelo de color, ya sea RGB o IHSL.
- Si la segmentación es óptima, pero aún no cumple con el total de imagen deseado, se almacena en memoria para luego ser sumado con otras segmentaciones adecuadas de la misma imagen.
- Vuelven a ejecutarse los pasos desde el principio y se van sumando las mejores imágenes para finalmente totalizar y aplicar morfología matemática con el fin de pulir la imagen obtenida.

6.2.2. RESULTADOS OBTENIDOS CON SEGMENTACIÓN POR DISTANCIA VECTORIAL

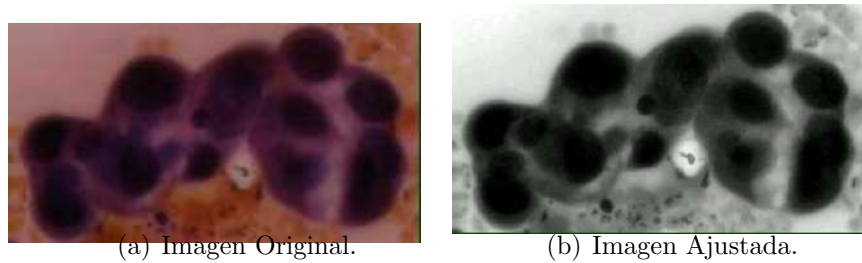


Figura 62: Imagen a segmentar por método de Distancia Vectorial.

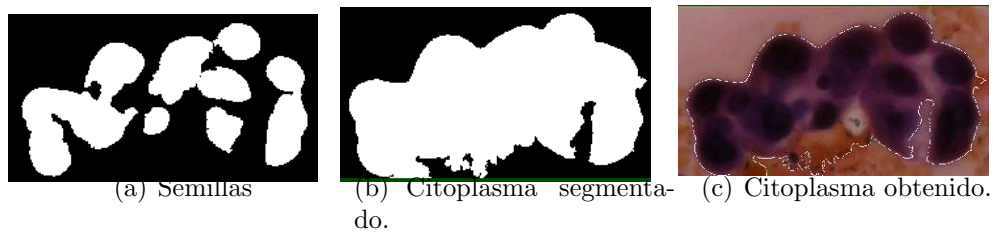


Figura 63: Segmentación obtenida por método de Distancia Vectorial en RGB con umbral=0.393446.

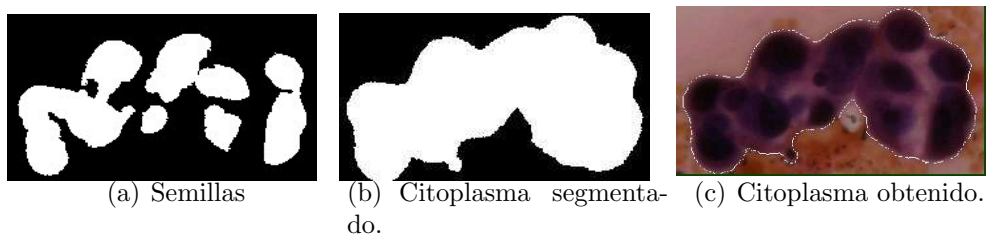


Figura 64: Segmentación obtenida por método de Distancia Vectorial en IHSL con umbral=0.3534356.

Como puede ser observado en las figuras anteriores se obtienen mejores resultados aplicando el método de segmentación sobre el modelo IHSL, sobretodo cuando las células presentan irregularidades cromáticas y texturales

de etiología neoplásica. Es el caso del citoplasma obtenido en la Figura 64, comparado con el obtenido mediante sistema RGB de la Figura 63.

7. DESCRIPCIÓN

Obtenidos núcleos y citoplasmas durante la etapa de Segmentación, tanto para células malignas como benignas, se procedió a extraer los parámetros que permitieron caracterizar cada célula mediante la utilización de imágenes binarizadas.

7.1. DESCRIPCIÓN DE CITOPLASMA

7.1.1. MEDIDAS MORFOLÓGICAS

Están relacionadas con la forma y distribución espacial del citoplasma.

Área: Se calcula utilizando el número total de píxeles en blanco contenidos en la imagen binaria que caracteriza al citoplasma. Para determinar la medida en micras cuadradas (μm^2) se multiplica por el factor de conversión descrito en el capítulo de Adquisición.

$$Area = \sum pixels_{blancos} \quad (35)$$

Perímetro: Se calcula mediante la suma de las distancias entre píxeles blancos que se encuentran en el borde de la imagen binaria del citoplasma (Figura 65). Para determinar la medida en micras (μm) se multiplica por el factor de conversión descrito en el capítulo de adquisición.

$$Perimetro = \sum_{i=1}^n \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} \quad (36)$$

Siendo a y b píxeles adyacentes y n el número total de píxeles en el borde.

Eje mayor y eje menor: Longitud o diámetro de los ejes mayor y menor de la elipse con el segundo momento igual a la región del citoplasma (Figura 66).

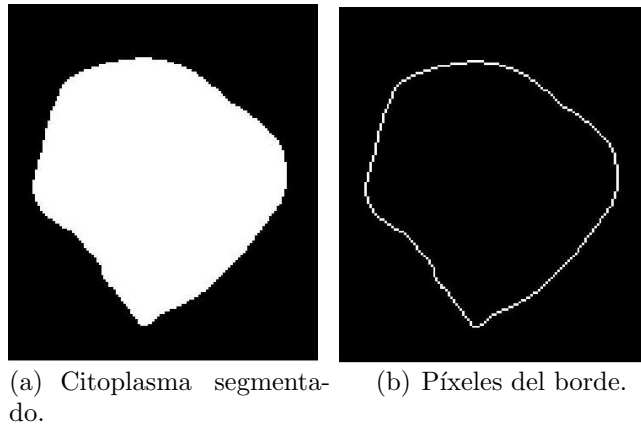


Figura 65: Descripción de Citoplasma: área y perímetro

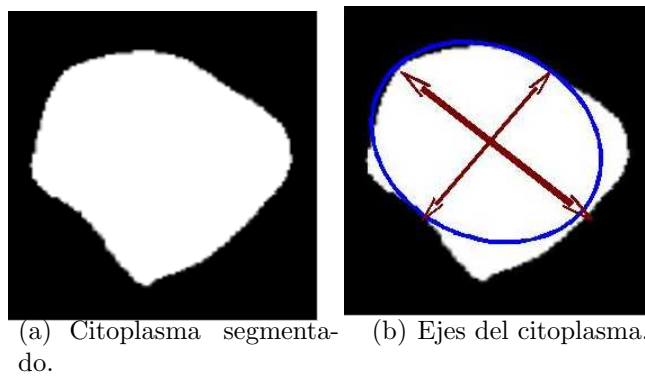


Figura 66: Descripción de Citoplasma: Eje mayor y menor

Relación Núcleo/Citoplasma: Es calculada como la proporción entre el tamaño del núcleo y el tamaño del citoplasma.



Figura 67: Distancia centroide citoplásmico y nuclear.

Varianza (forma): Es la variabilidad del vector que contiene la distancia de los puntos del contorno del citoplasma al centro del mismo, suponiendo un valor de varianza igual a cero para un círculo perfecto completamente regular [Delgado et al., 2006, p.75].

$$S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (37)$$

7.2. DESCRIPCIÓN DEL NÚCLEO

7.2.1. MEDIDAS MORFOLÓGICAS

Están relacionadas con la forma y la ubicación espacial del núcleo dentro de la célula.

Área: Se calcula por el número de píxeles que están en blanco dentro de la imagen binaria que contiene el núcleo.

$$Area = \sum pixels_{blancos} \quad (38)$$

Perímetro: Suma de las distancias entre píxeles blancos que se encuentran en el borde de la imagen binaria del núcleo.

$$Perimetro = \sum_{i=1}^n \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} \quad (39)$$

Siendo a y b píxeles adyacentes y n el número de píxeles que se encuentran en el borde del núcleo.

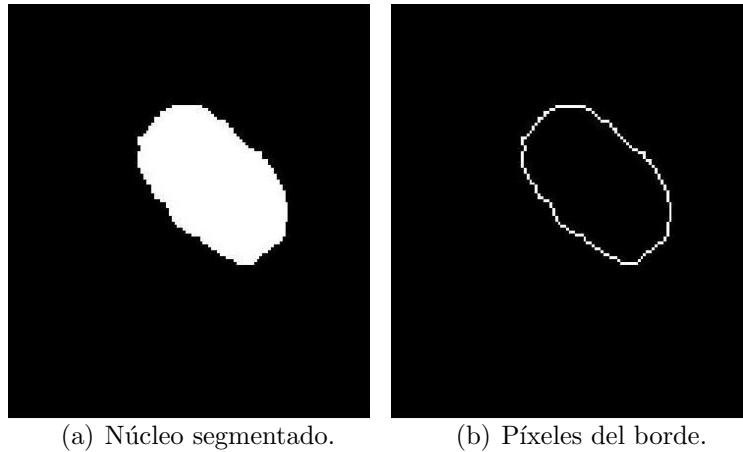


Figura 68: Descripción de Núcleo: área y perímetro.

Excentricidad o Varianza Distancia Núcleo Citoplasma

Varianza: Es la variabilidad del vector que contiene la distancia de los puntos del contorno nuclear al centro del mismos, suponiendo un valor de varianza igual a cero para un círculo perfecto completamente regular.

$$S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (40)$$

Circularidad: Se define por:

$$C = \frac{P^2}{4\pi A} \quad (41)$$

Es la relación que existe entre el cuadrado del perímetro del núcleo (píxeles adyacentes del borde) y el área del núcleo (píxeles blancos dentro de la imagen binaria que contiene el núcleo). Para un círculo perfecto el valor será uno.

7.2.2. MEDIDAS DE TEXTURA

En ésta categoría se agrupan las características relacionadas con la coloración del campo celular.

Hipercromatismo. Se pudieron observar altos grados de tinción en el plano de la luminancia por lo que se decide utilizar los descriptores de la matriz de co-ocurrencia en el plano de la luminancia con el fin de percibir cualquier tipo de variabilidad por muy sutil que se presente.

Pérdida de polaridad. Esta situación era provocada por cambios en la textura del plano Matiz, para lo cual se decide utilizar los descriptores de la matriz de co-ocurrencia en éste plano, explicados en el marco teórico. Se adicionaron la entropía y la desviación estándar de los niveles en el plano.

Distribución de cromatina nuclear. Como la saturación es un indicador de la cantidad de color, se utilizaron los mismos descriptores de textura mencionados para la pérdida de polaridad pero calculados en el plano de la saturación con el objetivo de modelar los cambios de la cromatina nuclear.

En la descripción nuclear no es utilizado el plano del Matiz, ya que los núcleos muestran el mismo color en toda su superficie debido a la tinción usada para efectos de la lámina.

En un comienzo se utilizaron descriptores de matriz de co-ocurrencia sólo para análisis de textura nuclear pero debido a malos resultados en la identificación de malignidad durante la ejecución del modelo computacional, además de observar rotundas mejorías al incluir el análisis de la textura citoplásmica, se decidió modelar también las variaciones texturales citoplásmicas tanto en el plano de la saturación como de la luminancia.

7.3. MODELO ENTIDAD-RELACIÓN

Las descripciones de las imágenes celulares, indispensables para el entrenamiento de la Red Neuronal con el fin de caracterizar y determinar posibles parámetros de malignidad, se almacenaron en una base de datos utilizando el motor de libre distribución Mysql, configurando las siguientes entidades:

- **Celula:** Contiene el identificador de la clase celular, el tipo celular (individual/racimo) y el diagnóstico (malignidad/benignidad). La relación con la tabla Nucleo y Citoplasma asigna los descriptores respectivos a la célula parametrizada.

- **Núcleo:** Almacena la información de las medidas extraídas del núcleo. Su relación con la entidad Celula le permite diferenciar al sistema las caracterizaciones malignas y benignas necesarias durante el entrenamiento de la red neuronal.
- **Citoplasma:** Contiene los parámetros medidos del citoplasma celular. Su relación con la entidad Celula permite diferenciar caracterizaciones de benignidad y malignidad necesarias en la etapa de entrenamiento neuronal.

La Figura 69 muestra el diagrama Entidad-Relación, donde las relaciones Celula - Núcleo y Celula - Citoplasma son uno a uno.

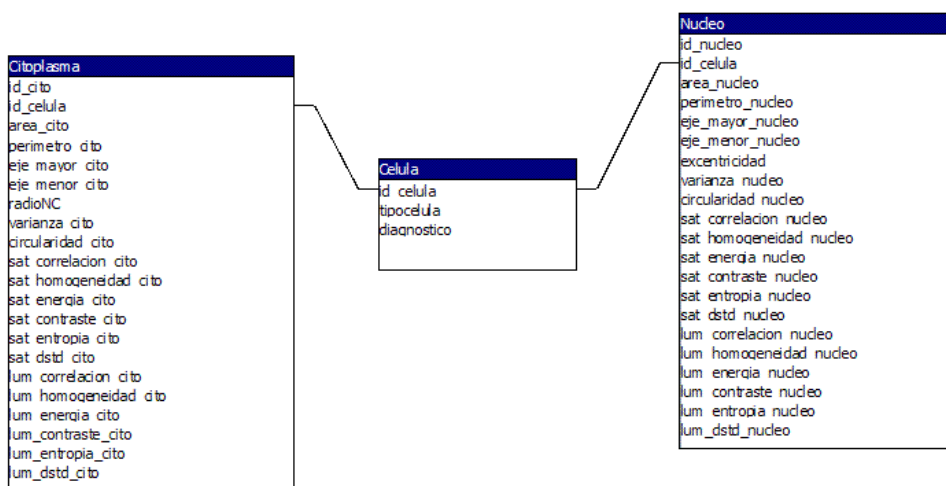


Figura 69: Modelo Entidad - Relación

8. INTERPRETACIÓN Y RECONOCIMIENTO

La finalidad de esta etapa es poder clasificar las células como benignas o malignas según los parámetros almacenados en la base de datos para éstos grupos celulares.

La Tabla 4 muestra el número de imágenes de células individuales que se utilizaron para la creación del patrón de clasificación de células malignas y benignas.

<i>Clasificación</i>	<i>Número de Imágenes</i>	<i>Número de Células</i>
Negativas	45	61
Positivas	57	96

Tabla 4: Número y tipo de imágenes celulares caracterizadas.

Con el propósito de establecer un patrón que lleve a cabo este proceso, se buscó identificar las posibles diferencias entre estas dos clases utilizando gráficas de dispersión.

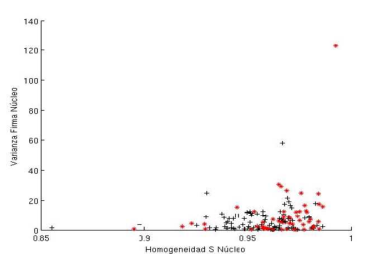
Se construyeron gráficas 2D y 3D con el fin hacer comparaciones entre 2 y 3 parámetros para observar cómo era su dispersión.

8.1. CARACTERÍSTICAS CELULARES

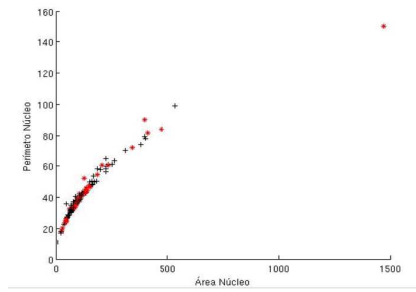
Se puede observar en las siguientes figuras la dispersión que existe entre la ubicación de Núcleos Benignos (rojo) y Malignos(N).

Posteriormente se ve la correlación existente entre las medidas, detallando de forma clara su comportamiento.

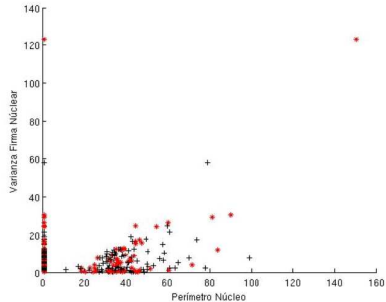
8.1.1. GRÁFICAS DE DISPERSIÓN



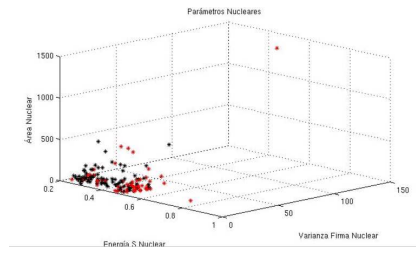
(a) Homogeneidad en el Plano de la Saturación - Varianza de la Firma Nuclear.



(b) Área Nuclear - Perímetro Nuclear.

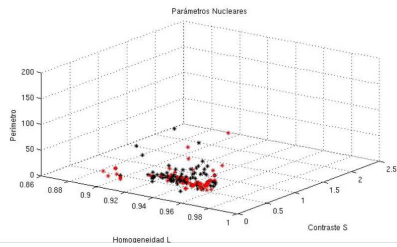


(c) Varianza de la Firma Nuclear - Perímetro Nuclear.

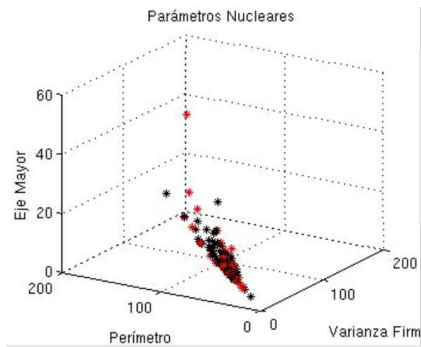


(d) Energía en el Plano Saturación.

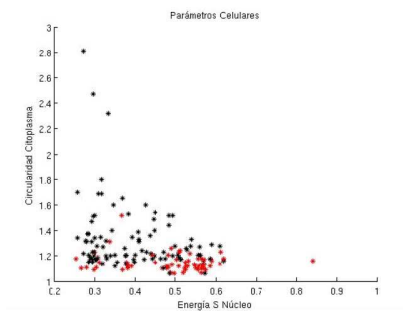
Figura 70: Gráficas de dispersión medidas morfológicas y plano saturación nuclear.



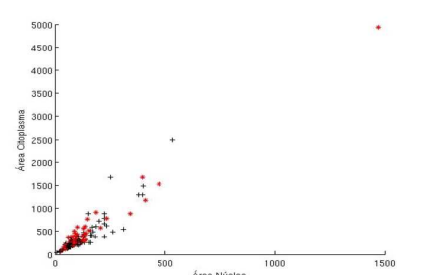
(a) Homogeneidad en el Plano de la Luminancia - Contraste en Saturación - Perímetro Nuclear.



(b) Varianza Firma Nuclear-Perímetro Nuclear-Diámetro Nuclear Mayor.



(c) Circularidad Citoplasma-Energía S Núcleo.



(d) Área Citoplasma-Área Núcleo.

Figura 71: Gráficas de dispersión Núcleo - Citoplasma.

8.1.2. INTERPRETACIÓN DE CARACTERÍSTICAS CELULARES SEGÚN GRÁFICOS DE DISPERSIÓN

Algunos parámetros no permitieron establecer claramente diferencias entre benignidad y malignidad. Resulta difícil separar linealmente éstas caracterizaciones dentro de las dos clases estudiadas. Por ejemplo, la Figura 70.b, que relaciona en el espacio bidimensional Área y Perímetro Nuclear, refleja la imposibilidad de separar linealmente puntos negros y rojos, malignidad y benignidad, respectivamente. Ésto último prueba la inexactitud del estudio citológico basado en la pericia visual del patólogo, en consecuencia las altas tasas de falsos positivos registrados en el marco conceptual del presente trabajo.

8.2. RED NEURONAL ARTIFICIAL DE CLASIFICACIÓN

Las anteriores gráficas de dispersión demostraron que el problema no podía ser solucionado usando métodos de separación lineal, caso típico de compuerta XOR que puede resolverse mediante una red neuronal de tipo retropropagación o Backpropagation, además, trabajos como el de Martínez (2007) han utilizado éste modelo frente a situaciones similares con óptimos resultados.

8.2.1. ARQUITECTURA Y ALGORITMO DE ENTRENAMIENTO

En la arquitectura backpropagation multicapa o de retropropagación, el error se propaga a medida que se avanza en cada una de las capas que conforman la red, por tal motivo se eligió una capa oculta; la capa de entrada estaría definida por las medidas extraídas a un elemento celular, mientras en la capa de salida se definió una neurona con función tangente hiperbólica en el intervalo $[-1,1]$ que indicaría la clase celular encontrada.

La Tabla 5 muestra la salida que tiene la Red Neuronal Artificial mientras la Figura 72 su arquitectura.

<i>Clasificación</i>	<i>Salida</i>
Negativas	-1
Positivas	1

Tabla 5: Salida de la RNA de clasificación

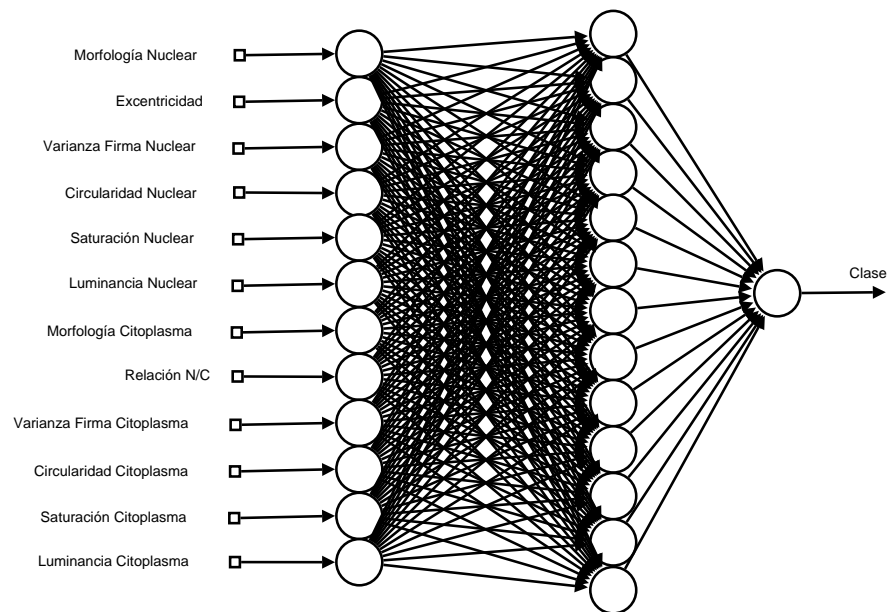


Figura 72: Arquitectura de la Red Neuronal Artificial.

Se definió un margen de error para la salida de $1 * 10^{-10}$. Para la función de entrenamiento se recurrió a la optimización por el método del Gradiente Conjugado Escalado. Finalmente se ejecutó el algoritmo de optimización y en la iteración 6649 se obtuvo la convergencia de la RNA tal como puede verse en la Figura 73.

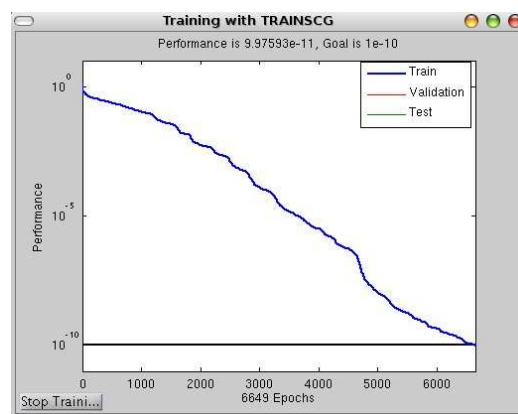


Figura 73: Convergencia del Algoritmo de Entrenamiento.

9. MODELO COMPUTACIONAL

Se utilizó el Proceso de desarrollo unificado de software como metodología para llevar a cabo la construcción del sistema [Jacobson et al., 2000], enlazándolo con la metodología del Tratamiento Digital de Imágenes.

A continuación se muestra el diagrama de casos de uso (Figura 74), posteriormente se describe el comportamiento del sistema mediante el flujo normal y el alterno (Tabla 6). (126-163).

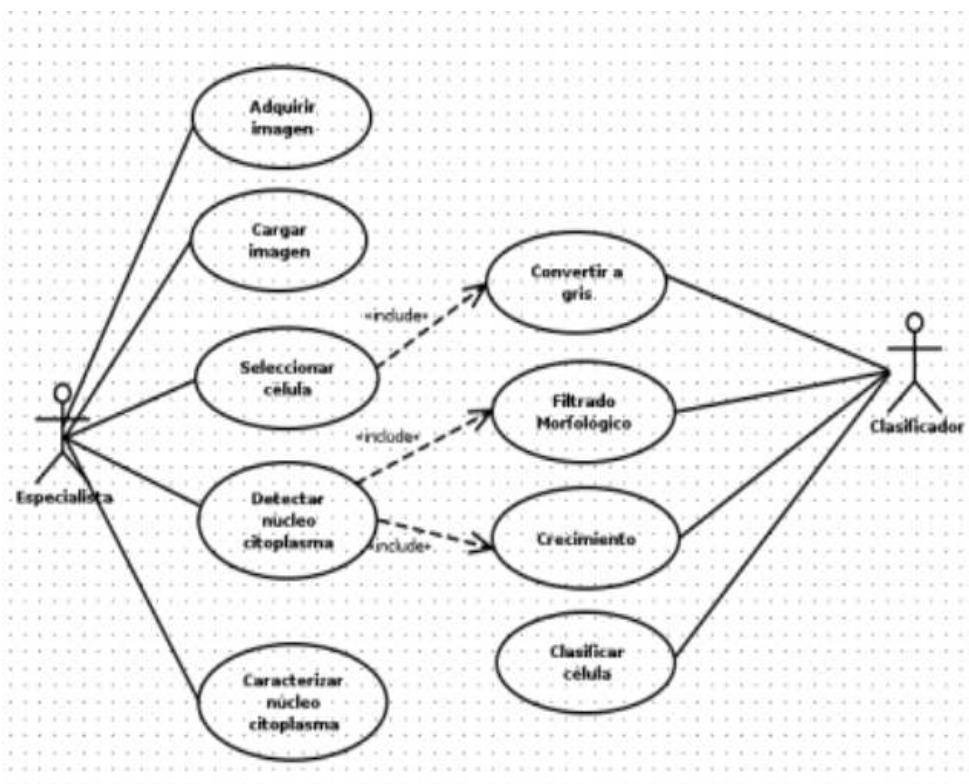


Figura 74: Diagrama de Casos de Uso

Tabla 6: Documento de descripción de los casos de uso.

Nombre:	Diseño e implementación de un sistema basado en Redes Neuronales Artificiales para la caracterización de células malignas en líquido pleural.
Autores:	Juan Francisco Remolina Caviedes, Adrián Fernando Rodríguez Vega.
Fecha:	Septiembre de 2007 .
Actores:	Especialista, Sistema Clasificador.
Precondiciones:	Precondiciones: El especialista debe contar con una lámina que contenga muestra de líquido pleural, una cámara fotográfica digital Olympus C7070 WZ, un microscopio trinocular Olympus CX31 y un acople Olympus C5060. El zoom óptico de la cámara debe estar ajustado en 4x y deshabilitado el digital. El objetivo a utilizar en el microscopio corresponde a 40x. La resolución de la imagen debe ser 1200x1600 píxeles y en formato jpeg. La aplicación debe ser ejecutada en la plataforma Matlab r2006a o superior.
Flujo Normal	<p>El especialista adquiere fotografías digitales de Células en líquido pleural.</p> <p>El especialista carga la imagen en la aplicación.</p> <p>El especialista selecciona una Célula para analizarla.</p> <p>El sistema clasificador convierte la Célula seleccionada a escala de grises.</p> <p>El especialista Selecciona el tipo de filtro para mejorar la imagen. El sistema clasificador procesa la imagen mediante un algoritmo que va de acorde con la opción que eligió el especialista. El especialista utiliza el botón de ajuste. El sistema clasificador utiliza un algoritmo para elevar el contraste en oscuros y blancos.</p> <p>El especialista fija una semilla sobre el núcleo ajustado. El sistema clasificador habilita un algoritmo de crecimiento de regiones. El especialista mediante una barra deslizante comienza a segmentar, por su parte el sistema clasificador recibe cada parámetro y utiliza el algoritmo para que en la imagen se vean los resultados de la segmentación.</p> <p>El especialista introduce parámetros en la opción de morfología matemática y por medio de barras deslizantes se van haciendo cierre y apertura. El sistema clasificador ejecutará el filtro basado en morfología matemática y los resultados se mostrarán en la imagen.</p>

Si el algoritmo de segmentación no muestra un resultado adecuado. El especialista puede recurrir a un método manual donde fija una semilla y un paso y selecciona segmentar, el sistema clasificador comienza a iterar hasta terminar o hasta que el especialista detenga las iteraciones. El especialista tiene la opción de llevar a cabo la caracterización de núcleos individuales o de citoplasmas utilizando la opción de fijar Cito/Núcleo. La caracterización de núcleo y citoplasma de una misma célula se presenta en una misma interfaz.

Al fijar núcleo o citoplasma en la interfaz de caracterización se visualizan todas las medidas. El especialista utiliza las acciones de saturación y luminosidad del núcleo y citoplasma. El sistema clasificador con el algoritmo de la matriz de co-ocurrencia halla los valores y los presenta en la misma interfaz.

La interfaz de caracterización de la célula seleccionada visualizará todas las medidas y la opción de clasificar. El especialista le solicita al sistema que efectúe la clasificación y se mostrará la clasificación en la misma interfaz.

Flujo alterno

1. El especialista carga una imagen ya seleccionada.

Se continúa con los pasos posteriores.

2. Se utilizan los métodos de segmentación por umbralización.

Se continúa con los pasos posteriores.

Post-condiciones

La Célula ha sido clasificada, el especialista confrontará el resultado del sistema con su criterio y confirmará o rechazará el resultado.

10. RESULTADOS

Posterior al modelo computacional se llevaron a cabo una serie de pruebas de validación para la clasificación de células. Durante las pruebas se utilizaron 32 células con diagnóstico positivo y 19 benignas, ninguna de estas células fue utilizada durante el desarrollo de la investigación, ni durante el entrenamiento de la RNA.

Para las pruebas se utilizaron tablas de verdad con el fin de determinar el número de células clasificadas correcta o incorrectamente.

10.1. VALIDEZ DE LA PRUEBA DIAGNÓSTICA

Se utiliza el cálculo de la Sensibilidad y Especificidad con el fin de establecer pautas de validez para el modelo computacional propuesto.

10.1.1. SENSIBILIDAD

Es la probabilidad de clasificar correctamente a un individuo enfermo, es decir, la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo. La sensibilidad es, por lo tanto, la capacidad del modelo computacional para detectar la enfermedad [Pita and Pértegas, 2003].

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (42)$$

Donde VP corresponde a Verdaderos Positivos y FN a Falsos Negativos.

10.1.2. ESPECIFICIDAD

Es la probabilidad de clasificar correctamente a un individuo sano, es decir, la probabilidad de que para un sujeto sano se obtenga un resultado negativo (121). En otras palabras, se puede definir la especificidad como la capacidad del modelo computacional para detectar células sanas.

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (43)$$

Donde VN corresponde a Verdaderos Negativos y FP a Falsos Positivos. De ahí que también se denomina *fracción de verdaderos negativos (FVN)*.

10.2. SEGURIDAD DE LA PRUEBA DIAGNÓSTICA

Los conceptos de sensibilidad y especificidad permiten valorar la validez de una prueba diagnóstica, sin embargo, carecen de utilidad en la práctica clínica [Pita and Pértegas, 2003, p.123]. Tanto la sensibilidad como la especificidad proporcionan información acerca de la probabilidad de obtener un resultado concreto (positivo o negativo) en función de la verdadera condición del enfermo con respecto a la enfermedad. Sin embargo, cuando a algún paciente se le realiza alguna prueba, el médico carece de información a priori acerca de su verdadero diagnóstico, y más bien la pregunta se plantea en sentido contrario: ante un resultado positivo (negativo) en la prueba, ¿cuál es la probabilidad de que el paciente esté realmente enfermo (sano)? Así pues, resulta obvio que hasta el momento sólo se ha abordado el problema en una sola dirección. Por medio de los valores predictivos se complementa ésta información.

10.2.1. VALOR PREDICTIVO POSITIVO

Es la probabilidad de padecer la enfermedad, o de la célula ser catalogada como maligna, si se obtiene un resultado positivo mediante el modelo computacional. El valor predictivo positivo puede estimarse, por tanto, a partir de la proporción de células con un resultado positivo en la prueba que finalmente resultaron ser malignas:

$$VPP = \frac{VP}{VP + FP} \quad (44)$$

Siendo VP Verdaderos Positivos y FP Falsos Positivos.

10.2.2. VALOR PREDICTIVO NEGATIVO

Es la probabilidad de que una célula con un resultado negativo en el modelo computacional esté realmente sana. Se estima dividiendo el número de verdaderos negativos entre el total de células con un resultado negativo a través del modelo propuesto:

$$VPN = \frac{VN}{FN + VN} \quad (45)$$

10.3. INFLUENCIA DE LA PREVALENCIA

A pesar de que los valores de sensibilidad y especificidad definen completamente la validez de la prueba diagnóstica, presentan la desventaja de no proporcionar información relevante a la hora de tomar una decisión clínica ante un determinado resultado de la prueba. Sin embargo, tienen la ventaja adicional de que son propiedades intrínsecas a la prueba diagnóstica, y definen su validez independientemente de cuál sea la prevalencia de la enfermedad en la población a la cual se aplica.

Por el contrario, el concepto de valores predictivos, a pesar de ser de enorme utilidad a la hora de tomar decisiones clínicas y transmitir a los pacientes información sobre su diagnóstico, presenta la limitación de que dependen en gran medida de lo frecuente que sea la enfermedad a diagnosticar en la población objeto de estudio. Cuando la prevalencia de la enfermedad es baja, un resultado negativo permitirá descartar la enfermedad con mayor seguridad, siendo así el valor predictivo negativo mayor. Por el contrario, un resultado positivo no permitirá confirmar el diagnóstico, resultando en un bajo valor predictivo positivo (123).

10.4. RAZONES DE PROBABILIDAD

Dado que la prevalencia es un factor determinante en los valores predictivos de una prueba diagnóstica, éstos no pueden ser utilizados como índices a la hora de comparar y/o extrapolar los resultados de otros estudios a datos propios (124). En consecuencia, resulta necesario calcular otros índices de valoración que sean clínicamente útiles y no dependan de la prevalencia de la enfermedad en la población a estudiar. Por tanto, se suele hablar de nuevos conceptos: razón de verosimilitudes, razón de probabilidad, o cociente de probabilidades [Dujardin et al., 1994]; midiendo cuánto más probable es un resultado concreto (positivo o negativo) según la presencia o ausencia de enfermedad:

10.4.1. RAZÓN DE VEROSIMILITUDES POSITIVA O COCIENTE DE PROBABILIDADES POSITIVO

Se calcula dividiendo la probabilidad de un resultado positivo en los pacientes enfermos entre la probabilidad de un resultado positivo entre los sanos. Es, en

definitiva, el cociente entre la fracción de verdaderos positivos (sensibilidad) y la fracción de falsos positivos (1-especificidad):

$$RV_+ = \frac{\textit{Sensibilidad}}{1 - \textit{Especificidad}} \quad (46)$$

10.4.2. RAZÓN DE VEROSIMILITUDES NEGATIVA O COCIENTE DE PROBABILIDADES NEGATIVO

Se calcula dividiendo la probabilidad de un resultado negativo en presencia de enfermedad entre la probabilidad de un resultado negativo en ausencia de la misma. Se calcula como el cociente entre los falsos negativos ($1 - \textit{Sensibilidad}$) y los verdaderos negativos ($\textit{Especificidad}$):

$$RV_- = \frac{1 - \textit{Sensibilidad}}{\textit{Especificidad}} \quad (47)$$

10.5. RESULTADOS Y VALIDACIÓN DEL MODELO

Los resultados obtenidos con el modelo computacional son presentados en las tablas 7 y 8.

<i>Clasificación Real</i>	<i>Clasificación del Modelo Computacional</i>
Negativa	19
Positiva	30

Tabla 7: Clasificación celular realizada por el modelo computacional.

<i>Falsos Negativos</i>	2
<i>Falsos Positivos</i>	0

Tabla 8: Número de Falsos Positivos y Negativos obtenidos con el modelo computacional.

Conforme las ecuaciones 42, 43, 44, 45, 46 y 47 se tiene:

$$Sensibilidad = \frac{30}{30 + 2} * 100 = \mathbf{93,75 \%} \quad (48)$$

$$Especificidad = \frac{19}{19 + 0} * 100 = \mathbf{100 \%} \quad (49)$$

$$VPP = \frac{30}{30 + 0} * 100 = \mathbf{100 \%} \quad (50)$$

$$VPN = \frac{19}{2 + 19} * 100 = \mathbf{90,48 \%} \quad (51)$$

$$RV_+ = \frac{0,9375}{1 - 1} * 100 = \textit{Indeterminado} \quad (52)$$

$$RV_- = \frac{1 - 0,9375}{1} * 100 = \mathbf{6,25 \%} \quad (53)$$

Se puede observar que los valores de RV_+ y RV_- , cociente de probabilidades positivo y cociente de probabilidades negativo, son el complemento de las probabilidades concerniente a *Especificidad* y *Sensibilidad*, respectivamente. Desde este punto de vista se puede considerar el valor RV_+ igual a cero.

Comparando los cocientes de probabilidad de dos pruebas diagnósticas existen cuatro posibles resultados que pueden ser reflejados en una gráfica como herramienta de validación [Molinero, 2002]. En el eje de las Y se representa la *Sensibilidad*, en el eje X se representa $1 - Especificidad$. El punto color rojo de la Figura 75 representa los valores de *Sensibilidad* y $(1 - Especificidad)$ observados en la prueba de examen citológico descrito en el marco teórico, cuya sensibilidad oscila entre 50 % y 80 % que para la presente validación se tomó el valor más alto, con especificidad del 97 %. Se traza una línea desde el origen $(0,0)$ que pase por el punto rojo y otra desde el valor $(1,1)$ que también pase por ese punto, quedando dividido el gráfico en cuatro zonas como se aprecia en la figura, de tal manera que cualquier otra prueba cuyo resultado se represente gráficamente estará en alguna de las cuatro zonas.

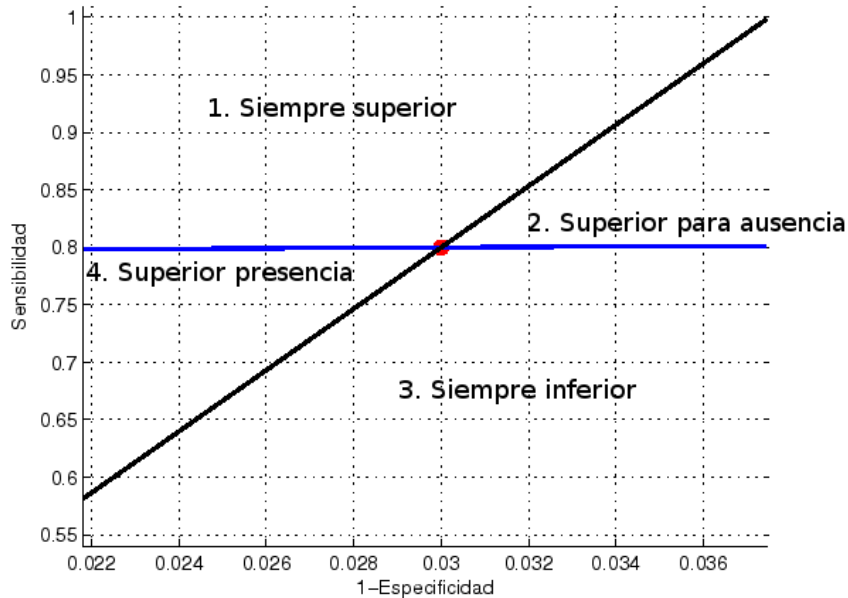


Figura 75: Gráfica de validación examen citológico.

Las zonas corresponden a las siguientes situaciones:

1. En ésta región la prueba alternativa será siempre superior al examen citológico, tanto para confirmar presencia como ausencia de enfermedad.
2. En ésta zona la prueba alternativa sólo es superior al examen citológico para confirmar ausencia de enfermedad.
3. Aquí la prueba alternativa siempre será inferior al examen citológico.
4. Si el resultado de la prueba alternativa está en ésta área, será superior al examen citológico para confirmar la presencia de enfermedad.

En la Figura 76 se compara la prueba de examen citológico con el resultado del modelo computacional propuesto cuya sensibilidad es 93,75 % y especificidad 100 %, representado por el punto azul.

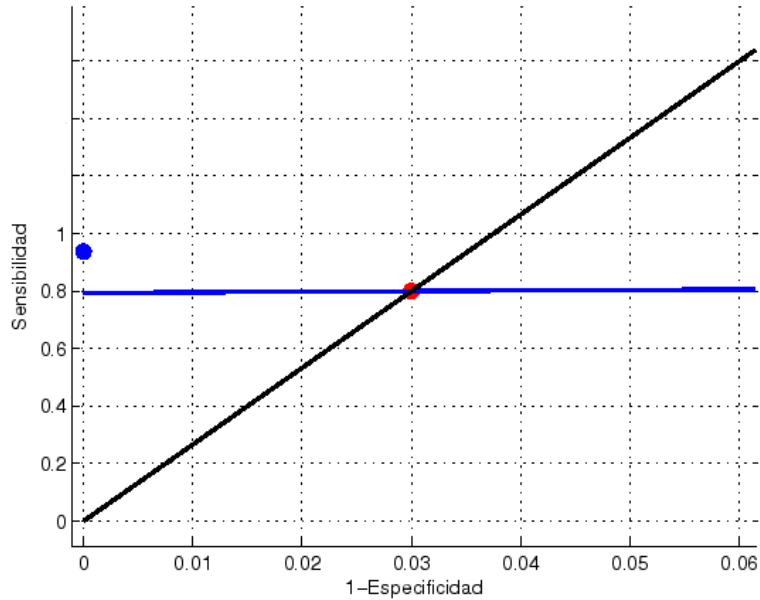


Figura 76: Gráfico de comparación prueba citológica y modelo propuesto.

La contundencia de los resultados obtenidos demuestra que la clasificación de células en fluido pleural aumenta significativamente su sensibilidad, especificidad, valor predictivo positivo y negativo si se soporta en el análisis y procesamiento de imágenes digitales, utilizando además las ventajas ofrecidas por las redes neuronales artificiales, haciendo del modelo computacional una herramienta muy superior frente al examen citológico tanto para confirmar presencia como ausencia de malignidad en imágenes celulares de líquido pleural.

11. CONCLUSIONES Y RECOMENDACIONES

Los algoritmos empleados y el procesamiento implementado mostraron resultados muy positivos comparado con el estudio citológico de pleura, lo cual se ve reflejado en el aumento de las tasas de sensibilidad y especificidad [Pertuze and Aparicio, 1997], [Maghfoor et al., 2000]. De ésta manera se cuenta con una herramienta computacional capaz de disminuir el número de falsos positivos y en el mejor de los casos también reducir la cantidad de falsos negativos.

Si bien se alcanzaron resultados satisfactorios, la continuidad de la investigación debería ampliar su universo de imágenes, abriendo horizontes hacia la detección de sitios primarios para células malignas de tipo adenocarcinoma presentes en fluido pleural, favoreciendo de ésta manera la detección oportuna de cáncer y su pronta intervención que redunde en beneficio del paciente y su círculo familiar.

El campo de la Inteligencia Artificial es un amplio terreno sobre el cual explorar y experimentar, buscando la creación de nuevos y mejores algoritmos que aplicados al campo del tratamiento de imágenes médicas impacte de manera positiva no solo en el dominio académico e institucional sino también en el más importante, la salud de la población más vulnerable quien en últimas resulta ser la más beneficiada con éstas técnicas de bajo costo económico y computacional.

BIBLIOGRAFÍA

- [Chávez et al., 2005] Chávez, J., González, M., González, R., Cortés, E., Sánchez, R., and Chávez, A. (2005). Patología Pleural, modalidad clínico-quirúrgica. *Revista Hospital Juárez de México*, 72(4):139–147.
- [Delgado et al., 2006] Delgado, S., Villamizar, D., Ojeda, O., and Martínez, V. (2006). Mesoft 1.0: Prototipo software para la caracterización de células mesoteliales no tumorales malignas en citología de líquido pleural. Master's thesis, Universidad Industrial de Santander, Bucaramanga.
- [Dujardin et al., 1994] Dujardin, B., Van der Ende, J., Van Gompel, A., Unger, J., and Van der Stuyft, P. (1994). Likelihood ratios: a real improvement for clinical decision making? *Eur J Epidemiol*, 10(1):29–36. En línea: <http://www.ncbi.nlm.nih.gov/sites/entrez>.
- [Fariña and Rodríguez, 1996] Fariña, J. and Rodríguez, J. (1996). *Citopatología Respiratoria y Pleural*. Madrid.
- [González and Woods, 1996] González, R. and Woods, R. (1996). *Tratamiento Digital de Imágenes*.
- [Hanbury and Serra, 2003] Hanbury, A. and Serra, J. (2003). Colour image analysis in 3d-polar coordinates. In *Dagm Symposium*, volume 25, pages 124–131, Magdeburg.
- [Haykin, 1999] Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*.
- [Jacobson et al., 2000] Jacobson, I., Booch, G., and Rumbaugh, J. (2000). *El Proceso Unificado de Desarrollo de Software*. Pearson Educación, S.A., Madrid.
- [Johnston, 1985] Johnston, W. (1985). The malignant pleural effusion: a review of cytopathologic diagnoses of 584 specimens from 472 consecutive patients. *Cancer*, 56(4):905–909.
- [Maghfoor et al., 2000] Maghfoor, I., Doll, D., and Yarbro, J. (2000). Effusions. *Clinical Oncology*, pages 922–949.

- [Martínez, 2007] Martínez, V. (2007). Modelo computacional para la caracterización de células endocervicales. Master's thesis, Universidad Industrial de Santander, Bucaramanga.
- [Mattison et al., 1997] Mattison, L., Coppage, L., Alderman, D., Herlong, J., and Sahn, S. (1997). Pleural Effusions in the Medical ICU: prevalence, causes, and clinical implications. *American College of Chest Physicians*, 111(4):1018–1023.
- [Moliner, 2002] Moliner, L. (2002). Valoración de pruebas diagnósticas. En Línea: <http://www.seh-lelha.org/pdiagnos.htm>. Consulta en Septiembre de 2007.
- [Oropesa and Soto, 2002] Oropesa, A. and Soto, S. (2002). Derrame pleural neoplásico. *Cuadernos de Cirugía*, 16(1):92–99.
- [Pertuze and Aparicio, 1997] Pertuze, J. and Aparicio, R. (1997). Derrame Pleural Neoplásico. *Boletín de la Escuela de Medicina*, 26(2).
- [Pita and Pértegas, 2003] Pita, F. S. and Pértegas, D. S. (2003). Pruebas diagnósticas. *Cad Aten Primaria*, (10):120–124.
- [Presutti, 2004] Presutti, M. (2004). La Matriz de Co-ocurrencia en la Clasificación Multiespectral: Tutorial para la Enseñanza de Medidas Texturales en Cursos de Grado Universitario. In *Recursos Didácticos para o Ensino de Sensoriamento Remoto*, São Leopoldo, RS, Brasil. 4ª Jornada de Educação em Sensoriamento Remoto no Âmbito do Mercosul.
- [Reina, 2003] Reina, M. (2003). Técnicas de contaje celular. <http://www.ub.es/biocel/wbc/tecnicas/contajecelular.htm>. Consulta: Septiembre de 2007.
- [Rodríguez, 1993] Rodríguez, F. (1993). *Tumores Pleurales Secundarios*. Patología Pleural. Barcelona.
- [Sampallo, 2003] Sampallo, G. (2003). Reconocimiento de Tipos de Hojas. *Revista Iberoamericana de Inteligencia Artificial*, (21):55–62.

ANEXOS

A. FUNCIONAMIENTO DEL MODELO COMPUTACIONAL

La herramienta software que soporta el modelo computacional se desarrolló totalmente en Matlab® R2007a. El motor de base de datos utilizado para soportar las reglas de la caracterización es MySQL. El primer paso para ejecutar la aplicación es ubicar el directorio donde se encuentran las funciones y ejecutar el comando para iniciar la interfaz gráfica, como se muestra en la siguiente figura.

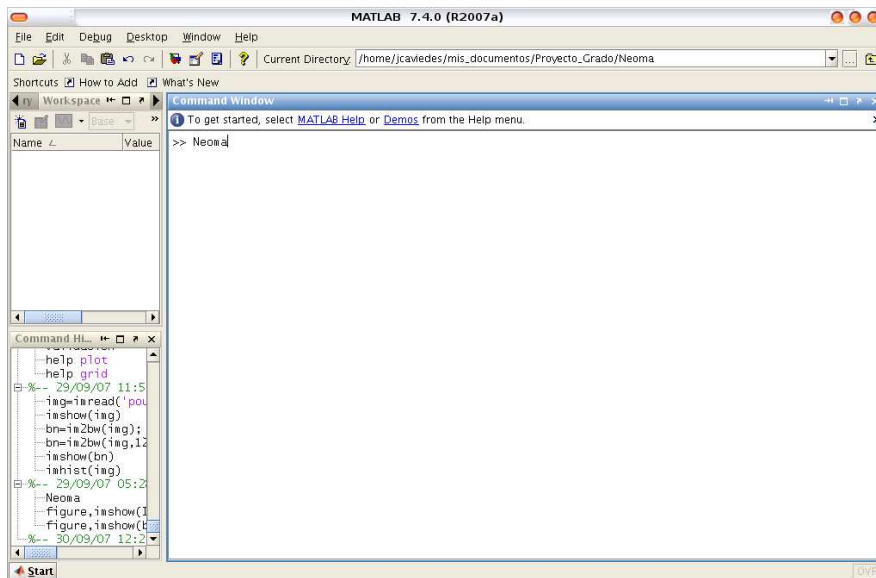


Figura 77: Inicio de la Herramienta Software.

Luego de ejecutar en la ventana de comandos la línea Neoma, se despliega la ventana principal, que habilitará las opciones de procesamiento mientras avanza la caracterización y clasificación celular. Se encuentran las opciones disponibles: abrir imagen y salir, del menú archivo si es la primera vez que se va a procesar una nueva célula, y la opción obtener célula del menú tratamiento si se encuentra guardada una célula seleccionada previamente.



Figura 78: Ventana principal de la herramienta software.

Al utilizar la opción *Abrir imagen* se despliega una ventana de exploración con el propósito de buscar la imagen celular que se desea analizar. El software sólo debe ser utilizado para el análisis de células en fluido pleural. La Figura 79 muestra la forma en que se selecciona una imagen celular.

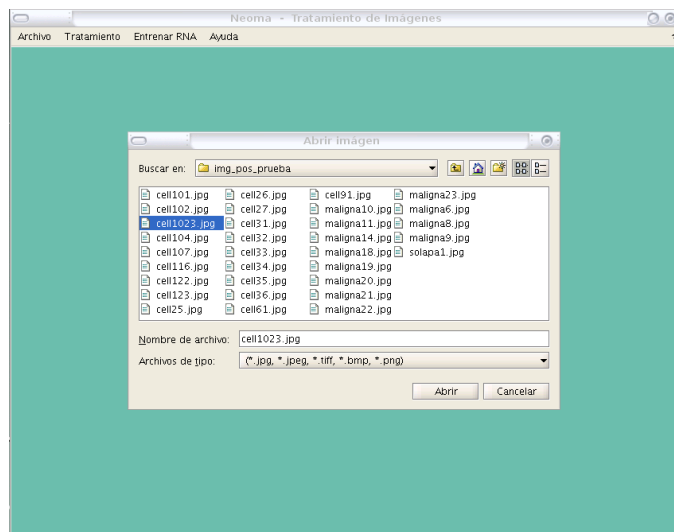


Figura 79: Opción abrir imagen.

Al abrir una imagen se presenta la opción de seleccionar célula como se presenta en la Figura 80.

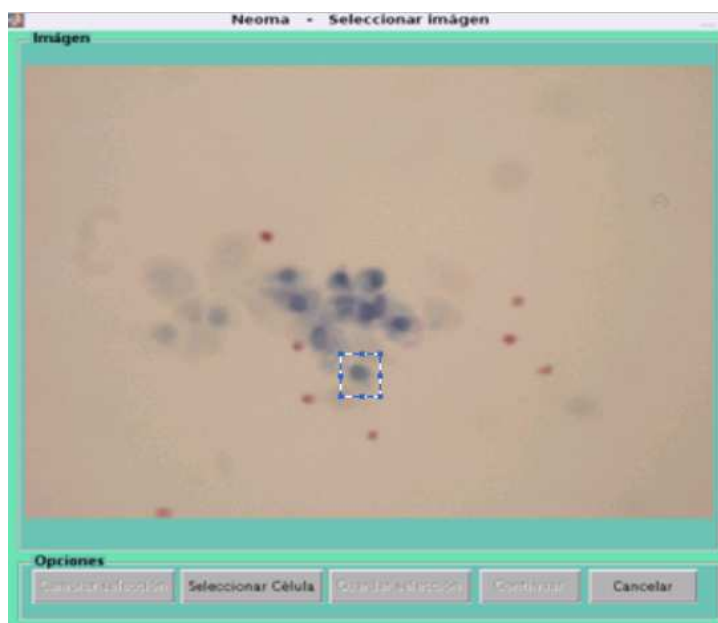


Figura 80: Opción seleccionar célula.

Una vez seleccionada la célula se presenta la opción de guardarla para un posterior análisis y abrirla directamente desde la opción obtener célula del menú Tratamiento. Si no se guarda se sigue con el proceso utilizando la opción Continuar, el software la mantendrá en memoria. La figura 81 presenta la célula seleccionada.

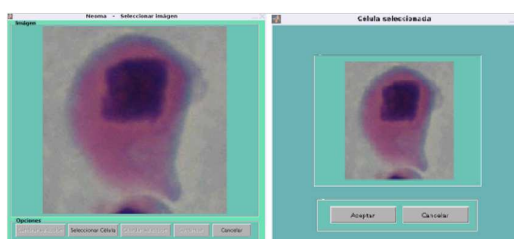


Figura 81: Célula seleccionada.

Después de estar conforme con la selección de la célula se despliega la ventana de tratamiento, la cual permite llevar a cabo todo el procesamiento de

la imagen. La ventana presenta tres áreas para imagen: la primera muestra la imagen original, la segunda muestra la imagen en escala de gris, la tercera presenta los resultados que se van obteniendo a medida que avanza el tratamiento de la imagen; y tres áreas para el procesamiento de la imagen: una para el filtrado, otra para la segmentación y otra para operaciones de morfología matemática. A continuación se presenta la figura 82 de la ventana Tratamiento.

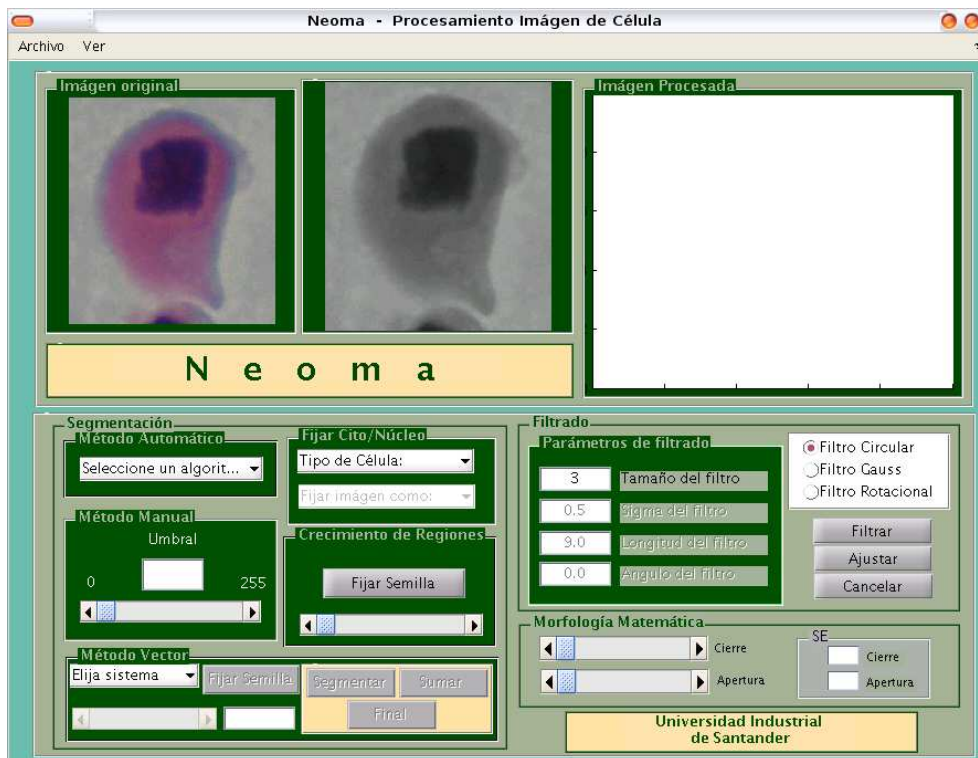


Figura 82: Ventana para Tratamiento de imagen celular.

Al desplegarse la ventana de Procesamiento de mejora de la imagen, el siguiente paso es seleccionar un tipo de filtro con algunos parámetros, accionar el comando Filtrar y posterior al resultado ajustar la imagen mediante el comando con el mismo nombre. La Figura 83 presenta este primer paso de procesamiento.

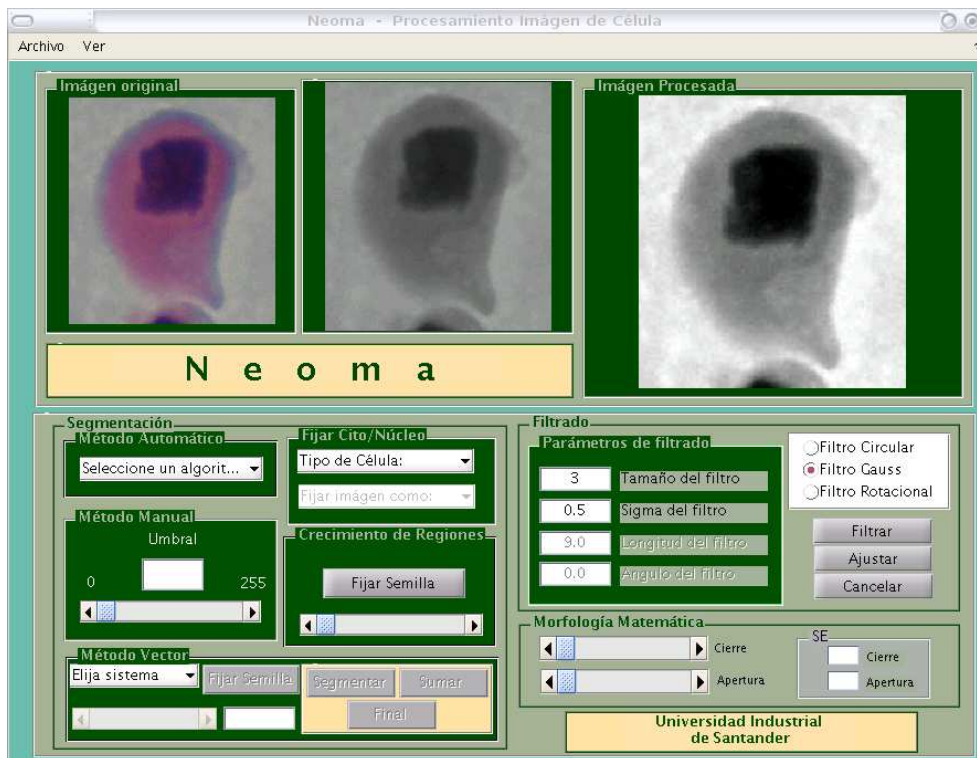


Figura 83: Imagen celular filtrada y ajustada.

El siguiente paso al filtrado es escoger un método de segmentación entre los cuales hay un método automático, uno manual, uno vectorial y el de crecimiento de regiones siendo este último el más utilizado para la segmentación, se acciona el comando fijar semilla permitiendo al usuario dar un clic sobre el área en que se encuentra la imagen procesada y posteriormente utilizar la barra deslizadora que se encuentra en la opción de crecimiento de regiones. Al comenzar el traslado de la barra en la segunda área de imagen aparece una imagen binaria como resultado de la segmentación que se está obteniendo y en la tercera área se muestra la segmentación sobre la imagen original así permitiendo al observador la mejor segmentación posible. El usuario escoge segmentar núcleo o citoplasma de manera arbitraria.

La Figura 84 presenta la segmentación de un núcleo.

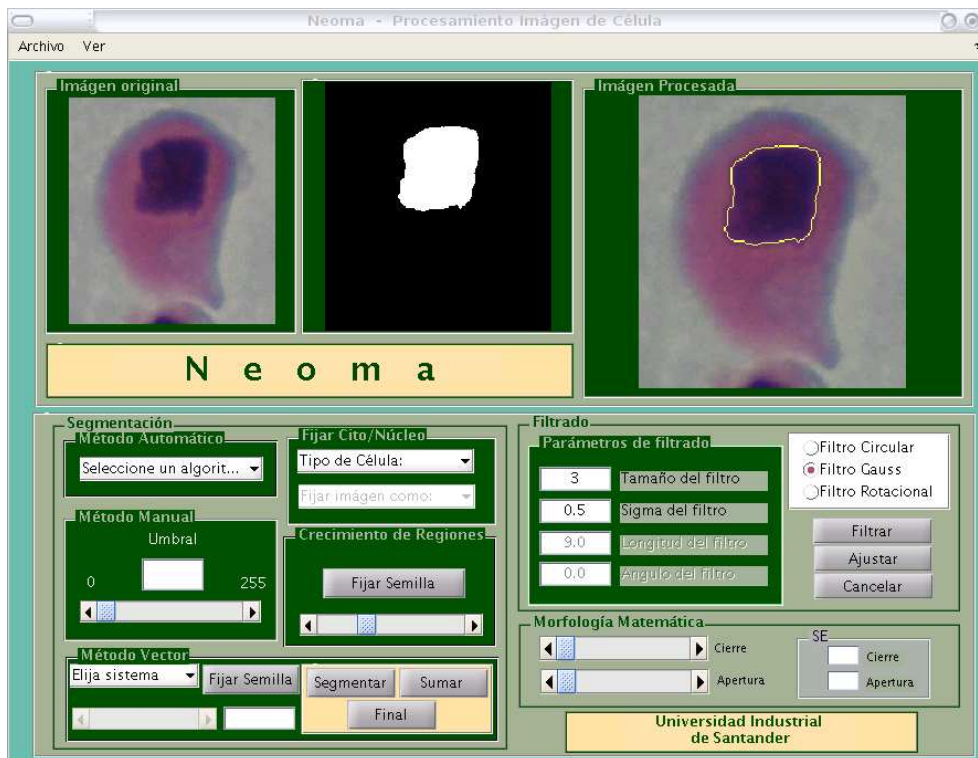


Figura 84: Segmentación de un núcleo.

Una vez segmentado el núcleo se acciona la opción de Fijar Núcleo y automáticamente se despliega una ventana que presenta las medidas del núcleo y los valores de la matriz de co-ocurrencia para saturación y luminosidad. La Figura 86 presenta las medidas y valores de un núcleo.

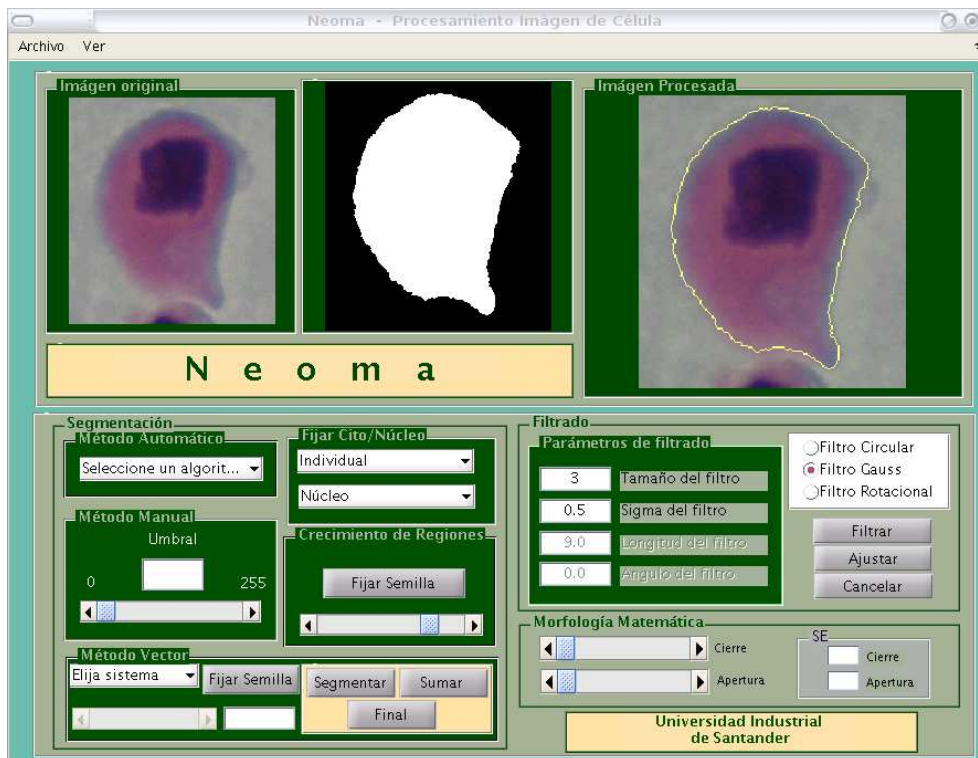


Figura 85: Segmentación de un citoplasma.

Al desplegarse la ventana de descripción de parámetros y obtener los valores para el núcleo se regresa a la ventana de procesamiento para segmentar el citoplasma como se presenta en la Figura 85.

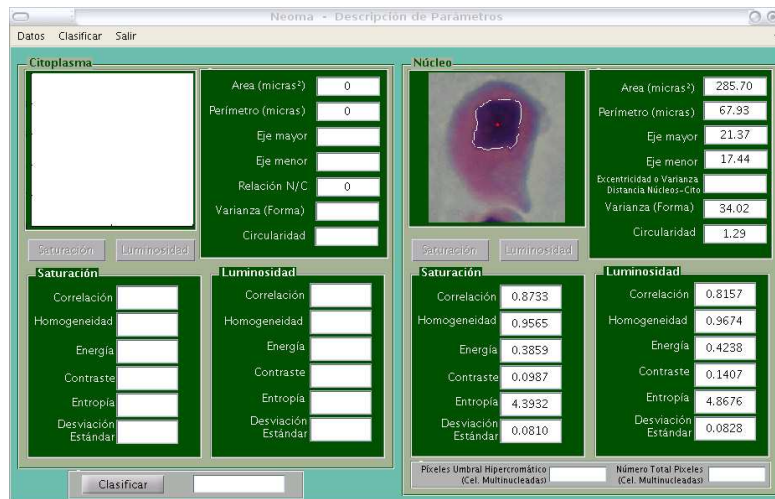


Figura 86: Descripción de parámetros nucleares.

Segmentado el citoplasma se pasa a la opción de Fijar Citoplasma y al igual que con el núcleo se despliega la ventana que presenta los valores de citoplasma y núcleo como lo muestra la Figura 87.

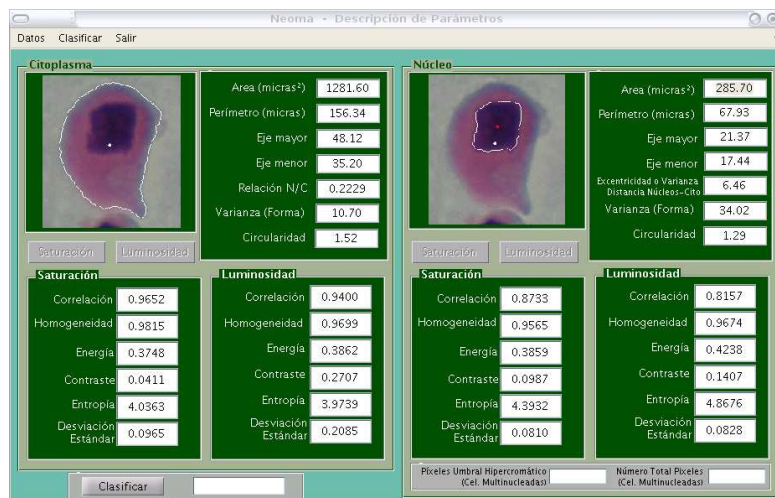


Figura 87: Descripción de parámetros Citoplasma y Núcleo.

Al obtener la descripción de parámetros de Citoplasma y Núcleo, finalmente el usuario utiliza el comando Clasificar, mostrando como resultado: Benigna o Maligna. Tal como se presenta en la Figura 88.

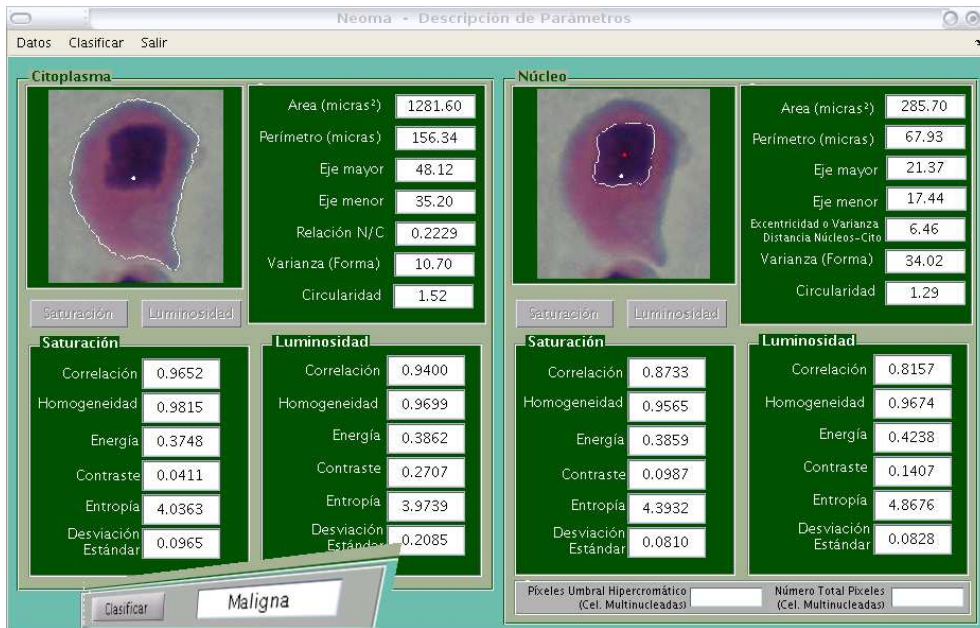


Figura 88: Resultado de la Clasificación Celular.