



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE ELÉTRICA Y ELECTRÓNICA

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y
TELECOMUNICACIONES**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN
DEL TÍTULO DE INGENIERO EN ELETRÓNICA Y
TELECOMUNICACIONES**

**TEMA: RECONOCIMIENTO DE CÁNCER DE MAMA
UTILIZANDO TÉCNICAS DE PROCESAMIENTO DIGITAL DE
IMÁGENES MAMOGRÁFICAS Y APRENDIZAJE DE MÁQUINA.**

AUTOR: CARRASCO LLERENA, CHRISTIAN ANDRÉS

DIRECTOR: PhD. CARRERA ERAZO, ENRIQUE VINICIO

SANGOLQUÍ

2017

CERTIFICACIÓN



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE ELÉTRICA Y ELECTRÓNICA

CARRERA DE ELECTRÓNICA Y TELECOMUNICACIONES

CERTIFICACIÓN

Certifico que el trabajo de titulación "RECONOCIMIENTO DE CÁNCER DE MAMA UTILIZANDO TÉCNICAS DE PROCESAMIENTO DIGITAL DE IMÁGENES MAMOGRÁFICAS Y APRENDIZAJE DE MÁQUINA," realizado por el señor **CHRISTIAN ANDRÉS CARRASCO LLERENA**, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar al señor **CHRISTIAN ANDRÉS CARRASCO LLERENA** para que lo sustente públicamente.

Sangoquí, 17 de Abril del 2017

A handwritten signature in black ink, appearing to read "Enrique Vinicio Carrera Erazo", written over a horizontal line.

PhD. Enrique Vinicio Carrera Erazo

DIRECTOR

AUTORÍA DE RESPONSABILIDAD



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE ELÉTRICA Y ELECTRÓNICA

CARRERA DE ELECTRÓNICA Y TELECOMUNICACIONES

AUTORÍA DE RESPONSABILIDAD

Yo, CHRISTIAN ANDRÉS CARRASCO LLERENA, con cédula de identidad N° 1719065102 declaro que este trabajo de titulación "RECONOCIMIENTO DE CÁNCER DE MAMA UTILIZANDO TÉCNICAS DE PROCESAMIENTO DIGITAL DE IMÁGENES MAMOGRÁFICAS Y APRENDIZAJE DE MÁQUINA" ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoria, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada

Sangolquí, 17 de Abril del 2017

Christian Andrés Carrasco Llerena

C. C. 1719065102

AUTORIZACIÓN



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE ELÉTRICA Y ELECTRÓNICA

CARRERA DE ELECTRÓNICA Y TELECOMUNICACIONES

AUTORIZACIÓN

Yo, **CHRISTIAN ANDRÉS CARRASCO LLERENA**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación "**RECONOCIMIENTO DE CÁNCER DE MAMA UTILIZANDO TÉCNICAS DE PROCESAMIENTO DIGITAL DE IMÁGENES MAMOGRÁFICAS Y APRENDIZAJE DE MÁQUINA**" cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 17 de Abril del 2017

A handwritten signature in black ink, appearing to read "CARRASCO LLERENA", is written over a horizontal dashed line.

Christian Andrés Carrasco Llerena

C. C. 1719065102

DEDICATORIA

A mis padres que son los gestores fundamentales de la excelente crianza y educación recibida, quienes siempre están dispuestos a regalarme un buen consejo, un regaño que muchas veces creo no merecer, pero que a pesar de todo son incondicionales en brindarme su amor, en compartir mis alegrías y tratar de simplificar mis tristezas.

A mi hermano, que es un amigo incondicional, de quien tengo el buen ejemplo a seguir, el cariño, el amor y su desinteresado apoyo.
A todas aquellas personas que siempre han estado junto a mí, cuando he celebrado mis triunfos y llorando mis derrotas.

Christian Andrés Carrasco Llerena

AGRADECIMIENTO

A Dios por darme la oportunidad de conocer a grandes amigos como son muchos de mis compañeros y todos mis maestros

A las autoridades, personal docente y empleados de esta noble institución por hacerme sentir que nuestra universidad es una familia, en la que cooperamos todos con el fin de buscar siempre el bienestar general. Al PhD. Vinicio Carrera quien con su sabiduría, conocimientos y experiencia me guio en la elaboración de este proyecto, gracias por exigirme el mejor esfuerzo y aportar siempre con sus consejos oportunos y certeros.

A mi familia que busca hacer de mí un profesional honesto y capaz de cambiar el destino de la patria

Christian Andrés Carrasco Llerena

ÍNDICE DE CONTENIDO

CARATULA	i
CERTIFICACIÓN	ii
AUTORÍA DE RESPONSABILIDAD	iii
AUTORIZACIÓN	iv
DEDICATORIA	v
AGRADECIMIENTO	vi
ÍNDICE DE CONTENIDO	vii
ÍNDICE DE TABLAS	x
ÍNDICE DE FIGURAS	xiv
RESUMEN	xvi
ABSTRACT	xviii
CAPÍTULO I	1
1. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN	1
1.1. Antecedentes	1
1.2. Justificación e Importancia	3
1.3. Alcance del proyecto	4
1.4. Objetivos	5
1.4.1. Objetivo general	5
1.4.2. Objetivos Específicos	5
CAPÍTULO II	6
2. MARCO TEÓRICO	6
2.1. Cáncer de Mama	6
2.1.1. Factores de riesgo	8
2.1.2. Diagnóstico	9
2.1.3. Mamografías	10
2.2. Procesamiento Digital de Imágenes	12
2.2.1. Imagen digital	13
2.2.2. Imagen en escala de grises	14
2.2.3. Segmentación	15

2.2.3.1.	Segmentación por Umbralización	16
2.2.3.2.	Método de Otsu	17
2.2.3.3.	Operaciones morfológicas	18
2.2.3.4.	Operaciones morfológicas básicas	18
2.2.3.5.	Erosión.....	19
2.2.3.6.	Dilatación.....	20
2.3.	Machine Learning.....	21
2.3.1.	Funcionamiento de Machine Learning	21
2.3.2.	Tipos de algoritmos.....	23
2.3.2.1.	Aprendizaje supervisado.....	23
2.3.2.1.1.	Técnica tipo supervisado Bayesiano Ingenuo.....	23
2.3.2.1.2.	Vecino más Cercano	25
2.3.2.1.3.	Árbol de decisiones	27
2.3.2.1.4.	Redes Neuronales Artificiales (ANN).....	28
2.4.	Evaluación	30
2.4.1.	<i>K-fold</i> Validación cruzada.....	30
2.4.2.	Matriz de confusión	31
2.4.2.1.	Sensibilidad.....	32
2.4.2.2.	Precisión.....	32
2.4.2.3.	Tasa de falso descubrimiento (FDR)	33
2.4.2.4.	Curvas ROC.....	33
CAPÍTULO III	35
3. DISEÑO DEL SISTEMA	35
3.1.	Introducción	35
3.2.	Diagrama de Bloques	35
3.3.	Imágenes Mamográficas	36
3.3.1.	Eliminación de etiquetas en la imagen	37
3.3.2.	Inversión de intensidad de la imagen	39
3.4.	Segmentación.....	40
3.4.1.	Valor umbral por el método de Otsu.....	40
3.4.2.	Operaciones morfológicas	44
3.5.	Características.....	47

3.6. Machine Learning.....	49
3.7. Validación	51
CAPÍTULO IV	52
4. ANÁLISIS DE PARÁMETROS DEL SISTEMA Y RESULTADOS OBTENIDOS..	52
4.1. Introducción	52
4.2. Obtención de parámetros.....	52
4.3. Discusión	106
CAPÍTULO V	107
5. CONCLUSIONES Y RECOMENDACIONES	107
BIBLIOGRAFÍA.....	109

ÍNDICE DE TABLAS

Tabla 1	Tipos de cáncer de mama	7
Tabla 2	Factores que propician el cáncer	8
Tabla 3	Tipos de procedimientos para el diagnóstico de cáncer de mama	10
Tabla 4	Operaciones morfológicas	19
Tabla 5	Partes de una matriz de confusión	31
Tabla 6	Exactitud vs (solidity 0.1, umbral y nivel) para ANN con red Fitnet	54
Tabla 7	Eficiencia vs (solidity 0.1, umbral y nivel) para ANN con red Fitnet	54
Tabla 8	Exactitud vs (solidity 0.2, umbral y nivel) para ANN con red Fitnet	55
Tabla 9	Eficiencia vs (solidity 0.2, umbral y nivel) para ANN con red Fitnet	55
Tabla 10	Exactitud vs (solidity 0.3, umbral y nivel) para ANN con red Fitnet	56
Tabla 11	Eficiencia vs (solidity 0.3, umbral y nivel) para ANN con red Fitnet	56
Tabla 12	Exactitud vs (solidity 0.4, umbral y nivel) para ANN con red Fitnet	57
Tabla 13	Eficiencia vs (solidity 0.4, umbral y nivel) para ANN con red Fitnet	57
Tabla 14	Exactitud vs (solidity 0.5, umbral y nivel) para ANN con red Fitnet	58
Tabla 15	Eficiencia vs (solidity 0.5, umbral y nivel) para ANN con red Fitnet	58
Tabla 16	Exactitud vs (solidity 0.6, umbral y nivel) para ANN con red Fitnet	59
Tabla 17	Eficiencia vs (solidity 0.6, umbral y nivel) para ANN con red Fitnet	59
Tabla 18	Exactitud vs (solidity 0.7, umbral y nivel) para ANN con red Fitnet	60
Tabla 19	Eficiencia vs (solidity 0.7, umbral y nivel) para ANN con red Fitnet	60
Tabla 20	Exactitud vs (solidity 0.8, umbral y nivel) para ANN con red Fitnet	61
Tabla 21	Eficiencia vs (solidity 0.8, umbral y nivel) para ANN con red Fitnet	61
Tabla 22	Exactitud vs (solidity 0.9, umbral y nivel) para ANN con red Fitnet	62
Tabla 23	Eficiencia vs (solidity 0.9, umbral y nivel) para ANN con red Fitnet	62
Tabla 24	Exactitud vs (solidity 0.1, umbral y nivel) para vecino más cercano	63
Tabla 25	Eficiencia vs (solidity 0.1, umbral y nivel) para vecino más cercano	64
Tabla 26	Exactitud vs (solidity 0.2, umbral y nivel) para vecino más cercano	64
Tabla 27	Eficiencia vs (solidity 0.2, umbral y nivel) para vecino más cercano	65

Tabla 28 Exactitud vs (solidity 0.3, umbral y nivel) para vecino más cercano.....	65
Tabla 29 Eficiencia vs (solidity 0.3, umbral y nivel) para vecino más cercano	66
Tabla 30 Exactitud vs (solidity 0.4, umbral y nivel) para vecino más cercano.....	66
Tabla 31 Eficiencia vs (solidity 0.4, umbral y nivel) para vecino más cercano	67
Tabla 32 Exactitud vs (solidity 0.5, umbral y nivel) para vecino más cercano.....	67
Tabla 33 Eficiencia vs (solidity 0.5, umbral y nivel) para vecino más cercano	68
Tabla 34 Exactitud vs (solidity 0.6, umbral y nivel) para vecino más cercano.....	68
Tabla 35 Eficiencia vs (solidity 0.6, umbral y nivel) para vecino más cercano	69
Tabla 36 Exactitud vs (solidity 0.7, umbral y nivel) para vecino más cercano.....	69
Tabla 37 Eficiencia vs (solidity 0.7, umbral y nivel) para vecino más cercano	70
Tabla 38 Exactitud vs (solidity 0.8, umbral y nivel) para vecino más cercano.....	70
Tabla 39 Eficiencia vs (solidity 0.8, umbral y nivel) para vecino más cercano	71
Tabla 40 Exactitud vs (solidity 0.9, umbral y nivel) para vecino más cercano.....	71
Tabla 41 Eficiencia vs (solidity 0.9, umbral y nivel) para vecino más cercano	72
Tabla 42 Exactitud vs (solidity 0.1, umbral y nivel) para árbol de decisión.....	73
Tabla 43 Eficiencia vs (solidity 0.1, umbral y nivel) para árbol de decisión	73
Tabla 44 Exactitud vs (solidity 0.2, umbral y nivel) para árbol de decisión.....	74
Tabla 45 Eficiencia vs (solidity 0.2, umbral y nivel) para árbol de decisión	74
Tabla 46 Exactitud vs (solidity 0.3, umbral y nivel) para árbol de decisión.....	75
Tabla 47 Eficiencia vs (solidity 0.3, umbral y nivel) para árbol de decisión	75
Tabla 48 Exactitud vs (solidity 0.4, umbral y nivel) para árbol de decisión.....	76
Tabla 49 Eficiencia vs (solidity 0.4, umbral y nivel) para árbol de decisión	76
Tabla 50 Exactitud vs (solidity 0.5, umbral y nivel) para árbol de decisión.....	77
Tabla 51 Eficiencia vs (solidity 0.5, umbral y nivel) para árbol de decisión	77
Tabla 52 Exactitud vs (solidity 0.6, umbral y nivel) para árbol de decisión.....	78
Tabla 53 Eficiencia vs (solidity 0.6, umbral y nivel) para árbol de decisión	78
Tabla 54 Exactitud vs (solidity 0.7, umbral y nivel) para árbol de decisión.....	79
Tabla 55 Eficiencia vs (solidity 0.7, umbral y nivel) para árbol de decisión	79
Tabla 56 Exactitud vs (solidity 0.8, umbral y nivel) para árbol de decisión.....	80
Tabla 57 Eficiencia vs (solidity 0.8, umbral y nivel) para árbol de decisión	80
Tabla 58 Exactitud vs (solidity 0.9, umbral y nivel) para árbol de decisión.....	81

Tabla 59	Eficiencia vs (solidity 0.9, umbral y nivel) para árbol de decisión	81
Tabla 60	Exactitud vs (solidity 0.1, umbral y nivel) para Naive Bayes	82
Tabla 61	Eficiencia vs (solidity 0.1, umbral y nivel) para Naive Bayes	83
Tabla 62	Exactitud vs (solidity 0.2, umbral y nivel) para Naive Bayes	83
Tabla 63	Eficiencia vs (solidity 0.2, umbral y nivel) para Naive Bayes	84
Tabla 64	Exactitud vs (solidity 0.3, umbral y nivel) para Naive Bayes	84
Tabla 65	Eficiencia vs (solidity 0.3, umbral y nivel) para Naive Bayes	85
Tabla 66	Exactitud vs (solidity 0.4, umbral y nivel) para Naive Bayes	85
Tabla 67	Eficiencia vs (solidity 0.4, umbral y nivel) para Naive Bayes	86
Tabla 68	Exactitud vs (solidity 0.5, umbral y nivel) para Naive Bayes	86
Tabla 69	Eficiencia vs (solidity 0.5, umbral y nivel) para Naive Bayes	87
Tabla 70	Exactitud vs (solidity 0.6, umbral y nivel) para Naive Bayes	87
Tabla 71	Eficiencia vs (solidity 0.6, umbral y nivel) para Naive Bayes	88
Tabla 72	Exactitud vs (solidity 0.7, umbral y nivel) para Naive Bayes	88
Tabla 73	Eficiencia vs (solidity 0.7, umbral y nivel) para Naive Bayes	89
Tabla 74	Exactitud vs (solidity 0.8, umbral y nivel) para Naive Bayes	89
Tabla 75	Eficiencia vs (solidity 0.8, umbral y nivel) para Naive Bayes	90
Tabla 76	Exactitud vs (solidity 0.9, umbral y nivel) para Naive Bayes	90
Tabla 77	Exactitud vs (solidity 0.9, umbral y nivel) para Naive Bayes	91
Tabla 78	Exactitud vs (solidity 0.1, umbral y nivel) para ANN con red Patternnet	92
Tabla 79	Eficiencia vs (solidity 0.1, umbral y nivel) para ANN con red Patternnet	92
Tabla 80	Exactitud vs (solidity 0.2, umbral y nivel) para ANN con red Patternnet	93
Tabla 81	Eficiencia vs (solidity 0.2, umbral y nivel) para ANN con red Patternnet	93
Tabla 82	Exactitud vs (solidity 0.3, umbral y nivel) para ANN con red Patternnet	94
Tabla 83	Eficiencia vs (solidity 0.3, umbral y nivel) para ANN con red Patternnet	94
Tabla 84	Exactitud vs (solidity 0.4, umbral y nivel) para ANN con red Patternnet	95
Tabla 85	Eficiencia vs (solidity 0.4, umbral y nivel) para ANN con red Patternnet	95
Tabla 86	Exactitud vs (solidity 0.5, umbral y nivel) para ANN con red Patternnet	96
Tabla 87	Eficiencia vs (solidity 0.5, umbral y nivel) para ANN con red Patternnet	96
Tabla 88	Exactitud vs (solidity 0.6, umbral y nivel) para ANN con red Patternnet	97
Tabla 89	Eficiencia vs (solidity 0.6, umbral y nivel) para ANN con red Patternnet	97

Tabla 90 Exactitud vs (solidity 0.7, umbral y nivel) para ANN con red Patternnet	98
Tabla 91 Eficiencia vs (solidity 0.7, umbral y nivel) para ANN con red Patternnet	98
Tabla 92 Exactitud vs (solidity 0.8, umbral y nivel) para ANN con red Patternnet	99
Tabla 93 Eficiencia vs (solidity 0.8, umbral y nivel) para ANN con red Patternnet	99
Tabla 94 Exactitud vs (solidity 0.9, umbral y nivel) para ANN con red Patternnet	100
Tabla 95 Eficiencia vs (solidity 0.9, umbral y nivel) para ANN con red Patternnet	100
Tabla 96 Eficiencia en algoritmos de aprendizaje de maquina	101
Tabla 97 Partes de una matriz de confusión	102
Tabla 98 Valores estadísticos aplicados en varios sistemas	103

ÍNDICE DE FIGURAS

Figura 1 a. En la primera región se presenta una masa benigna y en b) una masa maligna.....	2
Figura 2 Estructura interna de mama, enfocándose particularmente en los lobulillos.....	6
Figura 3 Mamografía.....	11
Figura 4 Procedimiento para realizar una mamografía.....	11
Figura 5 Imagen digital	14
Figura 6 a. Imagen original detectando cabeza y tórax b. División de imagen en tres áreas de interés c. Obtención de rostro	15
Figura 7 Ejemplo de Segmentación.....	16
Figura 8 Ejemplo de erosión	20
Figura 9 Ejemplo de dilatación.....	20
Figura 10 Funcionamiento de Machine Learning	22
Figura 11 Ejemplo de clasificación de datos aplicando técnicas de M.L.	22
Figura 12 Ejemplo de clasificación de Naive Bayes	25
Figura 13 Clasificación vecino más cercano	26
Figura 14 Ejemplo de clasificación vecino más cercano	26
Figura 15 Estructura y funcionamiento de árbol de decisiones.	28
Figura 16 Similitud entre neuronas biológicas y artificiales.....	28
Figura 17 Sistema de red multicapa para un sistema de una sola salida.....	29
Figura 18 Ejemplo de 4-fold validación cruzada.....	31
Figura 19 Curvas ROC	33
Figura 20 Diagrama de bloques del sistema	35
Figura 21 Imagen mamográfica de la base de datos en formato jpg.	36
Figura 22 Imagen mamográfica binarizada.	37
Figura 23 Imagen con áreas detectadas con medio de la función regionprop.....	38
Figura 24 Imagen resultante de esta etapa.	38
Figura 25 Inversión en la intensidad de la imagen.....	39
Figura 26 Diagrama de bloques de la segmentación.....	40

Figura 27 Imagen aplicada la técnica de segmentación de Otsu	41
Figura 28 Rellenar áreas de imagen	42
Figura 29 Imagen utilizando la función regionprops	43
Figura 30 Imagen de entrada para operaciones morfológicas.....	44
Figura 31: Implementación de dilatación	45
Figura 32 Eliminación de área de no interés	45
Figura 33 Aplicación de erosión.....	46
Figura 34 Resultado final de la segmentación.....	47
Figura 35 Primera característica por sumatoria de pixeles diferentes de blanco	48
Figura 36 Segunda característica por promedio de pixeles	48
Figura 37 Tercera característica se aplicó la varianza	49
Figura 38 Obtención de resultados.....	50
Figura 39 Sección de código que genera grupo de prueba y entrenamiento	51
Figura 40 Valor de exactitud vs k-iteraciones.....	52
Figura 41 Curva ROC para ANN (Fitnet)	103
Figura 42 Curva ROC para ANN (Patternnet)	104
Figura 43 Curva ROC para vecino más cercano.....	104
Figura 44 Curva ROC para árbol de decisiones.....	105
Figura 45 Curva ROC para Naive Bayes	105
Figura 46 Variación de umbral para datos que poseen error en la detección	106

RESUMEN

La elaboración del presente trabajo de investigación tiene como finalidad la identificación del cáncer de mama en imágenes mamográficas, debido a que esta patología afecta a una gran cantidad de mujeres en nuestro país. Para llegar a la identificación de células cancerígenas en la mamografía, es necesario un análisis previo en el área de procesamiento digital de imágenes y especialmente en técnicas de segmentación, las mismas que ayudan a buscar e identificar las regiones de interés por medio de la eliminación de pixeles adyacentes. Las técnicas de segmentación que se implementaron para la detección de cáncer son la dilatación y erosión que se fundamentan en la simplificación de las imágenes por medio de la forma geométrica que dichas células poseen. Ya obtenidas las regiones establecidas por las operaciones morfológicas se procedió a identificar la presencia de cáncer o no, mediante el uso de algoritmos de aprendizaje de máquina, los cuales son aptos en crear sistemas autónomos de decisión dependientes del modelamiento de la base de datos, para generar un resultado que se verifica con el análisis de exámenes de mamografías existentes, utilizando así diferentes algoritmos de aprendizaje de máquina como son: árbol de decisiones, bayes ingenuo, vecino más cercano y redes neuronales artificiales (ANN). Los resultados obtenidos en la etapa anterior, fueron analizados cuidadosamente mediante el uso de variables como: la exactitud, la sensibilidad, la eficiencia, etc., llegando a encontrar en todos los sistemas propuestos un valor cercano al 98% de exactitud en la identificación de cáncer de mama.

Palabras clave:

- **CÁNCER DE MAMA**
- **MAMOGRAFÍAS**
- **PROCESAMIENTO DIGITAL DE IMÁGENES**
- **APRENDIZAJE DE MÁQUINA**

ABSTRACT

The purpose of this research is to identify breast cancer in mammographic images, because this pathology affects a large number of women in our country. In order to arrive at the identification of cancer cells in the mammography, a previous analysis is necessary in the area of digital image processing and especially in segmentation techniques, the same ones that help to search and identify the regions of interest by means of the elimination of adjacent pixels. The techniques of segmentation that were implemented for the detection of cancer are the dilatation and erosion that are based on the simplification of the images by means of the geometric form that these cells possess. Once the regions established by the morphological operations, the presence of cancer was identified, using machine learning algorithms, which are able to create autonomous decision systems dependent on the modeling of the database, to generate a result that is verified by the analysis of existing mammography exams, using different machine learning algorithms such as: decision tree, naive bayes, nearest neighbor and artificial neural networks (ANNs). The results obtained in the previous stage were carefully analyzed through the use of variables such as: accuracy, sensitivity, efficiency, etc., reaching in all proposed systems a value close to 98% accuracy in the identification of Breast cancer

Keywords:

- **BREAST CANCER**
- **MAMMOGRAPHIES**
- **DIGITAL IMAGE PROCESSING**
- **MACHINE LEARNING**

CAPÍTULO I

1. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

1.1. Antecedentes

El cáncer de mama es responsable del cambio estructural y bioquímico de las funciones de las células y tejidos de los conductos mamarios, por consiguiente es la principal causa de cáncer, tanto en países en desarrollo como subdesarrollados. Así, en el Ecuador, la tasa de mortalidad que registra el Ministerio de Salud Pública, a causa del cáncer de mama, lo ubica como la décimo cuarta causa, y la tercera con relación a otros tipos de cáncer como el de cérvix y el de estómago (Organizaion World Health, 2016, pág. 4).

La vulnerabilidad que presentan los sistemas de salud, en el caso de los países subdesarrollados, debido a la escasez de sus recursos económicos, repercute en diagnósticos tardíos. El cáncer de mama es una de las afectaciones más tratadas por los especialistas del área de oncología, de manera que; la aparición de masas o tejidos benignos, tumores malignos, son los principales temas de estudio, ya que por la dificultad que representa la ejecución de un diagnóstico previo, son identificados en etapas tardías.

El índice de fallecimiento según el Instituto Nacional de Estadísticas y Censos (INEC), ubicó esta cifra durante el año 2010, en un total de 470 mujeres fallecidas, teniendo en cuenta que en la ciudad de Quito se encuentra la mayoría de casos de cáncer de mama, así las cifras rebelaron que durante el año 2008 a 2011, 36 de cada 100.000 quiteñas fueron diagnosticadas con este tipo de patología (Ministerio de Salud Publica, 2015).

Para Laya, D. el cáncer es “una masa anormal que se genera repentinamente en un tejido corporal” (Laya D., et al., 2008). El presente estudio está enfocado en el cáncer de mama, el cual se puede identificar a través de la mamografía, donde se lo observa como una mancha subnormal, de forma

irregular, con carácter benigno o maligno, como se observa en la Figura 1 (Yanfeng Li, 2016, pág. 4).

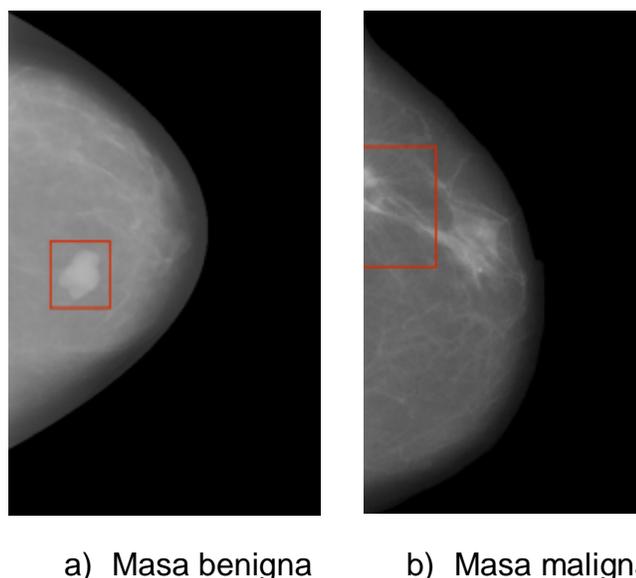


Figura 1 a. En la primera región se presenta una masa benigna y en b) una masa maligna.

Fuente: (Yanfeng Li, 2016, pág. 4)

La peligrosidad y celeridad con la que el cáncer ataca, requiere de un pronóstico rápido, preciso y efectivo, siendo la técnica más común para la detección del cáncer de mama el tamizaje de biopsias. Sin embargo, este examen médico, por su costo y procedimiento solo se lo realiza después de que el galeno valora la mamografía y tiene sospechas de la presencia de alteraciones en la imagen que pueden ser producto de la formación de un tumor.

Los avances tecnológicos en la investigación médica permiten desarrollar técnicas para el diagnóstico y la identificación del cáncer de mama, a través de la generación de grandes bases de datos que pueden ser utilizadas por otros profesionales para realizar nuevas investigaciones.

La mamografía digital permite la detección del cáncer de mama mediante la utilización de algoritmos para el procesamiento de imágenes digitales, mostrando así áreas sospechosas en la mama, los mismos que permiten evidenciar la

presencia de masas u otros signos prominentes de cáncer, lo que hace que se reduzca el número de casos de tumores malignos mal clasificados, a esto se conoce como segmentación de imágenes. Con la segmentación se pueden eliminar las áreas poco importantes en la región de la mama cercana al borde de la misma y el límite entre el músculo pectoral segmentado y el tejido mamario, en consecuencia esto hace que la estimación y el cálculo de las características sean más precisas.

Otras técnicas usadas son las de Máquinas de Vector Soporte (*Support Vector Machines*, SVMs), que con ayuda de funciones como *Ripley's*, permiten medir distancias a partir de datos conocidos, generando una clasificación y obtención de características, con resultados de por encima del 86% de exactitud en casos de estudio de mamografías que podrían presentar cáncer de mama, (Vierada, 2012). El presente estudio permitirá alcanzar un mayor índice de exactitud y en una menor escala de tiempo.

1.2. Justificación e Importancia

El implementar un sistema de reconocimiento de cáncer de mama brindará la oportunidad de detectar de manera confiable, segura, precisa, automática y con bajo costo la enfermedad. Lo que convertirá a éste sistema en un complemento de la evaluación médica que permitirá disminuir el error en el diagnóstico y acelerar la implementación de un tratamiento contra este grave padecimiento.

El proyecto cuenta con una amplia base de datos, como son las imágenes mamográficas que han sido diagnosticadas con cáncer, las mismas que son analizadas exhaustivamente en esta investigación y en base a éste estudio se ha diseñado el proyecto, que buscará más que nada beneficiar a mujeres de todo estrato social que atraviesan por esta grave enfermedad la misma que a más de problemas de salud ocasiona en la mayoría de veces también problemas económicos, sociales y personales.

El estudio propuesto podrá ser utilizado por cualquier profesional de la medicina, por su fácil manejo, por la versatilidad al momento de comparar imágenes, por la

eficiencia del proceso, por la rápida entrega de resultados confiables. La investigación se presentará a la sociedad como una contribución, por todo lo recibido de ella y como un pequeño aporte al avance de la tecnología en nuestro país.

1.3. Alcance del proyecto

El presente trabajo permitirá la identificación del cáncer de mama, aplicando técnicas de aprendizaje de máquina. Para realizar esta investigación se ha recurrido a la utilización de imágenes con características de alta resolución (1280x720 píxeles y 1920x1080 píxeles). Las imágenes han sido analizadas por profesionales de la medicina, donde se ha determinado la presencia o el descarte del cáncer de mama. El proyecto en su primera etapa inicia con el ingreso de la imagen original que fue extraída de la base de datos digital del repositorio de dominio público; “*Digital Database for Screening Mammography (DDSM)*” de la Universidad de Florida, (Universidad del Sur de Florida, 2006). La base de datos cuenta con 236 imágenes, dando un análisis más amplio y profundo al presente estudio. Así también, la variable de identificación que provee la base de datos, permite verificar los resultados médicos y compararlos con los obtenidos mediante el presente sistema permitiendo una validación del diagnóstico médico en la detección de cáncer de mama.

El procesamiento de las imágenes mamográficas tiene la finalidad de eliminar aquellos elementos que pueden distorsionar el proceso de identificación de micro calcificaciones reduciendo el área de trabajo y concentrar el estudio en la región de la mama donde se localiza el tumor; gracias a esto se obtendrán las características para el aprendizaje automático a través de la herramienta de análisis matemático como *MatLab*[®].

El presente estudio evalúa la eficiencia de las diferentes técnicas de predicción como: diagrama de árbol, vecino más cercano, redes neuronales artificiales, bayesiano ingenuo. Con la finalidad de obtener valores altos de

eficiencia, se generará una cantidad adecuada de características por medio de diferentes escenarios que pondrán a prueba a las mismas, ejemplo de estas pueden ser la varianza, la media, técnicas de segmentación espaciales que permiten un buen procesamiento para cada mamografía y otras técnicas que puedan aportar al sistema.

En base a los criterios seleccionados se podrá cuantificar la robustez de cada sistema para el reconocimiento del cáncer y generar un programa con alto grado de credibilidad para el personal que lo requiera justificando los resultados mediante un análisis de curvas y otros factores que se ampliarán en los siguientes capítulos.

1.4. Objetivos

1.4.1. Objetivo general

Implementar un sistema de reconocimiento de cáncer de mama utilizando técnicas de procesamiento digital de imágenes y *Machine Learning*.

1.4.2. Objetivos Específicos

- Estudiar y analizar los trabajos relacionados en técnicas de procesamiento de imágenes y aprendizaje de máquina para tener un enfoque más preciso al momento del diseño del sistema.
- Diseñar un sistema de identificación de cáncer de mama utilizando técnicas de segmentación y algoritmos de aprendizaje de máquina.
- Evaluar las propiedades de eficiencia, sensibilidad y matriz de confusión de cada sistema planteado.

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Cáncer de Mama

La glándula mamaria está compuesta por un conjunto de elementos en forma de racimos, denominados lóbulos mamarios, De ahí que, los lóbulos representan la estructura funcional de la glándula permitiendo producir la leche materna a través de células especiales que se albergan en su interior. Debido a esta conexión y estructura que presenta la mama hace que esta zona sea vulnerable, esto se ver mejor en la figura 2 (Sánchez, 2011, pág. 7).

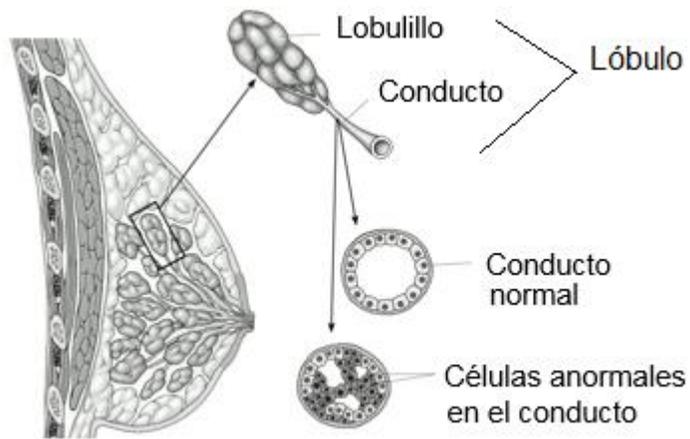


Figura 2 Estructura interna de mama, enfocándose particularmente en los lobulillos.

Fuente: (Yanfeng Li, 2016, pág. 4)

Según la Sociedad Española de Oncología Médica el “cáncer del seno es una enfermedad oncológica donde las células sanas se degeneran atacando principalmente a las glándulas mamarias formado zonas tumorales que, con el paso del tiempo llegan a expandirse por todo el cuerpo y provocando la muerte” (Roche Farma, 2011).

Los tipos de cáncer de mama que existen son generalmente por aspectos biológicos o genéticos de las células, a continuación se define las particularidades que presentan estos tipos de cáncer en la Tabla 1, donde se realiza un análisis por el grado de invasión, lugar donde se desarrolla, y sus características biológicas y genéticas de las células.

Tabla 1
Tipos de cáncer de mama

GRADO DE INVASIÓN DEL TUMOR	In situ: Ataca los conductos lácteos que conducen la leche. Infiltrante: Es el más general, y se propaga hacia los tejidos mamaros, la grasa de la mama, y sus conductos que lo rodean.
LUGAR DONDE SE FORMA	Lobular: Es el segundo tipo de cáncer más general, y se origina en los conductos lácteos. Ductal: Se origina en los revestimientos del conducto de la mama. Medular: Las células cancerosas se encuentran agrupadas, se observan como células cancerosas agresivas, pero no actúan. Coloide: Es un tipo de carcinoma ductal infiltrante. Tubular: Comienza dentro de los conductos lácteos, propagándose fuera de estos. Inflamatorio: Bloquean vasos linfáticos de la piel inflamando la mama, es de propagación rápida. Musinos: se forma en la acumulación de mucina, conocida como moco, normalmente esta mucosidad recubre la mayor parte del cuerpo humano como: tubo digestivo, pulmones, hígado, etc.
CARACTERÍSTICAS BIOLÓGICAS Y GENÉTICAS DE LAS CÉLULAS	Cribiforme: Invaden los tejidos conductivos de la mama, donde estas células, se ven y comportan de forma similar a las células sanas y normales. Paget: en este tipo de cáncer, las células cancerosas, se acumulan en el pezón, propagándose a la aureola. Metastático: es un tipo de cáncer recurrente, en el cual las células benignas han vuelto a aparecer.

Fuente: (Roche Farma, 2011)

2.1.1. Factores de riesgo

Existen determinados factores que pueden aumentar el riesgo de padecer cáncer de seno. Donde la enfermedad puede afectar a cualquier tipo de persona, siendo más vulnerable para el caso de la mujer, en gran medida debido a las hormonas femeninas denominados estrógenos, las cuales pueden promover el desarrollo de células cancerosas. Entre los principales factores que propician la aparición de cáncer de mama, se destaca las inherentes a su naturaleza donde en gran medida se deben a factores genéticos. El estilo de vida es otra de las causales para el cáncer de mama, determinado por el bajo nivel en los buenos hábitos de vida. Existe otro tipo de factores que repercuten, los cuales se han clasificado como efectos no definidos, los mismos que se mencionan y se explican de mejor manera en la Tabla 2 (American Cancer Society, 2016).

Tabla 2
Factores que propician el cáncer

INHERENTE A SU NATURALEZA	<ul style="list-style-type: none"> • Ser mujer, debido a los cambios de las hormonas femeninas. • Envejecimiento, con la edad las células se debilitan y se vuelven más vulnerables. • Por genética hereditaria, por la mutación hereditaria de los genes. • Raza y origen étnico, las mujeres blancas poseen más riesgo • Comienzo temprano de la menstruación.
RELACIONADOS CON EL ESTILO DE VIDA	<ul style="list-style-type: none"> • Consumo de bebidas alcohólicas • Consumo de cigarrillo • Sobrepeso u obesidad • Por control de natalidad, el uso de anticonceptivos de vía oral puede aumentar este riesgo • El alto grado de estrés
POR EFECTOS NO DEFINIDOS	<ul style="list-style-type: none"> • Alimentación y vitaminas, por una inadecuada alimentación y desequilibrios en la dieta. • Químicos en el ambiente

Fuente: (American Cancer Society, 2016, pág. 4)

2.1.2. Diagnóstico

Con el objetivo de verificar si un paciente puede poseer cáncer de mama, existen algunas técnicas, procedimientos e instrumentos que permiten determinar la presencia de agentes extraños como tumores y que mediante la intervención de especialistas de la medicina en el campo de la oncología, reconocen si estas presencias son benignas o malignas. Entre los principales se destaca el examen físico, el cual permite la identificación de protuberancias, también está la ecografía, donde técnicas de ondas sonoras permiten la identificación de cambios, otra es la biopsia que recurrirá al análisis microscópico, y finalmente están la resonancia magnética nuclear y las mamografías. A continuación en la tabla 3 se describe con mayor profundidad cada tipo de procedimiento que se implementa en el diagnóstico de cáncer de mama (American Cancer Society, 2016).

Tabla 3

Tipos de procedimientos para el diagnóstico de cáncer de mama.

TIPOS	DESCRIPCIÓN
EXAMEN FÍSICO	Mediante un examen minucioso de los senos, que pretende la identificación de protuberancias, se realiza una palpación, identificando su textura, tamaño, etc. Esta técnica es recomendable realizarla a partir de los 20 años.
ECOGRAFÍA	Denominada como monografía o ultrasonido, esta técnica utiliza ondas sonoras para delinear el cuerpo, permitiendo identificar cambios del seno, ya sea por la presencia de líquidos o masas sólidas, se utiliza en pacientes menores a 30 años de edad.
BIOPSIA	Después de haber detectado las células de cáncer se realiza un examen del tejido extraído, el cual será analizado y observadas a través del microscopio. Para la obtención de esta muestra se realiza una intervención quirúrgica, de manera que; esté procedimiento permita verificar el tipo de cáncer y su posible tratamiento.
RESONANCIA MAGNÉTICA NUCLEAR	Permite obtener imágenes de la mama utilizando ondas de radiofrecuencia generando una mayor resolución en los resultados, donde una computadora permitirá la traducción de los patrones de manera muy detallada. El proceso es efectivo, pero de alto costo.
MAMOGRAFÍAS	Denominado a la radiografía que se realiza al seno, la toma del monograma se lo realiza desde dos ángulos diferentes, y para la detección del cáncer en pacientes asintomáticos.

Fuente: (Roche Farma, 2011)

2.1.3. Mamografías

Las mamografías permiten obtener imágenes donde las partes más densas del cuerpo en estudio, aparecen en diferentes tonos y en escalas de todos grises; permitiendo la identificación de anomalías (Universidad del Sur de Florida, 2006).

En la Figura 3 se puede observar una imagen radiográfica del seno, esta técnica muestra sencillez y bajo costo comparado con otras técnicas.

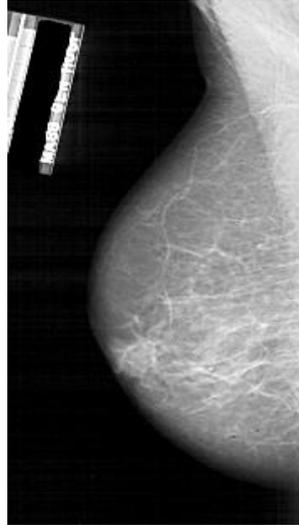


Figura 3 Mamografía

Fuente: (Universidad del Sur de Florida, 2006)

Para obtener las imágenes se utiliza un mamógrafo, el cual es un dispositivo médico en donde se coloca el seno, este se ubica bajo una placa plana y esta sobrepuesto por otra placa superior que permite la sujeción. La máquina realiza una baja concentración de rayos X atravesando la mama dejando señales que detecta la placa opuesta generando una captura fotográfica de carácter digital que se trasmite a un computador. En la Figura 4 se puede observar el equipo con el cual se realiza esta técnica.

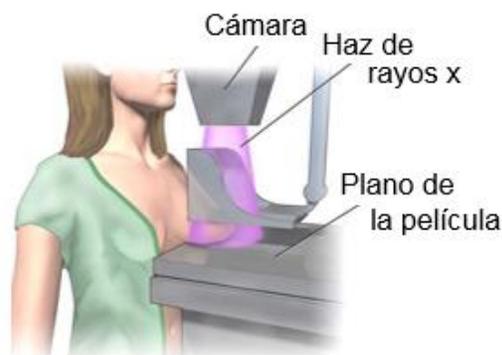


Figura 4 Procedimiento para realizar una mamografía

Fuente: (OPENLAB at City Tech, 2014)

Las imágenes mamográficas, permiten evidenciar la existencia de posibles tumores (Susan G, 2013). La detección temprana de dicho padecimiento contribuye a diagnósticos tempranos con el fin de poder salvar una vida.

Según la Sociedad de Radiología de Norte América “La mamografía es el sistema comúnmente más utilizado en la detección asistida por computadora” (Radiological Society of North America, 2015). Este examen permite la búsqueda de áreas con alta probabilidad de tener calcificaciones o quistes. Sin embargo este sistema debe poseer la asistencia del radiólogo que revisara cuidadosamente los resultados, la verificación de la mamografía queda limitada a la experticia del mismo.

En el presente estudio, las mamografías representan un instrumento primordial y necesario para implementarse en el estudio desarrollado. Motivo por el cual fue esencial recurrir a la base de datos de DDSM: “*Digital Database for Screening Mammography*”, presente en la Universidad del Sur de Florida (Universidad del Sur de Florida, 2006). De modo que, el proyecto tiene como finalidad ser un sistema capaz de identificar masas malignas en desarrollo, y que pueda servir como un soporte a los médicos oncológicos en mejorar su experiencia para no obviar un caso de cáncer de mama prominente.

2.2. Procesamiento Digital de Imágenes

El procesamiento digital de imágenes es un conjunto de procedimientos que tienen como finalidad modificar una imagen digital, para poder mejorar su calidad, cambiar sus dimensiones, u otro tipo de información que se desea obtener, logrando así una nueva imagen. De ahí que, el Procesamiento de Imágenes Digitales (PDI) tuvo una gran evolución en el año de 1959, donde la NASA al requerir enviar imágenes desde el espacio hacia la tierra en tiempos cortos y con alta definición, recurrió al desarrollo y procesamiento digital. (ILCE, 2011).

Esta rama se desarrolla en algunas ciencias como: informática, electrónica, medicina, óptica, matemáticas, etc. Uno de los factores más importantes para su uso, ha sido la mejora en el costo computacional y el bajo coste en su implementación de estos sistemas. Para aclarar mejor este concepto se recurrirá a comprender que es una imagen digital.

2.2.1. Imagen digital

La cámara fotográfica de Joseph-Nicéphore Niépce en 1826, fue el invento revolucionario de aquella época, dando origen a la fotografía y a la conceptualización de imagen. Una imagen no es más que “una representación o imitación de un objeto o cosa” (Sheets, 2013). La imagen es un generador de información que intenta representar la realidad que nuestros ojos ven, teniendo en mente que esta representación es ubicada en un determinado tiempo y espacio, haciendo que perdure a lo largo de los años.

La imagen digital está conformada por elementos denominados píxeles, los cuales individualmente tienen un valor propio de intensidad o brillo y al agruparlos en una matriz ordenada pueden formar la percepción de una imagen. Los valores de intensidad en las imágenes digitales se definen mediante bits. Un bit tiene dos valores posibles 0 o 1. Comúnmente los bytes están agrupados en cadenas de 8 bits los cuales pueden formar un rango de intensidad de 256 valores posibles, lo que hace que el usuario forme una idea de las posibles tonalidades. Al agrupar los píxeles se puede formar la idea de imagen, esto se aprecia de mejor manera en la figura 5 (Aparici, García, Fernández, & Osuna, 2009).

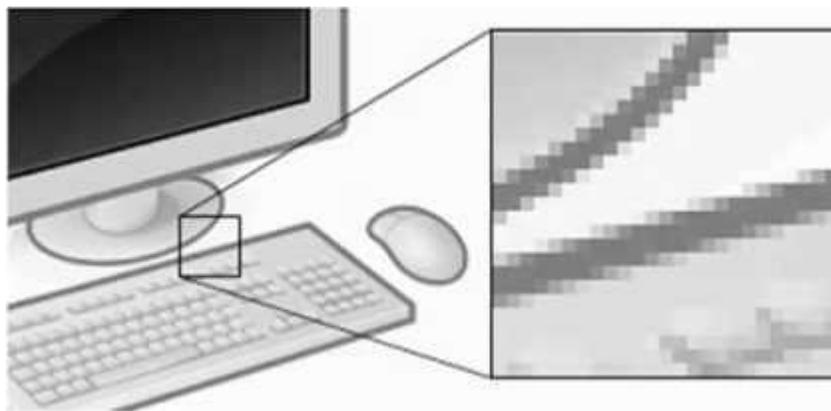


Figura 5 Imagen digital

Fuente: (Chaves Hernandez, 2011)

Las imágenes digitales pueden ser clasificadas por el número de matrices que contenga cada imagen, si consta de una matriz se las denomina imágenes monocromáticas, las cuales forman una escala de un solo color y si es de dos o más matrices se las denomina imágenes cromáticas, las mismas que forman una escala con la mezcla de más colores. Para el presente estudio utilizaremos imágenes monocromáticas en escala de grises debido a la facilidad que dichas imágenes brindan.

2.2.2. Imagen en escala de grises

Estas imágenes solo poseen colores en escalas de grises, es decir que se encuentran en una sola matriz de dos dimensiones y se componen de 256 posibles tonos de gris desde el negro al blanco, ya que cada pixel tiene un valor de luminancia único, dependientemente de un solo color. (Fisher, Perkins, & Wolfart, 2003). Debido a la facilidad que estas imágenes brindan son parte fundamental en la implementación de la etapa de segmentación, la cual se explica de mejor en las siguientes secciones.

2.2.3. Segmentación

La segmentación es un conjunto de técnicas que permiten dividir, eliminar, o sobreponer regiones, áreas o superficies que sean de interés dentro de una imagen. Permitted simplificar o aislar características esenciales de una imagen y así contribuir a la solución de un problema. La segmentación es comúnmente utilizado en sistemas de movimiento o estéreo, compresión de imágenes, edición de imágenes o búsqueda de bases de datos de imágenes (Jepson & Fleet, 2007). La segmentación es el primer paso del proceso de una imagen de bajo nivel a transformar a un alto nivel de descripción de la imagen en términos de características, objetos y escenas.

El ejemplo de la figura 6 tiene como objetivo identificar y segmentar el rostro de la persona en la fotografía, como primer paso detecta a la persona y divide la figura en dos regiones, la cabeza y el tórax como se visualiza en figura 6a, para luego dividir la misma imagen en tres regiones como vemos en figura 6b, en este caso utiliza azul para el rostro, verde para el tórax y rojo para el fondo, segmentadas estas regiones procede en la figura 6c a seleccionar y utilizar solo el color azul para obtener el rostro.

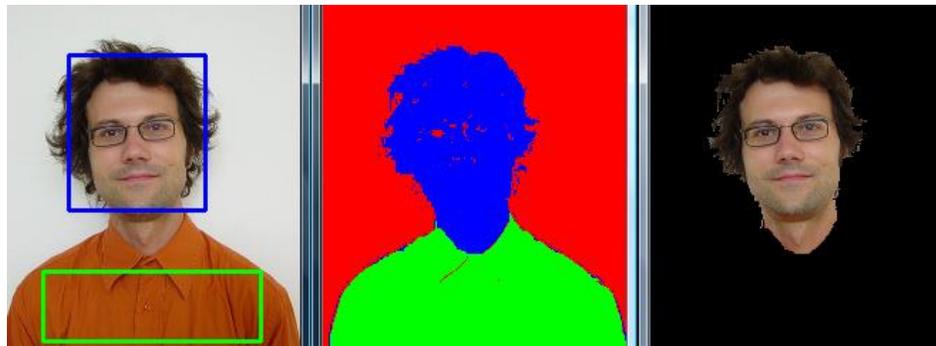


Figura 6 a. Imagen original detectando cabeza y tórax b. División de imagen en tres áreas de interés c. Obtención de rostro

Fuente: (Roy, 2010)

2.2.3.1. Segmentación por Umbralización

Es un método de segmentación que permite a una imagen con diferentes tonalidades, ser convertida a una imagen en escala de grises. En consecuencia, esta técnica es una de las más usadas, debido a que, permite comparar cada píxel de la imagen con un nivel de umbral. El índice umbral se puede buscar mediante el histograma de imagen o la intensidad del píxel ubicado en la zona de mayor interés. Una vez establecida el área que no forma parte de la segmentación se recomienda cambiar el valor umbral evitando que los pixeles se mezclen y se generen errores en etapas posteriores (UNAM, 2010).

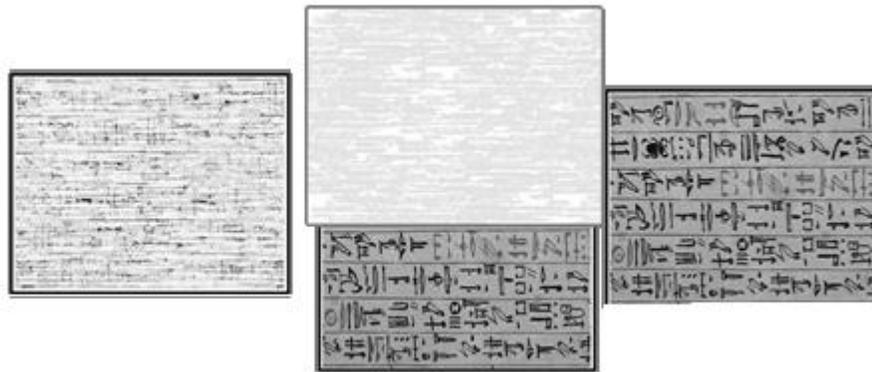


Figura 7 Ejemplo de Segmentación

Fuente: (Williams, 2010)

Como se puede observar en el ejemplo de la figura 7, la primera toma muestra una imagen con ruido, al utilizar la técnica de umbralización se logra dividir la imagen original en dos, la primera imagen que se presenta es la deseada, la segunda imagen representa el ruido que se puede detectar mediante el valor umbral

La segmentación por umbralización, está definida por la ecuación fundamental:

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

Así la umbralización se basa en el análisis de uno o varios umbrales, permitiendo la agrupación de los puntos que conforman la imagen, estas regiones poseen similares y determinadas características que están en función de sus niveles de gris.

- T : Representa el índice umbral.
- $f(x, y)$: Representa la intensidad o nivel de gris, es decir la imagen original representada en forma matricial.
- $g(x, y)$: Define la imagen que se obtendrá al realizar un proceso umbral. De modo que los pixeles marcados con 1 representan objetos, y los que se encuentran marcados con 0 pertenecen al fondo.

2.2.3.2. Método de Otsu

Este método fue desarrollado por Nobuyuki Otsu en el año de 1979, desde sus inicios este método ha sido de amplio uso por parte de la industria, el método permite la determinación automática de umbral de segmentación, de modo que; se aplica a objetos que sobresalen y se diferencian claramente de la imagen original. Su principio fundamental se basa en la similitud que pueden poseer sus áreas y las características comunes que tengan (Quilmes, 2005).

El método Otsu se usa comúnmente para encontrar valores umbrales de forma no paramétrica, debido a que varía sus parámetros estadísticos permitiendo maximizar las varianzas para encontrar el mejor umbral y seleccionar con mayor precisión los pixeles deseados. El método se aplica principalmente a imágenes que poseen un alto porcentaje de error o mala iluminación (Andrew, 2010, pág. 4).

2.2.3.3. Operaciones morfológicas

La morfología se basa en la geometría y forma de las áreas de interés mediante un procesamiento no lineal, permitiendo preservar las características esenciales eliminando aspectos que puedan ser irrelevantes. La morfología se fundamenta en la teoría matemática de los conjuntos que se encuentran como espacios matriciales binarios (Universidad Politécnica de Madrid, 2010).

La morfología matemática, es ampliamente utilizada en:

- La generación de límites o bordes en zonas de interés.
- Adelgazar, ampliar o reducir secciones
- Como filtro morfológico en la supresión de ruido
- Simplificación de formas
- Extracción los esqueletos
- Marcado de objetos,
- Generar cerco convexo

2.2.3.4. Operaciones morfológicas básicas

La morfología tiene su fundamentación matemática en la teoría de conjuntos, a continuación en la tabla 4 se puede observar sus principales operaciones básicas, las cuales son el fundamento teórico para el uso de operaciones más complejas como la erosión y la dilatación, de las que se habla en secciones posteriores.

Tabla 4
Operaciones morfológicas

OPERACIONES MORFOLÓGICAS BÁSICAS		
OPERACIONES	SIMBOLOGÍA	CONCEPTO
Unión	$A \cup B$	La unión es una operación que permite crear un nuevo conjunto a partir de todos los elementos que conformaban los dos anteriores
Intersección	$A \cap B$	Esta operación permite seleccionar elementos comunes entre los conjuntos A y B
Diferencia	$A - B$	Permite seleccionar los elementos que pertenecen A pero no B
Complemento	A^c	En esta operación se genera un nuevo conjunto que contiene todo el universo excepto los elementos que se encuentran en A

Fuente: (Gonzales & Woods, 2006)

2.2.3.5. Erosión

La erosión permite la disminución y minimización de los bordes en los objetos de interés, dependiendo fundamentalmente del tipo de estructura pudiendo ser: círculos, cuadrados, rombos, triángulos o cualquier polígono regular o irregular. La erosión es el conjunto resultante a partir de la comparación de los elementos que se encuentran en X con respecto a Y eliminando como conjunto vacío a los valores no seleccionados.

En imágenes en escala de grises en el cual el proyecto se enfoca, la erosión “es el mínimo valor de la función imagen en la ventana (dominio) definida por el elemento estructural cuando su origen se sitúa en x (señales unidimensionales) o (x, y) (imágenes)” (Ortiz, 2002). Además en estas imágenes el menor valor (negro) es aquel en el cual se trabaja debido a que existe mayor cantidad en las imágenes. En la Figura 8 podemos ver un ejemplo al aplicar esta técnica morfológica.

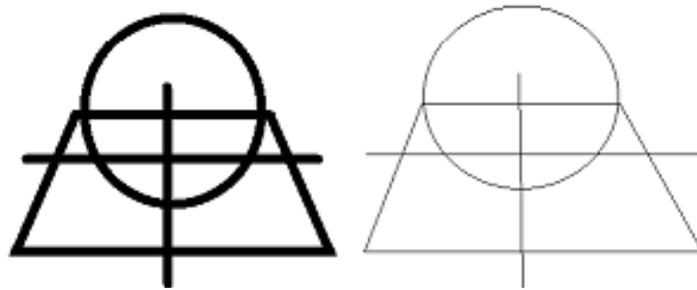


Figura 8 Ejemplo de erosión

2.2.3.6. Dilatación

Esta operación consiste en aumentar las dimensiones de las zonas de interés dependiendo fundamentalmente de la cantidad y forma seleccionada para la aplicación de ese sistema, como sabemos la dilatación es todo lo contrario a la erosión pero las dos son necesarias para seleccionar dichas áreas. Esta operación se la usa habitualmente en la unión de áreas de interés que tengan características comunes.

La dilatación en imágenes de escalas de grises “es generalmente, una señal de mayor valor, es decir, una imagen más clara, puesto que la dilatación maximiza el valor de la señal” (Ortiz, 2002). El procedimiento que se desea llegar en aumentar la definición de luminosidad de áreas claras con respecto a las oscuras. A continuación indicaremos un ejemplo de esta operación en la Figura 9.

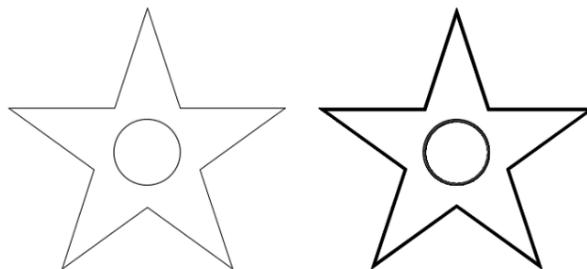


Figura 9 Ejemplo de dilatación

2.3. Machine Learning

El Aprendizaje de Máquina (ML) “es un proceso computacional que utiliza valores o conocimientos previos para generar resultados predictivos o clasificados” (Shai & Shai, 2014). El programa posee un aprendizaje de máquina, lo cual permite que sus resultados sean autónomos. De manera que, entre sus ventajas tenemos el no requerir de una programación preestablecida para el desarrollo de sus tareas así como, el utilizar modelos y códigos pequeños.

El aprendizaje de máquina tiene como objetivo utilizar características adecuadas para construir modelos adecuados que logren tareas correctas, estas tareas se aprenden de una manera supervisada que requiere datos de entrenamiento etiquetados. Como se menciona anteriormente las características tienen un papel muy importante ya que ningún modelo del aprendizaje automático puede existir sin características, ya veces una sola característica es suficiente para construir un modelo. Los datos no siempre vienen con características ya hechas, casi siempre tenemos que transformar características, debido a lo mencionado anteriormente el ML es a menudo un proceso iterativo, hasta encontrar satisfactoriamente los resultados deseados en el análisis.

La introducción de conocimientos previos en el aprendizaje de máquina, influye en el desarrollo de su proceso y son inevitables para el éxito de un algoritmo. Así, cuanto más fuerte es el conocimiento previo con el que se dará inicio al proceso de aprendizaje, más fácil resultará el aprendizaje.

2.3.1. Funcionamiento de Machine Learning

En la figura 10 se ejemplifica mediante un esquema, el funcionamiento del aprendizaje de máquina, la primera fase es la externa en la misma utilizamos el conocimiento del dominio para extraer nuevas características que serán utilizados por el algoritmo de aprendizaje automático, en la segunda fase se

encuentran las características las mismas que indicarán valores o etiquetas que son propias de cada elemento de la base de datos, como tercera etapa se encuentra el algoritmo del aprendizaje de máquina, el mismo que se explicará de mejor manera en secciones posteriores, y como fase final el sistema nos indicará a que clase pertenece los datos que están a prueba.

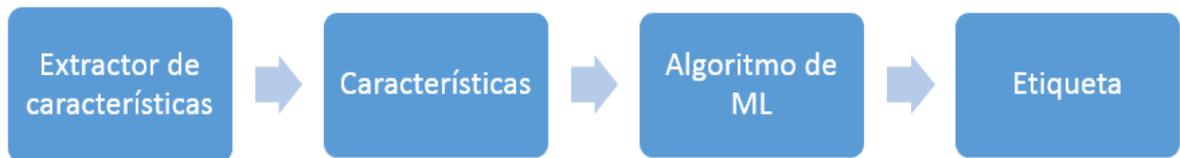


Figura 10 Funcionamiento de Machine Learning

Fuente: (Moujahid, 2016)

El aprendizaje de máquina utiliza algoritmos matemáticos que permiten la división o agrupación de valores de características similares y que están presentes en la base de datos y tienen como objeto la generación de patrones para facilitar un mejor resultado mediante la comparación de nuevos datos y los ya existentes. La figura 11 ilustra un ejemplo de clasificación de datos en un sistema bidimensional de dos conjuntos.

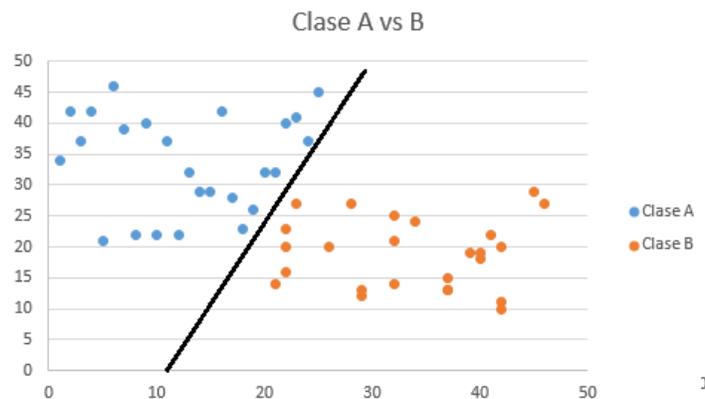


Figura 11 Ejemplo de clasificación de datos aplicando técnicas de M.L.

2.3.2. Tipos de algoritmos

El desarrollo de algoritmos matemáticos a lo largo del tiempo abarca un amplio dominio, permitiendo agruparlos en dos aprendizajes como son: supervisados y no supervisados. El presente estudio se concentra en el aprendizaje supervisado, ya que la investigación se basa en la premisa de la presencia de cáncer que aporta la base de datos y las dificultades que representa el desarrollo de una investigación multidisciplinar que requiere del conocimiento especializado de profesionales de la medicina que validan los resultados obtenidos. (Ben-David, 2014, pág. 50). En las siguientes secciones se explica de mejor manera el aprendizaje supervisado y sus diferentes clasificaciones.

2.3.2.1. Aprendizaje supervisado

El aprendizaje supervisado permite desarrollar un algoritmo computacional, a partir de una base de datos preestablecida conocida como: datos de entrenamiento. Generalmente la metodología del aprendizaje supervisado se aplica en el procesamiento de señales siendo fundamental para datos de imágenes.

El aprendizaje supervisado trabaja de tal forma que, “usa la experiencia para adquirir experiencia” de modo que, el aprendizaje supervisado describe un escenario en el que la “experiencia” es un ejemplo de entrenamiento, contiene información significativa permitiendo ser aplicada a un caso de estudio y adquirir más experiencia de lo ejecutado (Cunningham & Cord, 2008)

2.3.2.1.1. Técnica tipo supervisado Bayesiano Ingenuo

La técnica de bayesiano ingenuo, aplica un método probabilístico que consiste en la generación de un modelo de clasificación de los datos a partir del teorema de Bayes o por la distribución de probabilidad de dichos datos, este sistema soporta atributos continuos como discretos (Bonilla & Ojeda, 2006).

Esta técnica permite la estimación de las medias o varianzas de cada grupo de datos para generar funciones de distribución independientes entre sí que clasifican cada elemento por medio de su máxima probabilidad de ocurrencia presente en cada una de sus clases. El teorema fue publicado por Thomas Bayes y se expresa a través de la ecuación (Felgaer, 2005).

$$P(h|q) = \frac{P(q|h) * P(h)}{P(q)} \quad (2)$$

Donde:

$P(h)$; denota la probabilidad inicial que tiene la hipótesis h

$P(q)$; denota la probabilidad previa de que se observe el dato de entrenamiento q , es decir la probabilidad que existe de que q no tenga conocimiento acerca de cuál es la hipótesis.

$P(q|h)$; denota la probabilidad de observar datos q dados algún mundo en el cual la hipótesis h se sostiene.

$P(h|q)$; es la probabilidad posterior y refleja la influencia de los datos de entrenamiento q , en contraste con la probabilidad previa $P(h)$, que es independiente de q .

En la ecuación tenemos la ecuación del teorema de Bayes, la cual es la piedra angular para los métodos de aprendizaje bayesianos, ya que la expresión permite calcular la probabilidad posterior de $P(h|q)$, a partir de la probabilidad previa $P(h)$, junto con $P(q)$ y $P(q|h)$ (Mitchell, 2011)

La teoría de probabilidades nos permite la toma de decisiones bajo incertidumbre. Proporcionando una forma de calcular la probabilidad de una hipótesis basada en su probabilidad previa, así, en la clasificación el teorema de Bayes se utiliza para calcular las probabilidades de las clases permitiendo tomar decisiones racionales entre múltiples acciones minimizando el riesgo como se observa en la figura. (Mitchell, 2011, pág. 156).

El presente método de proceso representa grandes problemas debido a su alta complejidad matemática, ya que en algunos casos se presentan dificultades

de convergencia en el sistema, otra dificultad presente en el método de Naive Bayes es la generación en el dividendo de valores cero, lo que podría generar valores desvirtuados. En la Figura 12 se observa que existen dos clases de elementos en ese caso rojo y azul, el método de Bayes ingenuo genera dos funciones de distribución normal con diferentes medias (μ_1 y μ_2), para poder agrupar de mejor manera, para ello se dividió en 2 regiones este caso (R_1 y R_2), pero es necesario poseer un límite en la región donde exista la intersección de las dos funciones de probabilidad, en el ejemplo se puede ver que la línea verde cumple dicha función.

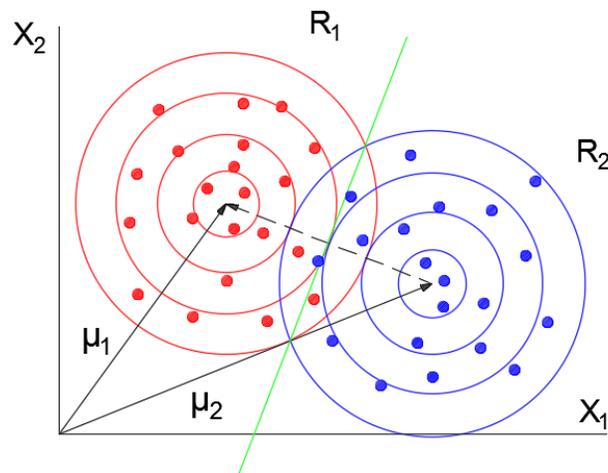


Figura 12 Ejemplo de clasificación de Naive Bayes

Fuente: (Allende, 2015)

2.3.2.1.2. Vecino más Cercano

Vecino más cercano es un estimador aún más simple que Naive Bayes ya que utiliza un bajo recurso computacional y es fácil de implementar. El método permite asignar una etiqueta de su vecino más cercano a una observación, por lo que, lo único que necesita este método es una medida de distancia entre pares de observaciones, teniendo en cuenta que no se requiere que exista distancias simétricas, lo que contribuye con su flexibilidad.

El método se basa principalmente en las distancias que se generan cuando un elemento va a ser clasificado en un grupo representado por una región, donde

a estas áreas se las ha denominado como células Voronoi-tessela las cuales son divisiones del espacio euclidiano por medio del teorema de distancia euclidiana (School of management, 2003) ver en figura 13. En Figura 14 se observa la generación de áreas irregulares para clasificar de mejor manera los datos requeridos.

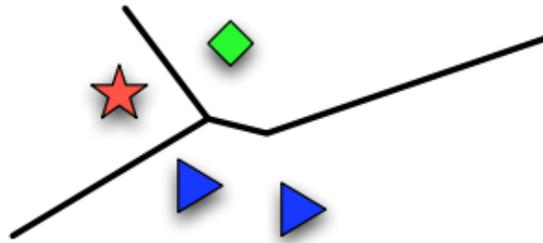


Figura 13 Clasificación vecino más cercano

Fuente: (Smola, Introduction to machine learning, 2010, pág. 25)

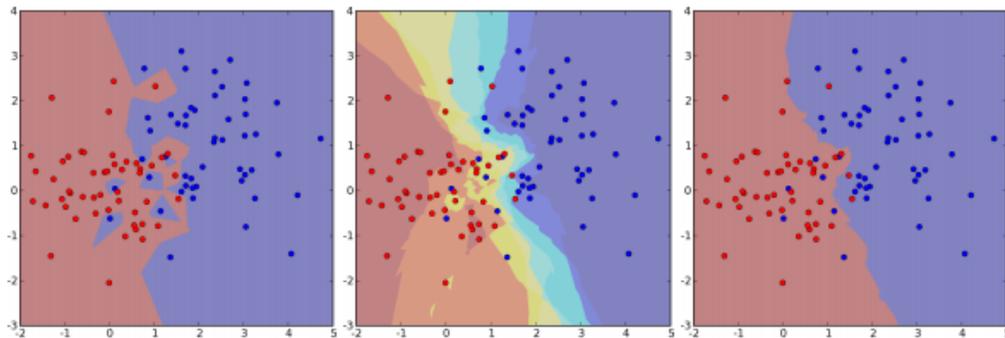


Figura 14 Ejemplo de clasificación vecino más cercano

Fuente: (Sutton, 2012)

A pesar de su simplicidad, los vecinos más cercanos han tenido éxito en un gran número de problemas de clasificación, incluyendo los dígitos escritos a mano o escenas de imágenes por satélite debido a que toman una decisión al límite entre diferentes clases. En el sistema desarrollado se utiliza KNN que es un sub-método de vecino más cercanos y tiene como prioridad asignar un valor de k puntos de datos permitiendo reducir el efecto “ruido” mediante la aplicación de una máxima probabilidad cuando los datos se encuentre entre los límites de

dos regiones y no se tenga la seguridad de donde debe ir. Gracias KNN se puede modelar regiones no paramétricas para determinar de mejor manera una región de interés (Columbia Electrical Engineering, 2015).

2.3.2.1.3. Árbol de decisiones

Es un método gráfico que compara varias alternativas para ir generando decisiones en todos los diferentes casos, lo que genera ramas de ramas de árbol por cada decisión que se vaya a tomar. Un árbol de decisiones se compone de nodos internos, ramas y hojas. Este método es un algoritmo básico y sencillo en su programación (Smola & Vishwanathan, Introduction to Machine Learning, 2008).

Ventajas del método:

- Disminuye la complejidad de un problema a través de toma de decisiones.
- Reduce considerablemente variables independientes.
- Brinda seguridad en los resultados obtenidos.
- Son sistemas secuenciales que generan más fácilmente a que grupo pertenecen los resultados (Olivas, 2007).

En la Figura 15 podemos ver el funcionamiento y la estructura de este método, la primera etapa contempla el problema a decidir, la segunda etapa una pregunta para diferenciar dos posibles casos, esto se realiza en el nodo de decisión, además se usa ramas para entender gráficamente los caminos que tomaron las diferentes decisiones, para llegar finalmente a los posibles resultados, en este caso habrá 4 resultados para el problema original.

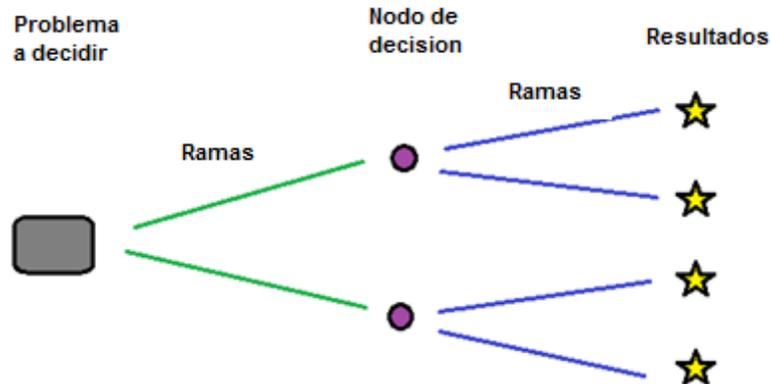


Figura 15 Estructura y funcionamiento de árbol de decisiones.

2.3.2.1.4. Redes Neuronales Artificiales (ANN)

El método de ANN, “trata de simular” las redes de células nerviosas (neuronas) y su sistema nervioso central biológico (humano o animal)” (Gershenson, 2010). El método ANN, plasma conjuntos de neuronas para generar un conocimiento neurofisiológico que permite generar mejoras o aceleración en las operaciones matemáticas de nuestro cerebro. De modo que la neurona artificial emula la biología, así por ejemplo las dendritas representan las entradas, el axioma representa los pesos y la sinapsis son las salidas, los pesos representan el valor de importancia del ingreso en el sistema; en la figura 16 se puede comparar la similitud entre neuronas biológicas y artificiales.

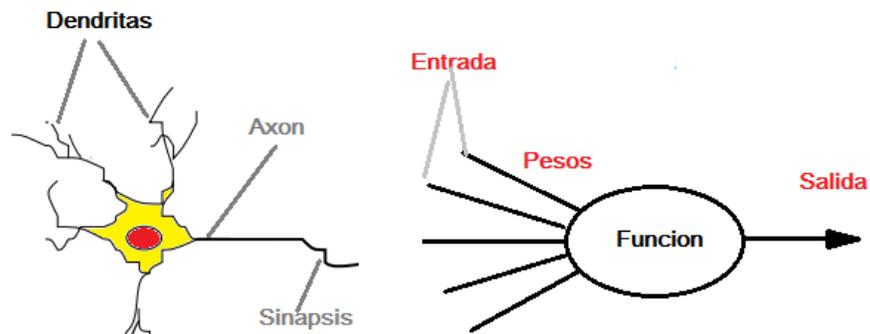


Figura 16 Similitud entre neuronas biológicas y artificiales

Las ANN tienen como base fundamental las siguientes características, estas fueron emitidas por McCulloch y Pitts desarrolladores de esta técnica:

- Las redes neuronales es lo más importante
- El sistema interno de conexión no cambia con los nuevos datos.
- Utiliza conocimientos de computación paralela y colectiva
- El sistema genera seguridad en operaciones matemáticas y resistentes a errores internos
- El sistema ajusta los valores de los pesos antes de ingresar al sistema
- Es muy importante la sincronización entre neuronas en este tipo de red
- El modelo permite utilizar multi-capas para poder resolver problemas más complejos.

Al unir varias neuronas se puede generar redes lo que permite a los sistemas resolver problemas de gran dificultad, por ende se subdivide al sistema en sistemas multi-capas y estos a su vez pueden ser internos permitiendo una clasificación más rápida. En la figura 17 se presenta el esquema de un sistema en red multicapa, primero se ingresa las características o capas de entrada que permiten definir a una clase, la capa oculta permitirá relacionar las diferentes clases hasta encontrar los valores más comunes y finalmente la capa de salida permite clasificar a que clase corresponde cada entrada, para este sistema obtendremos una salida única.

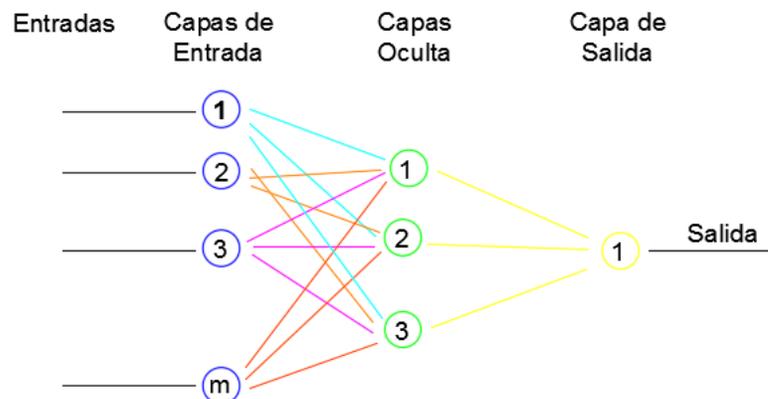


Figura 17 Sistema de red multicapa para un sistema de una sola salida

2.4. Evaluación

La evaluación permite medir el rendimiento que puede generar un sistema planteado, esta fase es primordial en la investigación ya que brinda seguridad en el entrenamiento debido al acierto que pueden generar sus predicciones. Además la evaluación permite definir algunos parámetros como son: la precisión, robustez, complejidad y eficiencia (Ericson, 2006). El método que se desarrolla es K-fold “validación cruzada”.

2.4.1. *K-fold* Validación cruzada

Este método permite la comparación y evaluación de algoritmos de aprendizaje de máquina mediante la división de la base de datos en dos grupos que permitirán evaluar y aprender al sistema. La validación cruzada permite la ejecución sucesiva aleatoria de varias rondas para poder contrastar las validaciones anteriores. Al usar *k-fold*, el índice k nos indicará la cantidad de veces que se repite la validación. Esta idea se originó en la década de 1930 gracias a la ayuda de Mosteller y Turkey, para poder contrastar sus resultados en cada una de sus investigaciones. El funcionamiento de esta validación consiste en generar un grupo de k elementos que se encarga de realizar las pruebas y lo restante se encargara del entrenamiento (Refaeilzadeh, Tang, & Liu, 2008). Esto se repetirá k veces y se aplicará la media aritmética para generar un resultado único. En la Figura 18 podremos entender de mejor manera lo expuesto anteriormente.

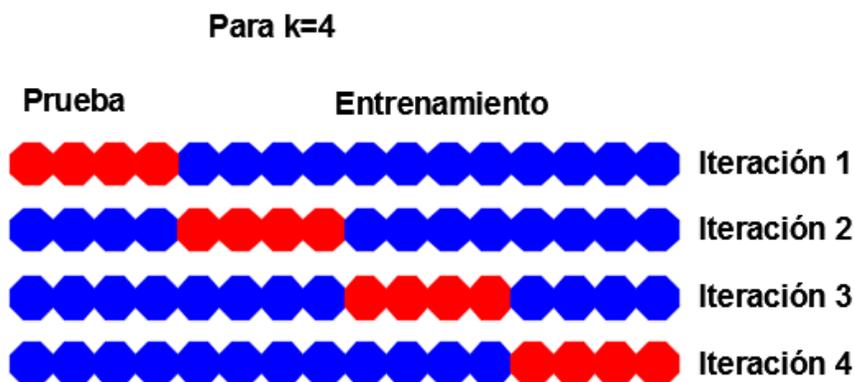


Figura 18 Ejemplo de 4-fold validación cruzada

2.4.2. Matriz de confusión

La matriz de confusión es donde se encuentra la predicción con datos de prueba y también los valores de información correcta para cada uno de los datos haciendo que la comprensión de los resultados sea más sencilla de realizar (Universidad Metodista del Sur, 2008). En la Tabla 5 se observa cómo se estructura esta matriz.

Tabla 5
Partes de una matriz de confusión

		Predicción	
		Positivo	Negativo
Datos Reales	Positivo	Verdadero-Positivo (TP)	Falso-Negativo (FN)
	Negativo	Falso-Positivo (FP)	Verdadero-Negativo (TN)

La matriz planteada permite encontrar otros parámetros que ayudarán a la verificación de los resultados, donde estos deberán estar enfocados al tema de investigación, de ahí que, se ha seleccionado los siguientes valores:

- Sensibilidad (S)

- Precisión (P)
- Tasa de falso descubrimiento (FDR)

2.4.2.1. Sensibilidad

Es un valor estadístico que mide el porcentaje de valores positivos que se encuentran bien clasificados (Universidad Metodista del Sur, 2008). El presente estudio, representa la proporción de casos que están presentes con la enfermedad y respecto a la base de datos referente.

En la ecuación se expresa su notación matemática.

$$S = \frac{TP}{TP + FN} \quad (3)$$

Donde:

TP; representa los verdaderos positivos

FN; denota los falsos negativos

2.4.2.2. Precisión

Es un valor estadístico que identifica el número de valores significativos a partir de un conjunto de datos (Ruiz, Garcia, & Mesa, 2012). Este valor representa la proporción de los casos que poseen esta enfermedad comparado con todos los valores de la base de datos que certifican la presencia de cáncer, En la ecuación 16 podemos ver su notación matemática

$$P = \frac{TP}{TP + FP} \quad (4)$$

Donde:

TP; representa los verdaderos positivos

FN; denota los falsos negativos

2.4.2.3. Tasa de falso descubrimiento (FDR)

Este valor estadístico permite relacionar las hipótesis nulas a partir de los casos positivos en la base de datos (Correa, 2010). Esta proporción permitirá identificar la cantidad de predicciones de cáncer aunque la monografía no la posea. En la ecuación 5 podemos ver su concepto matemático.

$$FDR = \frac{FP}{TP + FP} = 1 - P \quad (5)$$

Donde:

TP ; representa los verdaderos positivos

FN ; denota los falsos negativos

2.4.2.4. Curvas ROC

Son curvas que permiten evaluar los valores de predicción con los datos reales, es decir genera una curva entre la tasa positiva y los falsos positivos, relacionado la sensibilidad y la especificidad. Cuando la curva se encuentre en la región izquierda superior será más precisa será la prueba (Gamez, 2010). En la figura 19 podemos ver un ejemplo de estas curvas y como las curvas especifican la eficiencia de los sistemas

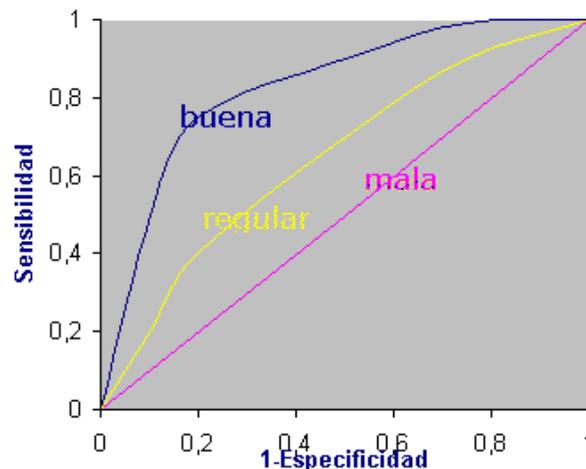


Figura 19 Curvas ROC

Fuente: (Barón López, 2016)

Una curva ROC representa:

- El equilibrio entre sensibilidad y especificidad.
- Mientras más cerca la curva se encuentre en el borde izquierdo superior, más precisa será la prueba.
- Cuanto más cercana esté la curva a la diagonal de 45 grados, menor será la precisión de la prueba.

CAPÍTULO III

3. DISEÑO DEL SISTEMA

3.1. Introducción

En esta sección se explica de mejor manera la implementación y las etapas para la elaboración de un sistema capaz de identificar cáncer de mama en las imágenes mamográficas, iniciando con la sección de segmentación la que permite definir la posible zona de interés de la cual se podrá encontrar las características del conjunto de entrenamiento y prueba para finalmente analizar las mismas con algoritmos de *Machine Learning* para determinar un pronóstico de la existencia o no esta enfermedad, por medio de las métricas de evaluación como exactitud, eficiencia, etc. Todo esto procedimiento se implementa en lenguaje C utilizando el software MatLab®.

3.2. Diagrama de Bloques

A continuación se indica en la Figura 20 las etapas que se implementan para la ejecución del nuestro programa



Figura 20 Diagrama de bloques del sistema

En el diagrama anterior las etapas iniciales tiene como finalidad la extracción de las áreas de interés, para luego obtener de dichas áreas las características, las cuales permitan ayudar en la identificación de cáncer por medio de algoritmos de aprendizaje de máquina. Este método está respaldado en el trabajo de investigación denominado “clasificación de mamografías usando la función de base radial de rango tipo M” del Ing. José Moreno. La cual fue de gran ayuda de gran ayuda en la resolución del problema planteado, ya que explica sencillamente el uso de operaciones morfológicas y como estas se puedan aplicar a las mamografías. (Moreno Escobar, 2007).

3.3. Imágenes Mamográficas

La investigación propuesta tiene como pieza fundamental las mamografías de la base de datos DDSM: Digital Database for Screening Mammography (Universidad del Sur de Florida, 2006), la cual posee una cantidad de 236 imágenes en el estándar JPG, de las cuales 141 poseen cáncer y las 95 restantes no poseen dicha enfermedad. El formato mencionado anteriormente, permite trabajar en una matriz de tres dimensiones la misma que representa en cada una de ellas el color rojo verde y azul, haciendo necesario la transformación de dicha imagen a escala de grises para poder trabajar en las siguientes etapas, en la figura 18 se puede observar una mamografía con el formato jpg.

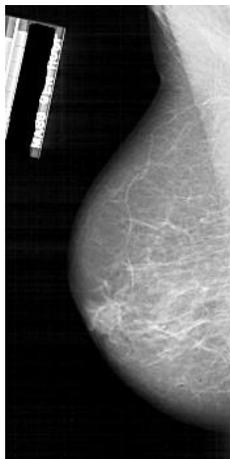


Figura 21 Imagen mamográfica de la base de datos en formato jpg.

Como se menciona anteriormente es necesario la transformación de la imagen RGB a escala de grises, debido a la facilidad que permite la implementación de operaciones morfológicas en la mamografía, esto se debe tomar muy en cuenta en el diseño, ya que así se detecta más fácilmente las anomalías que se presentan en la imagen. Las dimensiones que brinda la base de datos se encuentran entre (185-257 pixeles) x (330-370 pixeles), con un formato de 8 bits de resolución.

Con las imágenes ya obtenidas en escala de grises utilizando la función *rgb2gray*, la cual tiene como base la exclusión del tono y la saturación solo enfocando a la luminancia para generar dichos valores (MathWorks, 2016), en la ecuación 14 se representa la relación matemática de los tres colores primarios para la obtención de la imagen en escala de gris (SIVP Toolbox, 2006).

$$\mathbf{Imagen} = 0.299 * R + 0.587 * G + 0.144B \quad (6)$$

3.3.1. Eliminación de etiquetas en la imagen

Esta sección permite eliminar zonas de no interés alrededor de la mama, enfocándose principalmente en las etiquetas de los médicos utilizan para identificar la mamografía, para ello se implementó la función *im2bw* que permite binarizar la imagen la cual implementa 1 para todas las áreas diferentes del fondo, esta etapa implementada se ve en la figura 22 (MathWorks, 2016).



Figura 22 Imagen mamográfica binarizada.

Para la eliminación de las etiquetas se utiliza la función *regionprop* la misma que identifica todas las áreas que existen en la imagen, y almacena dicha información en su variable de salida, en la figura 23 se puede observar por medio de rectángulos todas las áreas que el sistema encuentra (MathWorks, 2016).

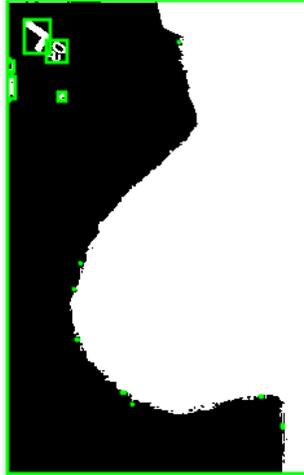


Figura 23 Imagen con áreas detectadas con medio de la función *regionprop*.

Como sección final se tomó la región de mayor medida que la función emite, en este caso solo selecciona la mama y elimina las etiquetas que se encuentra alrededor de la misma. Para ello se utilizó la función *ismember* que permite identificar los elementos que se desean obtener dentro de la imagen colocando valores de 1 en la nueva matriz. Esto se reemplaza en la imagen original para solo obtener la región de la mama, esto se ve en la imagen 24.

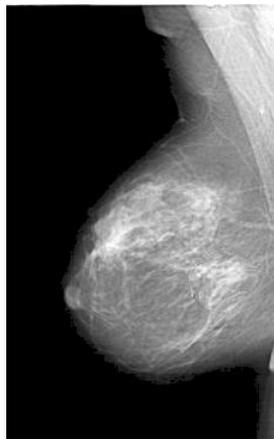


Figura 24 Imagen resultante de esta etapa.

3.3.2. Inversión de intensidad de la imagen

En esta sub-etapa es necesario, cambiar la saturación de intensidades bajas de la imagen, debido a que la etapa de segmentación requiere que las masas anormales se encuentren en pixeles con una intensidad alta, esto hace que aumente el contraste, esto se logra utilizando la función *imadjust*, los términos internos representan a la imagen, al primer vector indican los valores de contraste de la imagen de entrada y el segundo el contraste de salida que se desea, en este caso solo se necesita la inversión de intensidad de la mamografía, en la Figura 25 se observa este proceso de transformación resultante (MathWorks, 2016).

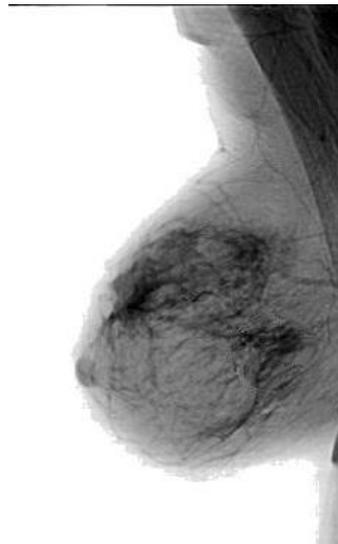


Figura 25 Inversión en la intensidad de la imagen

Con la imagen ya convertida es necesario enfocarse principalmente en el mejoramiento de la calidad de la misma, para poder identificar de mejor manera las zonas donde exista una probable masa anormal en el tejido, centrándose en la eliminación de áreas que no pertenecen a la región de la glándula mamaria, para ello se aplicará técnicas de segmentación.

3.4. Segmentación

El método a usarse en el presente estudio será el de segmentación completa para la obtención directa de las regiones de interés, de tal forma que elimina las áreas que son menos relevantes en el análisis. Al trabajar con mamografías en escala de colores grises, se utiliza principalmente dos propiedades básicas de la segmentación que son: la discontinuidad y la similitud (Muñoz, 2009).

Esta etapa es fundamental en el sistema propuesto, ya que identifica la superficie donde existe el padecimiento, como se mencionó en el capítulo anterior las técnicas de segmentación permiten dividir a la imagen, para obtener las áreas de un posible cáncer en la imagen, que luego puedan generar características propias y permitan clasificar correctamente cada caso, en la Figura 26 se muestra un esquema de las sub-etapas que se implementan.



Figura 26 Diagrama de bloques de la segmentación

Para imágenes médicas en este caso mamografías, los comportamientos de las capas de ruidos no son uniformes, es decir no existe un valor de umbral fijo para todos los casos; para ello se recomienda utilizar algoritmos de multicapa haciendo que aumente la complejidad en la resolución, para esto se recomienda métodos específicos, como puede ser el de Otsu.

3.4.1. Valor umbral por el método de Otsu

Para imágenes médicas en este caso mamografías, los comportamientos de las capas de ruidos no son uniformes, es decir no existe un valor de umbral fijo

para todos los casos; para ello se recomienda utilizar algoritmos de multicapa haciendo que aumente la complejidad en la resolución, para esto se recomienda métodos específicos, como puede ser el de Otsu.

Como se mencionó en la sección anterior, al momento de invertir la intensidad de la imagen nos permite fácilmente obtener un umbral acorde a las zonas de interés debido a que esta técnica es muy dependiente de la luminosidad de la imagen, en el programa se aplicó la función *multithresh* la que utiliza este método generando un umbral único para cada imagen (MathWorks, 2016), y además brinda la posibilidad de segmentar a la imagen en dos niveles de cuantización, pero si le acompaña con la función *imquantize* la cual aumenta más niveles permite encontrar de mejor manera las secciones que deseamos eliminar (MathWorks, 2016).

En la figura 27 se observa visualmente los diferentes niveles que van generando las capas en la imagen permitiendo una mejor diferencia de las zonas de no interés, con las cuales pueda confundir el sistema, generando así soluciones erradas (MathWorks, 2016).

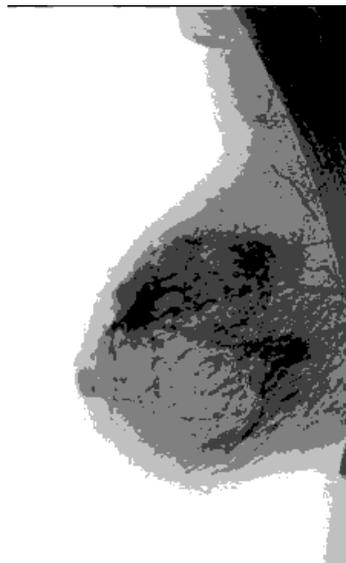


Figura 27 Imagen aplicada la técnica de segmentación de Otsu

Para poder eliminar las secciones diferentes al color blanco que corresponden a la glándula mamaria, se recomienda rellenar el área del seno para que no se produzca una eliminación de partes que contengan información acerca de la patología buscada, para esto se usó la función *infill* aquella que “realiza una operación de inundación de relleno en los píxeles de fondo de la imagen binaria de entrada BW, a partir de los puntos especificados en las locaciones” (MathWorks, 2016), en la Figura 28 se observa el relleno de información en las áreas interés de las cuales se va a trabajar.

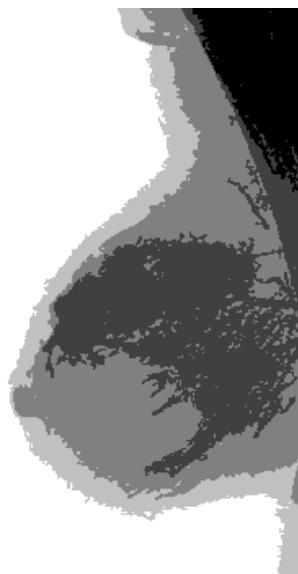


Figura 28 Rellenar áreas de imagen

Como último paso se realiza la eliminación de las áreas no deseadas, para esto se utilizó la función *regionprops(bw, {'Solidity'})* la misma que es “un escalar que especifica la cantidad de los píxeles en la envolvente convexa que también están en la región” (MathWorks, 2016), gracias a esto se genera una región en blanco, donde se encuentran elementos que forman ruido la cual es necesario eliminar, un ejemplo de esta función se lo observa en la figura 29. Es necesario utilizar un índice de grosor para la detección de dichas zonas de no interés, para ello se utiliza la función *ismember* la misma que compara los elementos de la

imagen del perímetro con la línea de índice del grosor para formar la zona real si eliminar información del seno por el grosor (MathWorks, 2016)



Figura 29 Imagen utilizando la función *regionprops*

La función *find* sirve para utilizar las regiones de negro como base para la imagen original y así obtener una nueva mamografía solo con la zona de interés, pero como eso genera una matriz sencilla de formato simple es necesario cumplir el formato de imagen para utilizarlas en las funciones de los siguientes etapas y por eso se usó la función *uint8* generando así el formato deseado (MathWorks, 2016), en la Figura 30 podremos ver la imagen de ingreso para la etapa de segmentación utilizando las operaciones morfológicas.

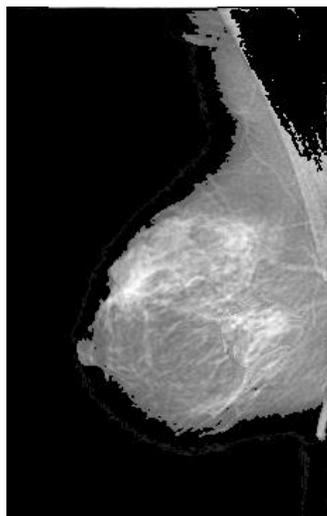


Figura 30 Imagen de entrada para operaciones morfológicas

3.4.2. Operaciones morfológicas

Esta etapa del sistema es la más importante ya que aquí se podrá encontrar posiblemente la existencia de melanomas en la imagen mamográfica, como sabemos las operaciones eliminarán áreas de no interés, generando así características de las mismas. Para nuestro sistema se implementó las operaciones de dilatación y erosión.

Antes de aplicar las operaciones es necesario utilizar la función *strel* (MathWorks, 2016), que permite crear elementos morfológicos de 2 dimensiones en las imágenes para generar un elemento estructurante en forma de: diamante, disco, línea, cuadrado, etc. En nuestro caso se implementó el tipo diamante debido a que solo acepta esta estructura nuestro sistema y con un radio de 1 pixel.

Ya con la forma establecida se procede a realizar la primera parte que es la dilatación, esto se logró gracias a la función *imdilate* (MathWorks, 2016) que permite generar múltiples dilataciones en la imagen a partir del elemento diamante seleccionado anteriormente, como vemos en la Figura 31 la imagen aumenta su contraste en los pixeles donde existe un posible melanoma mientras las otras regiones se difuminan disminuyendo así el contraste en estos pixeles.

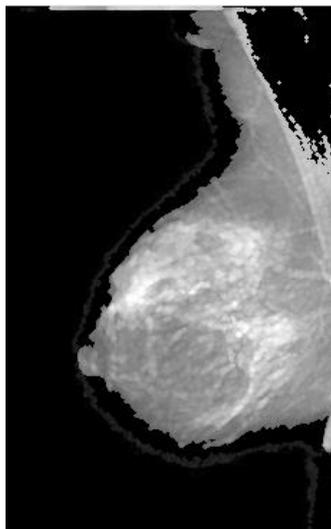


Figura 31: Implementación de dilatación

Es necesario eliminar las regiones de bajo contraste, para ello se utiliza la función *imclearborder* (MathWorks, 2016) la misma que permite cambiar la cromancia de los pixeles que se encuentran fuera del posible tumor detectado en las imágenes. Además se utilizó el máximo nivel de conectividad que es 8 el cual permite descartar todos los valores que estén fuera del diamante, ya que si se escogiera otro valor hubiera generado ruido innecesario en el sistema propuesto. En la Figura 32 podemos ver los resultados de esta función y como se genera la zona de interés.

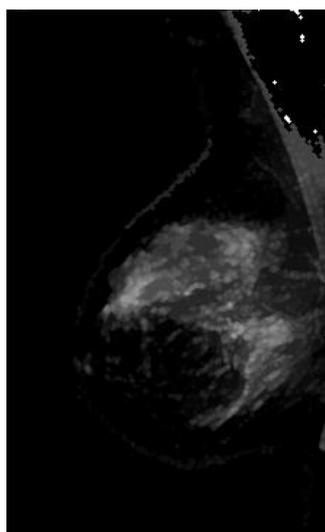


Figura 32 Eliminación de área de no interés

Para poder eliminar las áreas que no son parte de un posible melanoma se aplicará la erosión, por medio de la función *imerode* (MathWorks, 2016), la cual permite la selección de elementos en forma estructurada de diamante para así poder suprimir los valores que se encuentran fuera de los posibles zonas de interés. En la Figura 33 se podrá ver la aplicación de esta operación morfológica.



Figura 33 Aplicación de erosión

Por último aplicamos la segmentación por umbralización para definir más claramente la región donde existe cáncer, esto se logra comparando el umbral que posee cada pixel de la imagen con un valor preestablecido de umbral, para poder aceptar o desechar este valor. En la Figura 34 podremos ver el resultado final de todos los procesos mencionados anteriormente y estas imágenes serán las que nos brindan las características deseadas.



Figura 34 Resultado final de la segmentación

3.5. Características

Esta sub etapa es una de las más importantes ya que permitirá definir los atributos más importantes que cada mamografía posee. Las características permitirán diferenciar matemáticamente cada clase a la que pertenece cada caso en la base datos. Las características son los datos de ingreso en los algoritmos de aprendizaje de máquina, las cuales permiten construir el clasificador encontrando patrones en cada uno de los posibles casos. En nuestro caso utilizamos 3 características. A todas las características planteadas se recomienda normalizar los datos para tener resultados óptimos.

La primera característica que se implementó fue el número total de pixeles diferentes al color blanco que existen en la imagen, es decir es la sumatoria de pixeles que poseen diferente intensidad de color blanco en la imagen. En las imágenes que se cree un posible cáncer el número de pixeles es bajo mientras las que no poseen aumenta el número de pixeles ya que la etapa de segmentación genera toda la imagen del seno haciendo que se acumule mucho más pixeles en cada caso. En la Figura 35 podemos ver los valores que se generan en la primera característica.

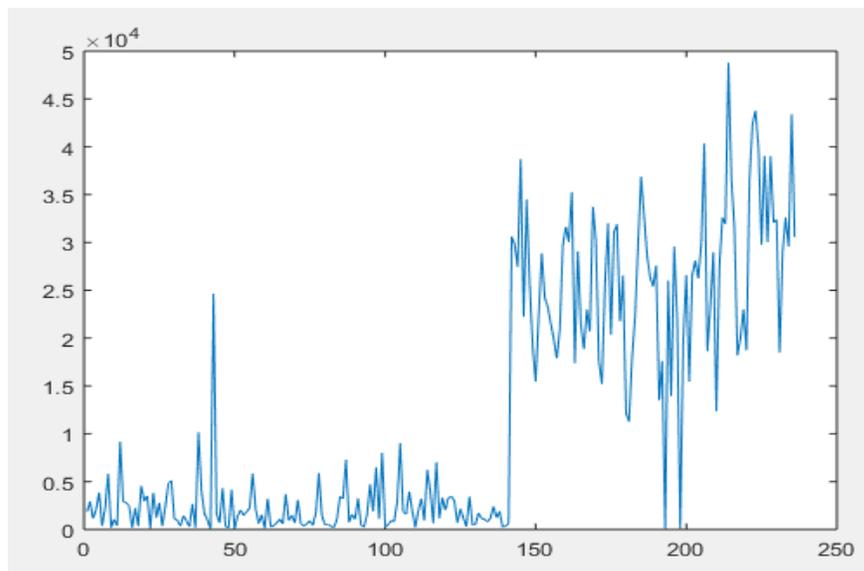


Figura 35 Primera característica por sumatoria de pixeles diferentes de blanco

La segunda característica consiste en realizar el promedio de los valores mínimos que se obtuvo de la sección anterior es decir, mientras más pequeño sea la cantidad de pixeles en la imagen su valor promedio de intensidad será mayor debido a que no cambiará entre pixeles y resultado obtenido será casi fijo, mientras en la otra imagen al tener una gama diferente de gris en cada una de las regiones generara un valor menor en su promedio debido a sus diferentes valores en toda la imagen. En la figura 36 podremos ver el promedio de intensidad de gris para cada caso de la base de datos.

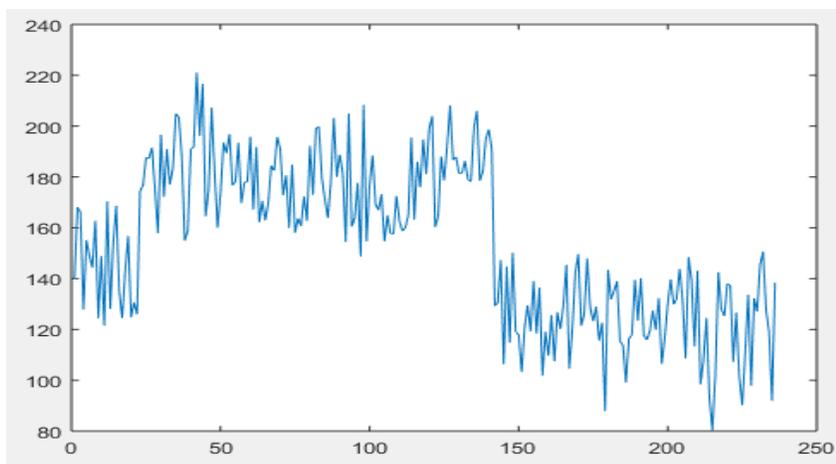


Figura 36 Segunda característica por promedio de pixeles

Como vemos en la Figura 37 se aplicó medidas de dispersión como la varianza, estas características se obtuvieron a partir de los píxeles diferentes al color blanco para poder encontrar las relaciones numéricas que se originan entre el número de píxeles y su intensidad de gris.

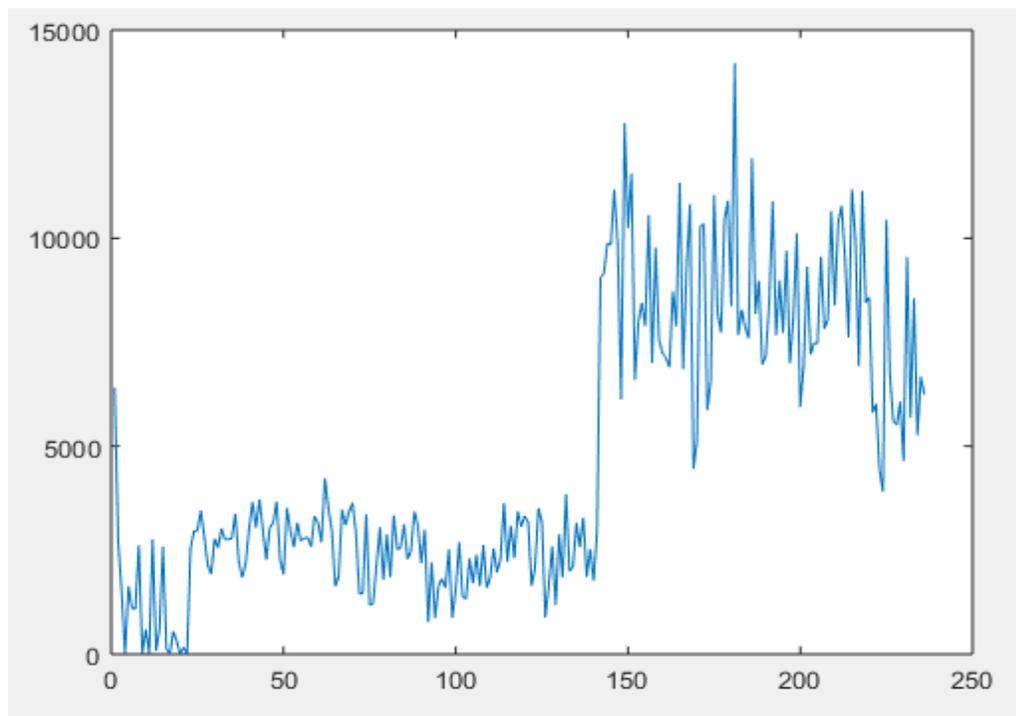


Figura 37 Tercera característica se aplicó la varianza

3.6. Machine Learning

Con los grupos ya obtenidos se aplica las funciones de clasificación de redes neuronales artificiales, como primera función se requiere generar una red neuronal en este caso se implementó *fitnet* la cual tiene una capa oculta de neuronas para una mejor clasificación (MathWorks, 2016). Luego se aplicó la función *train* que permite entrenar las características de la base de datos utilizando la red neuronal mencionada anteriormente.

Después de ingresar los valores requeridos se genera como salida una red de clasificación que permitirá a las nuevas imágenes de prueba determinar la existencia de cáncer o no, para ello se recomienda realizar una función *if* ya que

los resultados obtenidos van de cero a uno, por ende es necesario generar una condición que permita definir a que clase pertenece utilizando una condición en donde los valores mayores a 0.5 determinara la presencia de cáncer.

Como sección final se aplicara la comparación de los resultados obtenidos y los originales para poder determinar la eficiencia del sistema y gracias a ello comparar con las diferentes técnicas de *Machine Learning* como KNN, Naive bayes, o árbol de decisiones, de las mencionadas anteriormente se aplicara el mismo método que se implementó para ANN, generando un grupo de prueba y entrenamiento que ingresara al núcleo de cada uno de los sistemas propuestos.

Como se aplicó k-fold es necesario utilizar el valor promedio de todas las iteraciones ya que se podrán sustentar los valores en todos los posibles casos que puedan ir variando a lo largo de cada caso. En la Figura 38 se puede ver el sistema de comparación que se realiza para encontrar la exactitud del sistema y generar así los valores que permitirán sustentar los métodos que se implementaron para la identificación del cáncer en las imágenes, para profundizar de mejor manera se observa que en el primer if permite comparar los caos emitidos por el sistema con respecto a los datos reales de cada imagen, cuando este detecta un falso acumula un error, para luego repetir el procedimiento dependiendo de los datos de prueba obtenidos en la validación cruzada. Con el error ya obtenido se encuentra el error porcentual el cual dividimos el número de errores con respecto el número de casos de prueba.

```

for z=1:length(respuesta)
    if solucion(z,1)==datprueby(z,1)
        error=error+0;
    else
        error=error+1;
    end
end
u=i;
e=100*error/num1;
errortotal(u,1)=e;
eficiencia(u,1)=100-e;

```

Figura 38 Obtención de resultados

3.7. Validación

Como se mencionó en el capítulo anterior se explicó el funcionamiento de k-fold validación cruzada, lo cual nos permite entender la función *crossvalind* (MathWorks, 2016) en donde el primer parámetro se asigna el tipo de validación en nuestro caso k-fold, como segundo valor se ingresara la matriz con los valores donde se define a cada imagen la presencia de cáncer o no, y por último se dará el valor de k iteraciones que requiera nuestro sistema.

Esta función nos entregara una matriz de igual dimensión que N pero con la diferencia de que cada posición tendrá asignados un valor aleatorio de 1 a k, lo cual permitirá generar grupos de prueba y entrenamiento, como vemos en la Figura 36 se aplicó un ciclo *for* que permitirá encontrar los índices en cada iteración para dividir en los dos grupos mencionados anteriormente, las variables *pruebak* y *entrenak* emitirán la posición aleatoria de datos tanto de prueba como de entrenamiento, para acumular solo los valores en la condición k que se encuentre el ciclo *for*, y los vectores *prueba* y *entrena* permiten definir los casos que se usaran para las pruebas y los restante para entrenamiento.

```

for i = 1:k
    pruebak = (indices == i);
    prueba=find(pruebak==1);
    entrenak = ~pruebak;
    entrena=find(pruebak==0);

```

Figura 39 Sección de código que genera grupo de prueba y entrenamiento ANN

CAPÍTULO IV

4. ANÁLISIS DE PARÁMETROS DEL SISTEMA Y RESULTADOS OBTENIDOS

4.1. Introducción

En este capítulo evaluaremos los diferentes algoritmos de aprendizaje de máquina propuestos, por lo cual se lo pone a prueba por medio de las variables dependientes que el programa posee. Para así obtener más claramente el mejor identificador de cáncer de mama respaldado con su respectivo análisis de eficiencia, especificidad, sensibilidad, curvas ROC y matriz de confusión.

4.2. Obtención de parámetros

Antes del análisis de parámetros se debe encontrar cual valor de k para la validación se utilizara para ello se comparara el mejor valor de k con sus respectiva exactitud para determinar cómo variable independiente dicha condición. En nuestro caso utilizaremos el valor de 10 el mismo que se puede verificar en la Figura 35

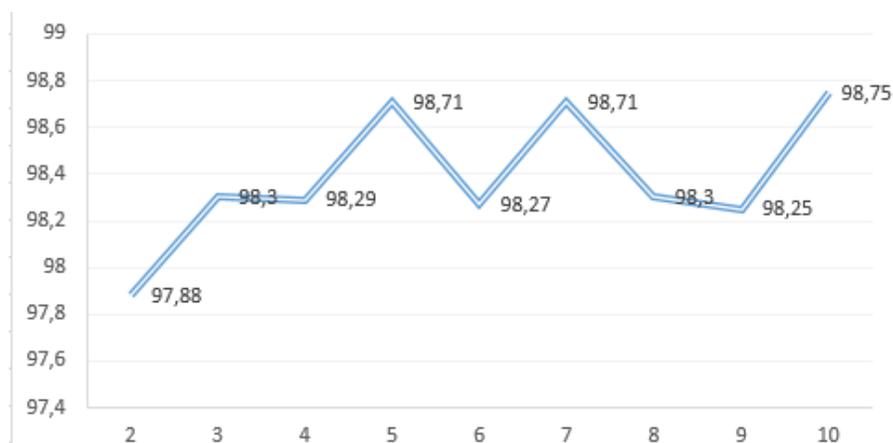


Figura 40 Valor de exactitud vs k-iteraciones

Para la obtención de los parámetros de segmentación mencionados anteriormente es importante definir que los mismos son dependientes entre sí en

el control de la exactitud de los sistemas, por ende es necesario ejecutar todo los rangos de los parámetros en las diferentes técnicas de aprendizaje de máquina para encontrar los máximos valores de exactitud para definir los mismos parámetros como absolutos al utilizarse en nuestro sistema.

Como se mencionó en la etapa de segmentación se requiero tres parámetros para identificar las áreas de un posible cáncer. La primera variable representa los niveles de cuantización que permite a la imagen dividir por diferentes capas, haciendo que se encuentre la región de musculo y hueso que se desea eliminar a esta variable la llamaremos Niveles que tienen valores de 1 a 10. La segunda variable representa el grado en porcentaje para la generación de un perímetro en la glándula mamaria, esta se lo representa con la variable Solidy y tiene un rango de 0.1 a 0.9.

La última variable representa el valor umbral de erosión que se utiliza para disminuir el ruido en el objeto detectado, la variable se representa con Umbral que va desde 0.1 a 0.9. Tenemos que tener presente que estas tres variables son relacionadas en si haciendo que si una cambie se modifique la eficiencia del sistema, por eso se propuso obtener los valores de eficiencia en función de las tres variables. Para llegar así a determinar los valores definitivos que se utilizara en nuestro sistema.

Se utiliza ANN como primer sistema de evaluación debido a su fácil implementación y seguridad que estos brindan. Desde las Tablas 6 a 23 se realizó el análisis de exactitud y eficiencia para encontrar los mejores parámetros que el sistema se brinda.

Tabla 6
Exactitud vs (solidity 0.1, umbral y nivel) para ANN con red Fitnet

Con solidity 0.1										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	98,768	99,524	98,75	98,885	99,6	98,828	98,23	98,032	98,102
	2	97,517	98,599	98,89	98,496	97,854	98,648	98,75	98,431	98,348
	3	99,009	98,532	99,16	99,333	98,706	99,31	99,57	98,047	97,996
	4	99,145	98,013	98,53	99,63	98,846	98,834	98,71	97,565	98,73
	5	99,412	97,83	99,09	98,88	98,929	98,296	99,63	98,615	98,899
	6	99,667	98,791	98,8	98,524	99,355	98,815	98,59	98,088	98,292
	7	98,704	98,889	98,66	99,111	98,591	99.8	98,26	97,977	98,705
	8	98,661	99,5	98,2	98,667	99,6	99,083	98,13	97,882	97,487
	9	99,091	99,259	98,25	99,444	99.8	99.8	99,58	98,402	97,727
	10	99,643	100	97,54	99,545	99,117	99,565	99.8	98,087	99,167

Tabla 7
Eficiencia vs (solidity 0.1, umbral y nivel) para ANN con red Fitnet

Con solidity 0.1										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	8,812	9,47	7,136	6,525	6,58	10,369	6,336	6,074	7,903
	2	11,51	18,53	9,624	9,622	9,15	16,496	9,276	8,699	11,36
	3	17,69	25,78	12,03	11,95	11,4	19,996	12,04	15,89	13,9
	4	22,33	28,94	14,71	14,44	13,77	22,467	20,01	19,7	16,31
	5	24,92	31,17	17,12	16,65	16,19	24,957	25,25	22,61	19,19
	6	31,7	33,62	19,54	19	18,64	32,13	27,78	25,56	21,75
	7	37,49	38,79	21,85	21,18	21,36	42,493	30,77	28,67	24,47
	8	41,29	41,38	24,18	23,56	27,26	48,752	34,05	35,96	27,11
	9	45,02	43,78	26,86	25,88	30	51,523	37,05	38,83	32,68
	10	50,71	48,26	29,46	28,32	32,25	58,339	39,61	41,41	37,75

Tabla 8
Exactitud vs (solidity 0.2, umbral y nivel) para ANN con red Fitnet

Con solidity 0.2										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,524	98,547	100	98,944	99,615	99,14	98,026	98,824	98,065
	2	99,23	99,259	99,545	98,641	99,697	99,524	98,078	96,225	97,336
	3	99,272	99,13	98,81	98,394	99,13	98,768	99.8	98,434	98,305
	4	99.8	98,421	99,643	98,85	98,99	98,171	98,929	98,297	96,822
	5	98,441	97,589	99,63	99,545	99.8	99,5	98,822	98,513	98,264
	6	97,318	98,101	99,255	98,99	99,565	99,583	98,239	97,866	99,124
	7	98,602	99.8	99,615	98,462	99.8	99,643	98,963	98,264	98,314
	8	99,545	99,315	97,405	98,2	99.8	99,583	98,416	98,81	97,692
	9	98,439	99.8	99,139	99,13	98,755	99,524	98,49	97,615	98,73
	10	99,545	98,333	99,201	99,089	98,564	98,655	99,583	96,421	96,912

Tabla 9
Eficiencia vs (solidity 0.2, umbral y nivel) para ANN con red Fitnet

Con solidity 0.2										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	9,5292	6,6431	5,306	7,639	5,695	5,66	5,791	5,495	5,7974
	2	12,553	9,1164	7,842	13,54	8,377	8,174	8,289	8,293	8,5322
	3	14,8	11,503	10,13	16,62	10,63	10,54	11,07	11,04	10,93
	4	17,292	14,031	12,64	18,9	13,4	12,95	13,51	13,94	13,455
	5	21,667	16,231	15	21,5	16,83	15,72	15,95	16,58	16,101
	6	25,555	18,476	17,25	23,73	19,15	18,46	18,37	19,34	18,482
	7	27,801	23,449	19,36	25,92	24,22	21,07	20,79	21,85	21,084
	8	30,458	27,358	21,68	28,34	27,84	23,55	23,28	24,4	23,667
	9	32,827	29,978	24,18	34,35	30,17	26,14	26,06	27	26,208
	10	35,34	32,493	26,52	36,98	32,83	28,63	28,68	29,4	29,239

Tabla 10
Exactitud vs (solidity 0.3, umbral y nivel) para ANN con red Fitnet

Con solidity 0.3										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,149	98,63	99,487	99,6	97,945	97,794	98,83	98,4	98,055
	2	98,178	97,963	99,2	99,048	98,333	99,286	98,49	98,677	98,75
	3	97,593	97,525	99,8	98,908	99,115	99,13	99,13	99,545	96,775
	4	98,443	99,63	99,706	98,634	99,245	99,722	98,94	99,165	96,74
	5	99,474	99,524	99,545	99,145	99,583	99,565	97,98	99,8	97,029
	6	98,187	99,63	99,091	98,276	98,071	98,261	98,31	98,5	99,524
	7	99,31	99,8	98,75	98,353	98,822	99,215	99	98,704	97,922
	8	99,615	98,656	99	99,259	98,75	98,676	99,62	99,155	97,761
	9	98,603	98,286	99,259	98,717	99,524	99,355	99,58	99,181	97,738
	10	99,677	99,23	98,92	98,808	99,5	98,855	98,25	96,555	96,473

Tabla 11
Eficiencia vs (solidity 0.3, umbral y nivel) para ANN con red Fitnet

Con solidity 0.3										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	6,7825	5,226	5,489	5,102	5,931	7,226	5,969	7,0237	6,8629
	2	9,0408	7,644	7,958	8,0359	8,696	9,848	8,426	11,771	9,6687
	3	11,229	9,899	10,71	10,666	11,46	12,17	11,15	14,37	12,211
	4	13,632	12,21	15,5	13,475	13,78	14,7	13,76	17,129	14,646
	5	16,221	14,6	18,03	15,864	16,18	17,19	16,39	19,844	16,941
	6	18,539	17,36	20,39	18,033	18,53	19,6	19,04	22,79	19,774
	7	20,834	19,71	22,76	20,581	20,78	22,26	21,59	26,081	22,035
	8	23,354	22,01	25,15	22,936	23,55	24,52	24,1	28,55	24,627
	9	25,743	24,35	27,6	25,193	26,13	27,08	26,6	31,032	29,592
	10	27,822	26,61	29,88	27,449	28,95	29,5	29,21	33,62	35,363

Tabla 12
Exactitud vs (solidity 0.4, umbral y nivel) para ANN con red Fitnet

Con solidity 0.4										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	98,194	99,655	98,194	98,649	99,375	98,519	99,6	99,286	98,183
	2	99,048	99,129	99,048	98,677	99,165	99.8	99,115	97,06	97,386
	3	99,167	99,643	99,167	98,377	98,807	99,615	99,298	99,208	97,933
	4	99,355	98,235	99,355	99,065	99,091	99,655	98,843	99,545	98,117
	5	98,419	99,215	98,419	98,307	99,111	99,667	97,482	98,533	98,013
	6	98,81	98,73	98,81	99,145	97,513	99,165	99,355	97,168	98,397
	7	98,291	99,286	98,291	98,361	98,786	99,655	99,474	98,322	98,348
	8	98,583	99,115	98,583	99,2	98,486	97,977	98,776	97,638	99,149
	9	99,041	98,245	99,041	99.8	98,908	98,952	98,339	99,091	98,38
	10	97,546	99,231	97,546	99.8	99,355	98,796	99,583	99,6	98,846

Tabla 13
Eficiencia vs (solidity 0.4, umbral y nivel) para ANN con red Fitnet

Con solidity 0.4										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	7,989	8,724	7,989	5,37	6,066	5,445	7,416	6,159	6,142
	2	10,79	12,72	10,79	7,842	8,695	8,202	11,004	9,246	8,791
	3	13,22	15,46	13,22	10,26	11,21	10,72	13,719	11,73	11,69
	4	15,69	18,11	15,69	12,55	13,86	13,17	16,665	14,45	14,33
	5	18,05	24,94	18,05	15,14	16,7	16,32	19,742	17,07	16,76
	6	20,6	29,78	20,6	17,59	19,24	18,85	25,124	19,9	19,76
	7	22,92	32,32	22,92	19,99	21,7	21,75	28,304	22,55	22,67
	8	25,69	34,74	25,69	22,57	24,14	24,16	31,017	25,21	28,61
	9	28,08	37,12	28,08	24,86	26,65	26,82	33,579	28,19	31,25
	10	30,62	39,7	30,62	27,56	29,48	29,64	36,444	31,15	36,84

Tabla 14
Exactitud vs (solidity 0.5, umbral y nivel) para ANN con red Fitnet

Con solidity 0.5										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,6	99,63	98,824	98,929	99,583	99,615	98,292	99,188	98,46
	2	99,197	97,895	99,081	99.8	99,254	98,02	99.8	99,545	96,88
	3	99,167	99.8	99.8	99,215	99.8	99,63	98,63	97,976	98,411
	4	99,032	99,212	99,667	99,048	97,565	98,424	97,37	96,956	97,481
	5	99,5	98,822	99,73	99,5	99,06	98,178	99,243	98,796	96,775
	6	97,403	98,364	97,547	96,291	97,72	98,118	99,063	98,441	98,122
	7	98,846	99,044	98,054	99,259	99,63	98,505	97,495	95,839	98,8
	8	97,335	98	99,474	97,592	98,555	99.8	98,31	97,983	97,794
	9	97,564	97,989	98,095	99,107	98,382	97,592	99	98,547	96,553
	10	99,009	98,246	97,468	98,148	97,973	98,783	99,545	99,208	97,028

Tabla 15
Eficiencia vs (solidity 0.5, umbral y nivel) para ANN con red Fitnet

Con solidity 0.5										
Efi. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	6,6845	7,5482	5,784	6,102	5,8204	6,655	6,342	6,2901	5,497
	2	9,3842	11,275	9,792	8,843	8,405	9,142	9,114	9,1741	8,275
	3	11,885	13,82	12,1	11,6	11,087	11,64	11,41	11,998	10,85
	4	14,135	16,361	14,65	14,11	13,867	15,43	13,98	14,759	13,44
	5	16,591	18,884	16,97	16,7	16,544	19,6	16,66	17,599	16,08
	6	18,999	21,501	19,47	19,08	19,149	22,29	19,37	20,825	18,82
	7	21,422	23,893	21,78	21,59	21,893	24,77	21,78	24,142	22,04
	8	23,848	26,424	24,1	23,87	24,469	27,4	24,41	26,933	24,84
	9	26,31	28,926	26,74	26,44	26,944	29,98	27,26	29,566	27,58
	10	29,177	35,259	29,52	29,54	29,405	32,55	30,03	32,201	30,44

Tabla 16
Exactitud vs (solidity 0.6, umbral y nivel) para ANN con red Fitnet

Con solidity 0.6										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	97,397	98,744	99,167	99.8	99,63	98,808	98,532	9,0271	97,68
	2	99,583	99,189	97,638	98,75	98,824	99,524	99,139	13,397	97,397
	3	99,28	98,439	99,583	99,277	99,195	99,048	98,947	15,933	98,341
	4	97,525	98,488	99,195	98,778	99,091	99.8	97,365	19,179	96,039
	5	98,856	99,333	97,44	97,344	98,63	98,565	98,783	21,687	97,686
	6	98,824	97,131	97,602	98,429	97,773	96,808	98,375	24,76	98,429
	7	98,367	98,512	98,571	98,6	98,533	98,096	99,126	27,322	98,565
	8	96,41	95,031	96,25	96,227	96,71	97,496	96,905	29,68	97,883
	9	97,37	97,133	98,565	98,565	95,627	96,684	97,307	32,382	95,542
	10	97,926	98,586	99,167	97,292	97,715	97,896	98,696	35,214	99,243

Tabla 17
Eficiencia vs (solidity 0.6, umbral y nivel) para ANN con red Fitnet

Con solidity 0.6										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	6,632	5,337	5,906	5,3527	5,5517	7,2487	5,7619	98,49	5,9502
	2	8,86	7,763	8,727	7,6491	8,3261	10,661	8,6503	99,181	8,5494
	3	11,3	10,47	11,09	10,433	10,831	13,477	11,301	97,983	11,085
	4	13,58	12,91	13,39	12,861	13,346	15,91	13,749	99,143	13,78
	5	16,23	15,52	15,93	15,086	15,852	18,371	16,217	98,966	16,383
	6	18,59	17,7	18,09	17,718	18,145	20,965	18,834	98,424	18,983
	7	20,92	20,3	20,7	20,272	20,655	23,552	21,687	98,078	21,85
	8	23,09	22,54	23,05	22,553	23,226	25,87	24,4	96,699	24,439
	9	26,16	24,86	25,81	25,073	25,631	28,788	27,121	96,258	26,84
	10	28,99	27,46	28,58	28,342	28,525	31,804	30,006	99,655	29,426

Tabla 18
Exactitud vs (solidity 0.7, umbral y nivel) para ANN con red Fitnet

Con solidity 0.7										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	98,974	98,721	99,213	98,656	97,846	99,272	99,23	98,097	97,719
	2	99,583	97,889	99.8	98,889	99,615	99,524	98,529	98,915	98,472
	3	97,72	98,776	97,589	98,389	99,524	99,667	99.8	99.8	98,166
	4	95,227	97,104	97,034	95,274	97,53	98,087	96,429	96,374	95,21
	5	97,122	96,033	95,008	98,077	95,468	95,98	97,33	97,5	96,31
	6	97,554	98,846	96,746	98,745	97,219	97,389	98,615	97,258	98,814
	7	96,876	95,619	98,016	96,902	95,79	97,342	99,188	97,944	98,593
	8	95,304	94,473	95,665	96,769	95,875	94,685	95,093	96,691	95,87
	9	96,735	96,261	96,274	97,744	94,128	95,1	97,437	95,609	95,795
	10	97,743	99	98,397	99.8	98,571	98,714	99,188	98,69	98,689

Tabla 19
Eficiencia vs (solidity 0.7, umbral y nivel) para ANN con red Fitnet

Con solidity 0.7										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	14,776	6,164	5,612	7,0613	5,804	8,1997	6,941	7,6846	7,547
	2	18,269	9,688	8,307	10,234	8,436	12,461	9,88	11,306	10,19
	3	20,852	17,52	10,79	12,601	11,19	15,181	17,03	14,105	13
	4	26,515	23,17	13,1	14,906	13,49	17,431	19,3	16,538	15,34
	5	32,649	28,66	15,52	17,331	15,88	19,68	21,73	19,017	17,75
	6	36,132	32,08	21,72	19,975	21,04	22,714	24,32	22,338	20,19
	7	38,677	34,74	24,08	22,385	24,4	25,428	30,71	25,311	22,99
	8	41,114	37,01	26,45	24,932	26,95	27,674	33,22	27,712	25,33
	9	43,704	41,1	28,72	27,419	29,32	29,983	35,97	30,152	27,59
	10	46,497	45,84	31,78	31,617	31,76	32,726	38,48	32,851	30,03

Tabla 20
Exactitud vs (solidity 0.8, umbral y nivel) para ANN con red Fitnet

Con solidity 0.8										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,5	99,444	99,524	99,183	99,1	99,213	97,538	99,63	96,877
	2	99,111	99,565	99,6	99,565	97,867	98,49	98,857	99,13	98,663
	3	98,504	98,796	98,654	99,8	99,63	99,655	99,6	98,346	98,547
	4	95,286	97,303	95,883	97,229	94,63	96,227	95,682	96,691	93,855
	5	97,088	95,85	96,022	96,515	96,017	95,871	94,959	97,037	95,681
	6	96,655	97,805	99,226	98,202	99,165	97,019	99,565	98,704	98,83
	7	98,781	97,262	96,075	95,692	99,333	98,593	98,417	99,655	97,22
	8	97,821	96,076	96,87	96,657	94,991	98,454	96,553	97,037	96,001
	9	96,197	96,477	96,781	96,457	95,515	97,461	95,568	96,786	97,88
	10	98,381	99,5	98,501	98,718	98,274	97,067	98,667	98,8	99

Tabla 21
Eficiencia vs (solidity 0.8, umbral y nivel) para ANN con red Fitnet

Con solidity 0.8										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	6,689	6,202	6,153	5,999	7,3937	7,3948	7,207	6,5715	6,488
	2	9,5	9,026	8,692	8,881	10,501	10,583	10,29	9,906	12,441
	3	15,3	11,33	13,97	11,48	13,021	13,453	12,96	12,666	15,053
	4	18,37	13,91	17,52	13,92	15,399	15,605	15,44	15,077	17,276
	5	20,57	16,1	19,7	16,3	17,979	17,789	17,92	17,583	19,584
	6	22,95	18,59	22,41	19,17	21,121	20,643	20,55	24,768	22,73
	7	25,75	21,23	24,94	21,6	24,001	23,347	23,55	28,406	25,47
	8	28,05	23,4	27,07	23,99	26,309	26,1	25,61	33,689	27,873
	9	30,23	26,01	29,45	26,13	28,583	28,638	28,13	36,468	30
	10	32,82	28,79	31,91	29,29	31,626	31,471	31,04	39,34	32,793

Tabla 22
Exactitud vs (solidity 0.9, umbral y nivel) para ANN con red Fitnet

Con solidity 0.9										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,048	98,712	99,5	98,667	99,167	99,63	99,167	98,783	97,585
	2	99,13	99,474	98,466	97,5	98,846	98,667	98,846	99,074	96,807
	3	98,462	99,5	98,783	98,968	97,703	98,998	97,703	98,696	98,188
	4	97,047	94,582	95,514	97,083	94,953	96,138	94,953	95,272	97,391
	5	93,977	95,753	96,552	94,877	95,74	96,267	95,74	96,946	96,123
	6	98,459	97,275	97,833	97,483	98,137	99,677	98,137	98,796	98,936
	7	99,545	98,63	98,35	98,976	97,242	98,245	97,242	98,875	98,447
	8	95,311	95,344	97,826	95,205	96,891	97,273	96,891	96,214	94,818
	9	96,117	95,518	96,127	95,658	94,352	95,604	94,352	95,272	95,765
	10	99,677	95,533	98,359	97,513	97,992	98,674	97,992	99,145	99,149

Tabla 23
Eficiencia vs (solidity 0.9, umbral y nivel) para ANN con red Fitnet

Con solidity 0.9										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	7,0929	5,734	5,791	5,8447	5,638	6,304	5,638	6,1615	8,8042
	2	9,62	8,488	8,558	8,3582	8,373	8,819	8,373	9,0997	13,212
	3	12,428	11,15	11,18	10,979	11,55	11,45	11,55	11,708	15,565
	4	14,75	13,61	13,6	13,123	13,82	13,72	13,82	13,98	17,835
	5	17,079	15,87	15,76	15,808	16,03	16,14	16,03	16,465	20,5
	6	19,43	18,55	18,42	18,699	19,28	18,97	19,28	19,487	23,664
	7	25,32	21,28	20,74	21,669	22,43	21,8	22,43	22,418	26,644
	8	27,761	23,57	23,08	23,984	24,61	24,25	24,61	24,906	28,899
	9	30,264	26,12	25,35	26,532	26,79	26,71	26,79	27,42	31,306
	10	32,968	28,44	27,85	30,81	29,99	29,26	29,99	30,312	33,892

Debido a que algunos casos comparten la máxima exactitud de 99.8% y el menor tiempo que brinda la eficiencia de 5.35 segundos, se optó por escoger uno de los primeros, debido a que el costo computacional es menor al ir desde valores inferiores hasta superiores, haciendo que los parámetros que se usan para la comparación entre sistemas son los siguientes:

- *Umbral* = 0.4
- *Solidy* = 0.6
- *Nivel* = 1

A continuación se realizó el mismo procedimiento implementado la técnica de vecinos más cercanos para encontrar la máxima eficiencia y exactitud, esto se puede visualizar en las Tablas 24 a la 41.

Tabla 24
Exactitud vs (solidity 0.1, umbral y nivel) para vecino más cercano

Con solidity 0.1										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,375	98,963	99.8	99.8	99,474	99.8	98,933	98,053	96,306
	2	99.8	98,958	98,636	99,565	99,412	99.8	98,548	98,824	98,636
	3	99,317	98,975	99,474	99,089	99,165	99.8	98,843	97,841	97,158
	4	99,474	99,643	99,667	99,13	99,524	98,162	98,35	97,415	96,222
	5	98,179	99,288	99,524	99.8	99,045	99.8	99,583	97,682	97,677
	6	99.8	99,117	99,615	98,69	99,5	99,19	98,81	95,898	97,791
	7	99,087	98,831	98,888	99	99,25	99,195	98,571	97,507	97,791
	8	99.8	99,545	98,696	98,889	99,63	98,373	98,75	97,771	97,652
	9	99,065	98,052	98,889	98,19	99.8	99,6	98,245	98,565	96,966
	10	98,443	98,796	99,215	99,655	99,153	99,565	99,375	96,659	95,932

Tabla 25
Eficiencia vs (solidity 0.1, umbral y nivel) para vecino más cercano

Con solidity 0.1										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,748	0,3645	0,3619	0,3628	0,3638	0,3645	0,3638	0,3632	0,3584
	2	2,195	0,728	0,7287	0,7287	0,7325	0,7313	0,7289	0,729	0,7256
	3	2,5584	1,1038	1,0877	1,0886	1,0891	1,0832	1,0868	1,0839	1,0761
	4	2,9157	1,3956	1,4471	1,4468	1,4456	1,3689	1,3828	1,4389	1,4264
	5	3,2542	1,6819	1,7957	1,7799	1,7612	1,6534	1,6696	1,7565	1,7542
	6	3,5402	1,9159	2,0824	2,1113	2,0462	1,8738	1,9037	2,0393	2,0348
	7	3,8148	2,1379	2,3668	2,4681	2,3003	2,0941	2,1237	2,2973	2,3057
	8	4,0342	2,3559	2,5877	2,8242	2,5203	2,3042	2,3435	2,5153	2,5238
	9	4,2534	2,5474	2,8087	3,162	2,7399	2,4947	2,5356	2,7357	2,7932
	10	4,4615	2,7394	3,0197	3,4474	2,9396	2,6853	2,7273	3,0507	3,1465

Tabla 26
Exactitud vs (solidity 0.2, umbral y nivel) para vecino más cercano

Con solidity 0.2										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	97,824	98,712	99,065	99,8	99,565	99,167	98,459	96,546	98,49
	2	99,13	99,344	99,8	99,1	99,039	98,786	99,5	98,315	97,139
	3	99,565	98,444	99,13	99,286	99,312	99,8	98,309	97,887	96,795
	4	99,333	98,219	99,5	99,333	99,06	98,956	99,63	96,131	96,32
	5	98,656	98,886	99,8	99,8	99,063	99,6	99,129	96,96	97,974
	6	99,8	99,286	99,5	98,655	99,8	98,783	99,8	96,795	95,984
	7	99,167	99,615	99,8	99,444	98,81	99,8	99,183	97,159	97,521
	8	99,333	99,13	99,565	99,565	99,474	98,782	99,474	95,804	96,462
	9	98,831	99,583	99,167	99,677	98,63	99,167	98,947	98,833	96,279
	10	99,524	99,8	98,625	99,5	99,655	99,5	99,524	98,031	96,648

Tabla 27
Eficiencia vs (solidity 0.2, umbral y nivel) para vecino más cercano

Con solidity 0.2										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,951	0,3687	0,3659	0,3643	0,3661	0,3665	0,3628	0,362	0,3805
	2	2,327	0,7393	0,7336	0,732	0,7334	0,7332	0,7302	0,7264	0,7436
	3	2,555	1,0923	1,0952	1,0912	1,0946	1,0922	1,0904	1,0844	1,0988
	4	2,775	1,3819	1,4544	1,4503	1,4551	1,4514	1,4457	1,4389	1,4215
	5	2,989	1,6676	1,7654	1,7783	1,7736	1,7932	1,7749	1,7436	1,705
	6	3,18	1,8904	2,0528	2,0663	2,0614	2,0797	2,1069	2,0278	1,9649
	7	3,372	2,1132	2,3031	2,332	2,3166	2,3599	2,4625	2,2723	2,1843
	8	3,561	2,3251	2,5257	2,5534	2,539	2,5802	2,8181	2,4929	2,4717
	9	3,73	2,5187	2,7467	2,775	2,7616	2,8	3,1536	2,7123	2,8269
	10	3,899	2,7123	2,9423	2,9776	2,9599	3,0085	3,4371	2,906	3,1816

Tabla 28
Exactitud vs (solidity 0.3, umbral y nivel) para vecino más cercano

Con solidity 0.3										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,091	98,706	98,25	99.8	98,724	99,565	97,977	97,417	96,109
	2	99,13	99,019	99,231	99,545	99.8	99,615	98,91	98,103	97,807
	3	98,573	99,226	99.8	99,048	99,63	99,333	99,188	97,863	98,258
	4	99,565	99,545	99,024	98,16	99.8	99,706	98,745	97,667	96,143
	5	99,349	99.8	98,908	99.8	98,519	99,643	99	97,257	97,273
	6	98,977	98,9	98,442	99,045	98,682	98,889	97,348	96,672	95,972
	7	99,583	99,565	97,946	99,13	99,188	99,5	98,783	96,482	95,554
	8	99,412	99,655	99,643	99,63	98,677	99,6	98,613	98,018	98,788
	9	99,615	98,631	99,13	98,782	99.8	98,5	99,565	98,746	96,88
	10	98,591	98,975	99,565	99,615	99,253	98,888	98,054	97,531	98,32

Tabla 29
Eficiencia vs (solidity 0.3, umbral y nivel) para vecino más cercano

Con solidity 0.3										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,8368	0,3644	0,3643	0,3658	0,3615	0,3595	0,3638	0,3612	0,3661
	2	2,1242	0,7314	0,7295	0,7315	0,7271	0,7257	0,7296	0,7266	0,7396
	3	2,3494	1,0902	1,0869	1,0882	1,084	1,083	1,0889	1,0655	1,1034
	4	2,5507	1,4475	1,4448	1,4439	1,4293	1,4399	1,4275	1,35	1,4681
	5	2,7432	1,7831	1,7673	1,7912	1,7157	1,796	1,7135	1,6235	1,7942
	6	2,9345	2,0694	2,1243	2,0755	1,9954	2,0833	1,9901	1,8419	2,0814
	7	3,1142	2,3402	2,4847	2,3594	2,2153	2,3674	2,2091	2,0658	2,3404
	8	3,2828	2,5599	2,8437	2,5803	2,4351	2,5938	2,428	2,2779	2,564
	9	3,4513	2,7797	3,1745	2,7997	2,6462	2,813	2,6347	2,4757	2,7827
	10	3,6199	2,9868	3,4611	3,0109	2,8372	3,0266	2,8243	2,6717	2,9795

Tabla 30
Exactitud vs (solidity 0.4, umbral y nivel) para vecino más cercano

Con solidity 0.4										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,32	99,474	98,889	99,565	99,8	99,8	99,2	97,435	97,466
	2	99,8	98,665	99,565	99,474	98,462	98,944	98,307	98,616	96,88
	3	99,091	99,8	99,545	99,583	99,8	98,731	99,149	99,181	96,621
	4	99,8	98,531	99,444	99,565	98,459	98,759	99,2	96,75	97,112
	5	98,833	98,966	99,048	99,667	99,23	98,5	99,213	97,481	96,873
	6	98,824	99,267	98,759	98,61	98,235	98,875	97,417	95,892	98,093
	7	98,663	99,8	98,489	99,545	99,63	98,8	98,16	95,704	98,413
	8	98,938	99,412	99,545	99,615	98,94	99,065	97,555	97,763	98,649
	9	98,042	96,958	99,6	99,005	99,048	98,99	99,583	98,16	98,244
	10	98,288	97,416	98,182	98,385	97,647	99,165	96,836	99,124	96,768

Tabla 31
Eficiencia vs (solidity 0.4, umbral y nivel) para vecino más cercano

Con solidity 0.4										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,0273	0,3641	0,363	0,3649	0,3623	0,3647	0,364	0,3589	0,3599
	2	2,4054	0,7304	0,733	0,73	0,7272	0,7289	0,7297	0,7253	0,7227
	3	2,6701	1,0886	1,092	1,0873	1,0856	1,0846	1,0877	1,0802	1,0771
	4	2,8897	1,446	1,45	1,4421	1,4438	1,4398	1,4446	1,4349	1,4304
	5	3,1096	1,7833	1,784	1,7651	1,7834	1,7748	1,7746	1,7569	1,7842
	6	3,31	2,0694	2,071	2,0483	2,0697	2,0571	2,0598	2,0395	2,0953
	7	3,5004	2,3466	2,34	2,3123	2,3442	2,3329	2,3256	2,2982	2,3768
	8	3,6907	2,5665	2,561	2,5311	2,5639	2,5503	2,5454	2,5201	2,6314
	9	3,87	2,7859	2,783	2,7497	2,7831	2,7678	2,7643	2,7389	2,8484
	10	4,0383	2,9917	2,989	2,9509	2,9903	2,9739	2,9675	2,941	3,065

Tabla 32
Exactitud vs (solidity 0.5, umbral y nivel) para vecino más cercano

Con solidity 0.5										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,1	98,75	98,947	99,8	99,6	98,573	98,103	97,915	97,582
	2	99,6	97,981	99,333	99,63	99,6	99,13	98,805	97,875	96,168
	3	99,124	99,615	99,8	99,706	98,628	98,929	99,155	98,747	98,384
	4	98,471	99,8	99,333	99,286	99,643	99,6	99,565	98,438	97,426
	5	98,451	98,3	99,8	99,8	99,286	99,655	99,583	98,447	96,492
	6	96,717	97,889	98,613	98,675	99,245	97,137	97,579	97,467	96,154
	7	99,524	99,13	98,248	99,429	98,783	99,8	96,591	97,947	99,231
	8	98,279	98,605	97,018	98,452	98,675	97,382	96,989	97,078	96,695
	9	98,506	97,815	99,31	98,5	98,439	99,667	98,076	99,259	96,738
	10	94,94	97,039	96,79	99,583	95,921	97,643	97,989	99,024	97,31

Tabla 33
Eficiencia vs (solidity 0.5, umbral y nivel) para vecino más cercano

Con solidity 0.5										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,876	0,3666	0,361	0,3629	0,3641	0,3637	0,3597	0,3923	0,3593
	2	2,19	0,7359	0,7264	0,7274	0,7299	0,7297	0,7297	0,7576	0,7188
	3	2,414	1,0939	1,0848	1,0854	1,0858	1,0866	1,0848	1,1142	1,0561
	4	2,627	1,4507	1,4408	1,4455	1,4404	1,4439	1,4409	1,4682	1,3377
	5	2,818	1,7735	1,7965	1,7509	1,7716	1,7522	1,773	1,8234	1,6119
	6	3,01	2,0598	2,1499	2,0363	2,0538	2,0354	2,0557	2,1087	1,8278
	7	3,199	2,3177	2,4559	2,2784	2,3258	2,2808	2,325	2,3924	2,0442
	8	3,368	2,5369	2,8105	2,4973	2,5436	2,4996	2,5431	2,6183	2,2518
	9	3,536	2,7561	3,1635	2,7162	2,7625	2,8174	2,7606	2,8361	2,4393
	10	3,705	2,9571	3,5172	3,0642	2,9669	3,1704	2,9654	3,0516	2,6266

Tabla 34
Exactitud vs (solidity 0.6, umbral y nivel) para vecino más cercano

Con solidity 0.6										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,161	99.8	99,22	99,643	99,565	99,5	98,325	98,396	97,496
	2	99,045	99,63	100	99,545	99,444	99.8	99,474	95,167	97,283
	3	99,087	99,524	99,643	99,153	99.8	99,111	99,6	96,009	97,937
	4	99,5	99,545	99,048	99,6	99,091	99,117	97,763	96,421	97,417
	5	98,265	99.8	99,087	99.8	98,797	99,23	98,513	97,914	98,201
	6	97,948	96,612	95,275	96,699	95,715	96,946	96,5	95,868	96,456
	7	99	96,405	97,09	98,485	99.8	97,873	96,981	97,056	95,651
	8	95,536	96,935	97,358	96,899	96,392	98,051	96,634	95,103	95,782
	9	96,755	95,564	97,943	99,048	95,481	98,947	96,477	97,802	95,529
	10	95,609	93,386	97,267	98,889	96,941	99,286	97,682	97,951	97,888

Tabla 35
Eficiencia vs (solidity 0.6, umbral y nivel) para vecino más cercano

Con solidity 0.6										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,8584	0,365	0,3616	0,3654	0,366	0,3643	0,3621	0,3694	0,3622
	2	2,1538	0,7338	0,7256	0,7345	0,7312	0,7311	0,7299	0,7399	0,7277
	3	2,3795	1,0919	1,0642	1,0929	1,0905	1,0797	1,0876	1,0949	1,082
	4	2,5852	1,4463	1,3481	1,4505	1,4473	1,3658	1,4431	1,4665	1,4347
	5	2,7773	1,7764	1,6216	1,8087	1,7721	1,6474	1,7765	1,8217	1,7565
	6	2,9694	2,0612	1,8403	2,0966	2,0623	1,8658	2,0583	2,1738	2,0383
	7	3,1523	2,333	2,0592	2,3828	2,3265	2,086	2,3278	2,5301	2,2987
	8	3,322	2,5496	2,2637	2,6085	2,5616	2,3424	2,5865	2,8839	2,5152
	9	3,4906	2,7666	2,4522	2,9555	2,7854	2,6974	2,9405	3,2099	2,7318
	10	3,6605	2,9721	2,6437	3,3121	2,9901	3,0543	3,295	3,4963	2,9335

Tabla 36
Exactitud vs (solidity 0.7, umbral y nivel) para vecino más cercano

Con solidity 0.7										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,615	99,167	98,571	99,643	99,583	99,667	99,545	97,562	98,485
	2	99	99,677	99,565	99,124	99,6	99,5	98,379	98,947	96,988
	3	97,545	97,95	97,88	98,472	99,722	99,6	99,069	98,531	97,569
	4	96,203	97,348	96,697	95,744	95,509	96,198	95,25	98,367	96,406
	5	96,713	96,674	97,947	97,324	96,655	97,614	98,462	96,226	95,595
	6	97,363	95,8	98,733	96,695	96,369	95,773	97,32	98,584	97,475
	7	97,12	97,577	96,543	98,262	96,596	96,068	96,946	95,132	97,607
	8	96,029	96,706	96,004	97,222	96,695	96,38	96,464	97,018	97,48
	9	96,441	96,007	94,784	96,073	94,153	95,695	95,654	95,609	94,378
	10	96,481	97,429	97,425	96,483	96,857	98,365	98,766	98,732	98,082

Tabla 37
Eficiencia vs (solidity 0.7, umbral y nivel) para vecino más cercano

Con solidity 0.7										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,8826	0,3652	0,3642	0,3637	0,3705	0,366	0,3647	0,3624	0,3619
	2	2,1914	0,7322	0,7295	0,7301	0,7496	0,7317	0,7302	0,727	0,7252
	3	2,417	1,0901	1,0857	1,0803	1,1058	1,088	1,087	1,0861	1,0802
	4	2,6272	1,3839	1,4367	1,3629	1,4568	1,44	1,4392	1,4405	1,4323
	5	2,8168	1,6667	1,7354	1,6439	1,7687	1,7671	1,7583	1,7556	1,7692
	6	3,0076	1,9044	2,0187	1,8633	2,0497	2,0491	2,0419	2,0376	2,0516
	7	3,195	2,1247	2,2642	2,0833	2,3025	2,3155	2,3042	2,2927	2,3315
	8	3,3615	2,3405	2,4794	2,29	2,5169	2,5304	2,5207	2,5078	2,5492
	9	3,5272	2,53	2,6947	2,4783	2,7316	2,7454	2,7371	2,7226	2,7649
	10	3,6954	2,7208	2,889	2,6684	2,9308	2,948	2,9375	2,9216	2,9734

Tabla 38
Exactitud vs (solidity 0.8, umbral y nivel) para vecino más cercano

Con solidity 0.8										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	98,7	98,876	99,8	99,545	98,079	99,545	99,375	97,977	96,434
	2	98,456	99,643	99	99,8	99,303	99,251	98,746	96,838	96,709
	3	98,048	99,091	98,888	99,31	99,545	99,8	99,615	98,767	96,269
	4	97,669	95,692	95,758	95,341	96,812	97,382	96,214	95,935	96,054
	5	96,308	95,569	97,302	96,925	97,672	95,626	96,56	96,464	94,877
	6	95,861	98,34	95,551	96,848	96,575	98,065	98,019	96,299	98,429
	7	98,171	94,747	96,291	97,219	98,486	98,377	97,725	96,886	97,496
	8	96,221	95,667	95,265	94,697	95,344	94,731	96,175	94,556	97,029
	9	95,578	95,499	95,523	96,335	97,581	94,857	96,658	95,455	96,926
	10	98,027	97,263	98,118	96,853	99,143	99,13	98,76	98,73	99,545

Tabla 39
Eficiencia vs (solidity 0.8, umbral y nivel) para vecino más cercano

Con solidity 0.8										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,9635	0,3654	0,3642	0,3659	0,3642	0,3638	0,3648	0,3633	0,3616
	2	2,3162	0,7323	0,7427	0,7337	0,7302	0,7303	0,7431	0,7263	0,7248
	3	2,5436	1,092	1,0995	1,0899	1,0862	1,0854	1,098	1,0803	1,0769
	4	2,8402	1,4153	1,4474	1,4381	1,4359	1,4335	1,4464	1,4274	1,425
	5	3,195	1,6987	1,7514	1,759	1,7546	1,7408	1,7531	1,7305	1,7521
	6	3,5549	1,9849	2,0337	2,0409	2,0355	2,0213	2,0332	2,0108	2,0321
	7	3,9177	2,3128	2,2828	2,3063	2,3018	2,2727	2,2881	2,2598	2,302
	8	4,2246	2,6649	2,4966	2,5203	2,5155	2,4862	2,5016	2,4729	2,515
	9	4,5081	3,0168	2,7116	2,7345	2,7416	2,6996	2,7149	2,686	2,7281
	10	4,7598	3,3623	2,9099	2,9375	3,0359	2,8957	2,9128	2,8831	2,9314

Tabla 40
Exactitud vs (solidity 0.9, umbral y nivel) para vecino más cercano

Con solidity 0.9										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,065	99,165	99,1	98,67	99,615	98,176	98,571	97,44	96,567
	2	99,8	98,936	99,149	99,545	98,591	99,13	99,375	98,083	97,852
	3	95,043	98,722	98,413	97,874	98,667	98,885	98,221	98,675	98,136
	4	95,099	95,791	94,668	96,612	94,087	95,988	94,964	96,482	95,779
	5	95,785	95,513	96,221	96,8	96,48	97,424	94,823	93,264	96,497
	6	96,343	95,833	97,513	97,937	97,635	99,048	98,438	98,199	98,29
	7	96,958	94,271	96,548	96,783	97,457	97,957	96,515	97,378	97,609
	8	96,155	94,508	96,236	97,367	95,771	95,739	95,412	96,191	95,691
	9	95,758	95,461	95,421	94,991	96,219	94,989	96,053	94,557	97,388
	10	98,036	99,139	97,619	98,258	97,981	97,865	98,412	96,89	98,822

Tabla 41
Eficiencia vs (solidity 0.9, umbral y nivel) para vecino más cercano

Con solidity 0.9										
Efi. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,0001	0,36	0,3642	0,3636	0,3652	0,3637	0,3647	0,3633	0,3597
	2	2,4425	0,7244	0,7297	0,7296	0,7326	0,7315	0,7299	0,7276	0,7212
	3	2,7325	1,0802	1,0864	1,0883	1,0927	1,0876	1,0861	1,0836	1,0733
	4	3,0008	1,4269	1,4357	1,4341	1,4435	1,4343	1,4343	1,4304	1,4205
	5	3,2151	1,7412	1,7545	1,7719	1,7766	1,7441	1,7831	1,7441	1,7467
	6	3,434	2,0253	2,0366	2,0543	2,0623	2,0261	2,1359	2,0245	2,0259
	7	3,6419	2,286	2,3027	2,3355	2,3374	2,2825	2,4673	2,2817	2,3006
	8	3,8326	2,4996	2,5169	2,5498	2,5599	2,5689	2,7458	2,4958	2,5163
	9	4,0189	2,7141	2,7309	2,7647	2,7804	2,9198	3,0116	2,7089	2,7295
	10	4,2026	2,9137	2,9349	2,9766	2,989	3,2748	3,2284	2,9094	2,9348

De todos los valores buscados se encontró la máxima exactitud de 99.8% con una eficiencia de 0.3619 segundos con los siguientes parámetros:

- *Umbral* = 0.3
- *Solidy* = 0.1
- *Nivel* = 1

Ahora trabajaremos con la técnica de árbol de decisión, para encontrar los parámetros deseados, en las Tablas 42 a 59 se podrá ver dichas simulaciones para encontrar la máxima eficiencia y exactitud con dichas condiciones.

Tabla 42
Exactitud vs (solidity 0.1, umbral y nivel) para árbol de decisión

Con solidity 0.1										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	99,161	99.8	98,947	98,593	97,7	99,183	98,254	98,6	98,095
	2	98,178	98,783	98,429	99,107	98,636	98,214	98,664	98,35	97,147
	3	99,688	98,429	98,476	98,667	98,426	97,631	97,242	99.8	95,167
	4	98,184	99,643	97,921	98,639	98,717	97,222	99,145	99,565	98,201
	5	98,729	98,575	97,292	97,649	98,831	98,544	99,2	99,195	97,215
	6	99,655	97,827	98,252	97,825	97,843	97,886	99,286	98,69	98,082
	7	98,814	99.8	97,292	96,667	98,73	97,436	98,389	96,806	98,028
	8	98,613	98,894	98,336	97,929	97,336	98,297	98,333	98,613	96,864
	9	98,528	99,583	99,081	98,446	98,539	98,137	97,745	99,223	98,123
	10	97,982	98,426	97,862	99,13	97,137	98,083	98,75	98,486	98,674

Tabla 43
Eficiencia vs (solidity 0.1, umbral y nivel) para árbol de decisión

Con solidity 0.1										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	6,2685	0,6464	0,6135	0,4265	0,5444	0,4936	0,6433	0,4253	0,328
	2	6,8387	1,2365	1,0898	0,7826	1,022	0,9081	1,2518	0,7805	0,596
	3	7,4596	1,831	1,5455	1,1352	1,4717	1,3154	1,894	1,1357	0,81
	4	8,0921	2,2767	2,2346	1,4891	1,9107	1,696	2,4173	1,4971	1,016
	5	8,6478	2,5824	2,8072	1,8433	2,3793	2,0505	2,8879	1,8445	1,235
	6	9,0417	2,8838	3,3463	2,1563	3,055	2,3878	3,2979	2,1604	1,452
	7	9,4123	3,1884	3,898	2,4678	3,6081	2,7008	3,67	2,4765	1,666
	8	10,041	3,8352	4,4034	2,7797	4,2065	2,953	4,0287	2,7366	1,881
	9	10,601	4,3826	4,8625	3,465	4,8939	3,1963	4,3455	3,0423	2,094
	10	11,103	4,8881	5,2959	4,0207	5,5537	3,44	4,653	3,2926	2,312

Tabla 44
Exactitud vs (solidity 0.2, umbral y nivel) para árbol de decisión

Con solidity 0.2										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	98,397	99,145	98,469	98,861	98,134	98,476	98,947	98,346	97,321
	2	97,583	98,615	99,2	98,304	97,458	98,965	99,615	96,096	99,183
	3	98,8	99,024	98,696	98,246	97,823	97,939	99,255	97,97	98,365
	4	96,814	98,776	98,784	98,216	98,052	97,674	98,503	99	97,103
	5	99,63	98,947	99,048	97,421	98,537	98,933	98,632	96,46	97,171
	6	98,591	98,814	97,657	98,868	97,957	98,775	97,355	98,783	97,447
	7	97,89	98,661	98,042	99,13	98,615	98,699	98,477	98,2	98,276
	8	98,696	98,464	97,258	97,318	97,629	97,649	96,9	98,476	97,083
	9	98,086	99,107	98,341	97,601	97,963	97,533	98,822	98,333	96,282
	10	99,63	98,027	98,977	98,2	98,49	99,15	98,977	98,493	97,705

Tabla 45
Eficiencia vs (solidity 0.2, umbral y nivel) para árbol de decisión

Con solidity 0.2										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,029	0,2038	0,5947	0,6145	0,606	0,366	0,6757	0,5762	0,2307
	2	2,401	0,3986	1,0205	1,0976	1,174	0,657	1,275	1,0617	0,4435
	3	2,658	0,6004	1,4177	1,5697	1,66	0,915	1,7624	1,4774	0,6665
	4	2,904	0,8001	1,7781	2,1322	2,123	1,111	2,2408	1,8668	0,8896
	5	3,138	0,9979	2,0706	2,5304	2,581	1,316	2,58	2,2306	1,1112
	6	3,345	1,198	2,3647	2,9049	3,038	1,531	2,9281	2,5711	1,3289
	7	3,549	1,4023	2,6308	3,2476	3,451	1,741	3,2256	2,8988	1,5455
	8	3,755	1,5995	2,9014	3,5703	3,856	1,951	3,5236	3,1649	1,7649
	9	3,959	1,7921	3,1173	3,8631	4,27	2,162	3,7592	3,4285	1,9822
	10	4,156	1,9821	3,3196	4,1209	4,661	2,372	3,9734	3,6593	2,1996

Tabla 46
Exactitud vs (solidity 0.3, umbral y nivel) para árbol de decisión

Con solidity 0.3										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	94,397	96,515	93,191	94,835	97,429	98,429	98,783	99,583	97,7
	2	96,231	93,593	94,82	94,046	98,196	97,965	98,235	99,005	97,771
	3	94,803	94,795	95,012	95,159	97,05	97,262	99,706	99,5	96,543
	4	93,003	94,751	94,124	95,434	97,139	97,546	98,676	99,259	96,371
	5	96,086	96,535	94,765	93,028	99,286	97,446	98,03	98,226	96,833
	6	95,292	94,615	93,851	94,438	97,656	99,243	97,265	97,104	99,183
	7	95,351	93,673	94,684	93,889	98,365	98,869	97,403	98,513	97,167
	8	95,171	94,103	94,032	95,049	99,13	99,091	97,191	98,619	98,32
	9	96,111	95,016	94,762	92,908	97,917	98,648	99,012	97,806	97,076
	10	95,236	91,836	91,825	94,15	97,679	97,814	98,696	96,926	96,717

Tabla 47
Eficiencia vs (solidity 0.3, umbral y nivel) para árbol de decisión

Con solidity 0.3										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,21	0,4409	0,4461	0,2917	0,5606	0,4466	0,7277	0,2393	0,4682
	2	2,857	0,8636	0,8333	0,5672	1,0253	0,8635	1,2993	0,4468	0,9216
	3	3,328	1,2179	1,1846	0,832	1,4503	1,2098	1,859	0,6538	1,2947
	4	3,796	1,5498	1,5266	1,1107	1,794	1,5765	2,5349	0,8648	1,6044
	5	4,262	1,8625	1,8294	1,3975	2,077	1,9825	3,248	1,0798	1,9029
	6	4,712	2,147	2,1212	1,6037	2,3671	2,3572	3,909	1,2985	2,1991
	7	5,16	2,3964	2,37	1,8001	2,6776	2,6662	4,4374	1,5129	2,4949
	8	5,606	2,6488	2,6133	2,0021	2,9603	2,9617	4,9186	1,7227	2,7166
	9	6,025	2,8638	2,8299	2,2015	3,1711	3,2498	5,3264	1,9288	2,9352
	10	6,666	3,0846	3,0297	2,3974	3,3742	3,5771	5,7123	2,1406	3,1547

Tabla 48
Exactitud vs (solidity 0.4, umbral y nivel) para árbol de decisión

Con solidity 0.4										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,129	95,875	94,618	94,743	93,272	92,787	97,014	95,592	90,518
	2	96,095	96,411	94,649	93,361	93,973	93,239	96,889	94,389	91,157
	3	96,327	96,147	92,771	94,25	92,761	94,466	95,459	96,244	91,465
	4	93,732	96,463	94,468	95,002	94,738	94,13	95,112	94,623	90,637
	5	96,066	96,32	94,642	94,331	94,268	94,429	96,393	95,795	89,886
	6	91,486	95,413	93,511	95,176	93,367	93,581	92,298	92,23	90,863
	7	95,146	92,704	92,539	93,097	95,359	97,578	93,921	94,7	93,805
	8	94,416	92,976	91,214	93,514	95,105	96,51	95,973	97,148	92,922
	9	96,154	94,784	96,43	95,46	94,011	94,116	95,696	97,285	93,487
	10	89,926	91,552	92,78	93,354	96,563	94,719	94,297	92,851	92,487

Tabla 49
Eficiencia vs (solidity 0.4, umbral y nivel) para árbol de decisión

Con solidity 0.4										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	1,9083	0,394	0,2347	0,5598	0,5712	0,3734	0,2295	0,4753	0,2248
	2	2,6483	0,733	0,4288	1,0504	1,2697	0,628	0,4297	0,8838	0,434
	3	3,1884	1,263	0,632	1,5613	1,9705	1,0104	0,6304	1,2976	0,6494
	4	3,6322	1,95	0,8391	1,9536	2,6197	1,7069	0,8363	1,6887	0,8681
	5	4,0475	2,624	1,0386	2,3256	3,3068	2,417	1,0386	2,0617	1,0903
	6	4,4562	3,187	1,245	2,6747	4,0436	3,1152	1,2517	2,4181	1,3112
	7	4,8304	3,646	1,4444	2,9863	4,6717	3,6351	1,4604	2,7479	1,5247
	8	5,1823	4,071	1,6479	3,2665	5,1899	4,1124	1,6793	3,0187	1,7419
	9	5,556	4,468	1,8494	3,5292	5,6364	4,5224	1,8876	3,2758	1,9566
	10	5,8816	4,85	2,057	3,7782	6,0425	4,9031	2,0926	3,5128	2,1669

Tabla 50
Exactitud vs (solidity 0.5, umbral y nivel) para árbol de decisión

Con solidity 0.5										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	96,007	96,068	94,666	95,012	94,129	93,72	96,077	95,522	88,677
	2	96,434	95,846	94,241	95,471	92,898	93,974	95,992	94,676	90,976
	3	95,278	95,064	94,798	94,737	92,23	93,387	96,578	95,074	91,856
	4	95,498	96,615	96,485	95,092	95,297	97,121	95,661	96,402	91,406
	5	95,432	94,589	95,909	94,047	94,916	92,778	96,618	96,455	92,114
	6	91,953	92,318	92,853	93,052	92,575	94,57	93,706	94,068	94,691
	7	95,363	95,549	94,577	93,458	92,693	96,492	92,935	94,166	91,458
	8	91,727	91,852	92,069	90,452	90,279	94,493	92,623	95,91	93,887
	9	92,266	92,767	96,727	96,417	94,625	94,327	96,115	97,163	93,943
	10	92,29	94,103	93,505	92,769	91,386	93,47	94,08	94,744	92,337

Tabla 51
Eficiencia vs (solidity 0.5, umbral y nivel) para árbol de decisión

Con solidity 0.5										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,683	0,6176	0,2929	0,5663	0,3869	0,2205	0,6255	0,3217	0,2304
	2	3,2231	1,1649	0,5446	1,0253	1,053	0,4223	1,1201	0,5953	0,4427
	3	3,5288	1,7298	0,7569	1,4671	1,6199	0,6262	1,607	0,8711	0,6632
	4	3,7732	2,1838	0,9516	1,8412	2,0725	0,8232	2,0373	1,1272	0,8869
	5	4,0178	2,6315	1,1588	2,1998	2,5025	1,0236	2,4516	1,3896	1,106
	6	4,2765	3,0468	1,3691	2,5433	2,9088	1,2279	2,8002	1,6497	1,3197
	7	4,5267	3,442	1,5772	2,8178	3,2643	1,4232	3,1244	1,8951	1,5268
	8	4,7417	3,8231	1,7916	3,036	3,6232	1,6311	3,4032	2,1861	1,7515
	9	4,9584	4,1691	1,9948	3,2344	3,9538	1,835	3,6065	2,448	1,9633
	10	5,1689	4,4844	2,2007	3,4468	4,2308	2,0419	3,8154	2,7665	2,1658

Tabla 52
Exactitud vs (solidity 0.6, umbral y nivel) para árbol de decisión

Con solidity 0.6										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	96,735	95,033	95,765	94,039	94,782	91,778	96,629	95,888	91,807
	2	96,13	94,652	92,436	94,898	93,883	92,787	94,297	94,475	91,618
	3	94,392	96,443	96,036	96,455	94,105	93,679	95,827	95,81	91,244
	4	95,174	95,822	95,763	94,001	95,187	96,073	94,774	92,183	88,697
	5	94,903	96,04	95,113	95,547	94,699	96,399	95,664	96,182	91,231
	6	89,421	89,586	91,457	91,356	94,726	92,38	95,546	94,099	94,298
	7	94,13	92,255	90,581	94,319	94,709	94,391	93,927	94,446	94,945
	8	88,063	86,705	87,835	86,707	88,522	89,116	86,176	86,352	86,203
	9	93,787	92,878	92,976	93,362	95,461	95,178	92,475	90,848	89,652
	10	95,399	94,715	95,06	95,21	93,836	95,124	92,183	95,042	93,301

Tabla 53
Eficiencia vs (solidity 0.6, umbral y nivel) para árbol de decisión

Con solidity 0.6										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,5626	0,5086	0,4559	0,2164	0,2615	0,5474	0,6074	0,6394	0,2359
	2	3,029	0,9053	0,8513	0,421	0,5089	1,0151	1,0742	1,3051	0,4502
	3	3,4222	1,25	1,2322	0,6213	0,7625	1,419	1,4525	2,0066	0,6671
	4	3,7973	1,5908	1,618	0,8232	1,0014	1,802	1,8324	2,6403	0,8913
	5	4,1532	1,9022	1,9843	1,0316	1,2362	2,1675	2,187	3,1564	1,1119
	6	4,5277	2,2089	2,3307	1,243	1,4749	2,5796	2,5924	3,6316	1,3309
	7	4,8936	2,4791	2,6643	1,4435	1,7198	2,9948	3,0022	4,0534	1,5427
	8	5,272	2,7501	2,94	1,6589	1,978	3,3541	3,3387	4,4773	1,7702
	9	5,6509	2,9746	3,2087	1,8723	2,2406	3,613	3,6638	5,0013	1,9924
	10	5,9885	3,1955	3,4335	2,0865	2,4631	3,8628	3,9271	5,512	2,202

Tabla 54
Exactitud vs (solidity 0.7, umbral y nivel) para árbol de decisión

Con solidity 0.7										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,695	95,058	96,082	94,012	93,644	92,524	95,635	94,537	89,514
	2	94,931	95,246	94,27	95,806	93,862	92,966	96,144	93,006	91,715
	3	94,541	92,416	95,302	95,858	94,832	95,624	95,189	95,904	89,357
	4	89,374	89,358	89,185	87,879	89,918	87,896	87,011	86,967	83,948
	5	90,11	90,819	91,934	89,29	90,743	88,896	88,546	89,708	84,632
	6	89,168	89,061	89,656	91,369	92,225	93,642	94,183	96,033	95,181
	7	90,289	90,043	90,131	90,329	89,064	93,277	94,398	94,759	97,227
	8	83,869	82,853	84,171	82,006	84,749	83,645	82,12	81,231	79,315
	9	89,781	90,468	90,991	92,052	92,344	92,33	89,98	90,25	85,963
	10	94,775	95,684	94,904	94,782	95,701	95,004	94,935	94,528	93,459

Tabla 55
Eficiencia vs (solidity 0.7, umbral y nivel) para árbol de decisión

Con solidity 0.7										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,6644	0,4917	0,7453	0,2729	0,5205	0,4402	0,4201	0,3384	0,2332
	2	3,5228	1,0318	1,4439	0,4781	0,9845	0,7953	0,6991	0,6873	0,4502
	3	3,9998	1,7533	2,1667	0,6806	1,462	1,0555	1,0142	1,0182	0,663
	4	4,5315	2,3747	2,8998	0,9027	1,9651	1,7953	1,3845	1,3498	0,889
	5	5,2519	2,8706	3,4775	1,196	2,4257	2,4378	1,7368	1,7015	1,1101
	6	6,0307	3,3334	4,0697	1,4097	3,1162	3,0441	1,9505	1,9643	1,32
	7	6,681	3,7401	4,8491	1,6237	3,6997	3,6402	2,1599	2,2402	1,5257
	8	7,2846	4,1458	5,4831	1,8475	4,3346	4,3445	2,3877	2,5424	1,7557
	9	7,7899	4,5008	6,0108	2,0707	4,8379	5,1344	2,6105	2,8569	1,9848
	10	8,1871	4,8175	6,422	2,2729	5,246	5,7169	2,8106	3,063	2,1812

Tabla 56
Exactitud vs (solidity 0.8, umbral y nivel) para árbol de decisión

Con solidity 0.8										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	96,093	96,427	95,028	94,414	94,183	94,424	95,746	94,741	90,151
	2	96,522	95,281	92,331	94,729	92,356	92,711	95,504	95,2	92,001
	3	96,054	93,602	94,372	95,643	94,958	95,96	98,351	95,459	92,987
	4	85,631	83,586	84,672	83,586	83,807	84,908	82,895	84,498	81,896
	5	87,588	88,575	88,214	86,405	87,515	87,393	87,244	87,453	83,997
	6	91,513	91,15	92,13	93,675	95,526	93,689	92,29	95,215	96,669
	7	90,738	89,565	88,572	89,522	93,15	94,713	93,851	94,208	96,953
	8	83,847	82,826	84,834	82,518	83,228	82,994	84,067	81,582	83,704
	9	87,901	88,046	89,856	88,387	89,102	88,805	90,438	88,747	86,841
	10	95,082	95,068	95,363	94,939	95,63	95,112	94,491	94,524	94,013

Tabla 57
Eficiencia vs (solidity 0.8, umbral y nivel) para árbol de decisión

Con solidity 0.8										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,7919	0,6305	0,4072	0,6973	0,5081	0,2661	0,4728	0,6079	0,2615
	2	3,5616	1,2642	0,7443	1,2721	0,9368	0,5432	0,9246	1,209	0,5174
	3	4,2958	1,7455	1,0769	1,826	1,3287	0,8117	1,2657	1,7636	0,7304
	4	4,9843	2,3235	1,4216	2,5492	1,7274	1,0861	1,6642	2,2938	0,9637
	5	5,5926	2,7928	1,7197	3,21	2,0958	1,3112	2,0133	2,7345	1,1868
	6	6,1725	3,2592	1,9955	3,7391	2,4104	1,5288	2,2968	3,1402	1,3976
	7	6,778	3,6344	2,2589	4,2172	2,7108	1,7362	2,5321	3,5163	1,6056
	8	7,5783	3,9484	2,5082	4,6693	2,9842	1,9719	2,7609	3,9284	1,8394
	9	8,2918	4,2344	2,8057	5,0673	3,2512	2,2016	2,9855	4,3159	2,0675
	10	8,8246	4,4693	3,1029	5,407	3,4495	2,409	3,1966	4,6506	2,2755

Tabla 58
Exactitud vs (solidity 0.9, umbral y nivel) para árbol de decisión

Con solidity 0.9										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,769	95,302	94,864	94,385	94,535	93,059	95,506	96,48	90,655
	2	96,281	95,913	94,795	93,622	92,69	93,33	96,246	95,276	90,341
	3	92,63	91,451	92,571	94,474	96,398	94,896	95,852	96,339	95,361
	4	84,207	83,837	83,094	82,775	82,692	81,91	83,984	81,148	81,39
	5	86,216	86,872	84,918	86,791	86,734	86,441	85,382	88,996	85,495
	6	91,102	91,471	93,356	95,577	95,677	95,331	96,829	96,617	97,085
	7	93,504	91,428	90,25	91,389	93,724	94,495	94,576	95,788	96,969
	8	84,799	84,385	86,222	83,453	84,323	86,796	83,633	84,582	86,471
	9	89,251	88,452	89,492	89,517	89,278	88,639	88,341	90,78	88,384
	10	94,458	96,12	95,298	95,646	95,79	94,766	94,812	94,557	93,178

Tabla 59
Eficiencia vs (solidity 0.9, umbral y nivel) para árbol de decisión

Con solidity 0.9										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,393	0,6445	0,2116	0,5379	0,2124	0,2212	0,5014	0,7247	0,2682
	2	2,8213	1,2044	0,4115	0,9909	0,4094	0,4713	0,9172	1,3056	0,496
	3	3,2213	1,6525	0,6176	1,4566	0,5991	0,761	1,3013	1,8267	0,717
	4	3,5951	2,1093	0,8482	1,862	0,8333	1,0723	1,706	2,4048	0,9634
	5	3,9371	2,5071	1,0702	2,1725	1,0564	1,3783	2,0754	2,9268	1,1948
	6	4,2329	2,879	1,2807	2,4745	1,2581	1,6173	2,3749	3,3606	1,3978
	7	4,5072	3,2288	1,5033	2,7413	1,4723	1,8671	2,6508	3,8268	1,6179
	8	4,803	3,5253	1,732	2,9952	1,7017	2,138	2,9353	4,2805	1,855
	9	5,1372	3,8063	1,9546	3,2194	1,9286	2,398	3,1956	4,695	2,0917
	10	5,4805	4,0473	2,1572	3,4257	2,1317	2,6338	3,4332	5,0478	2,3046

Con los valores simulados se encontró que la máxima exactitud es de 99.8% y con una eficiencia de 0.6464 segundos con los siguientes parámetros:

- *Umbral* = 0.2
- *Solidy* = 0.1
- *Nivel* = 1

De nuevo simularemos los datos obtenidos anteriormente pero aplicando la técnica de Naive Bayes, para encontrar la máxima eficiencia esto se podrá ver en las Tablas 60-77.

Tabla 60
Exactitud vs (solidity 0.1, umbral y nivel) para Naive Bayes

Con solidity 0.1										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,712	95,18	95,961	95,065	95,286	93,487	93,466	93,216	91,463
	2	95,206	94,688	96	95,282	95,346	92,847	92,845	92,538	91,677
	3	95,422	95,781	95,503	95,18	94,662	93,781	92,819	93,576	91,962
	4	94,967	95,49	95,977	95,702	95,652	93,652	92,128	93,462	90,069
	5	94,638	94,705	95,528	95,171	94,491	93,244	93,407	93,535	91,36
	6	95,659	95,836	95,407	94,679	94,535	93,676	93,431	93,312	92,154
	7	95,849	96,179	95,846	95,077	95,012	92,548	93,146	92,871	91,133
	8	96,633	95,821	94,581	94,586	94,713	93,283	93,248	94,013	92,955
	9	95,718	94,736	95,393	94,29	94,344	92,72	93,987	93,803	91,777
	10	95,87	95,655	94,141	95,505	95,003	93,066	93,368	92,043	91,833

Tabla 61
Eficiencia vs (solidity 0.1, umbral y nivel) para Naive Bayes

Con solidity 0.1										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	2,4872	0,0662	0,0879	0,1206	0,0779	0,0676	0,1293	0,0635	0,0571
	2	2,606	0,1363	0,1812	0,2705	0,147	0,1378	0,2286	0,1272	0,1088
	3	2,712	0,2098	0,2498	0,41	0,2208	0,2199	0,3291	0,196	0,1592
	4	2,8104	0,2669	0,3053	0,5172	0,2881	0,2894	0,4071	0,2525	0,2317
	5	2,9112	0,3251	0,3575	0,6193	0,3507	0,3647	0,4852	0,3078	0,3065
	6	3,0112	0,3879	0,4118	0,7158	0,4078	0,4346	0,5634	0,365	0,3817
	7	3,1137	0,4482	0,4782	0,8247	0,4648	0,5081	0,6418	0,4198	0,4454
	8	3,2173	0,5066	0,5302	0,9318	0,5217	0,586	0,7209	0,493	0,5079
	9	3,3195	0,5659	0,5932	1,0294	0,5776	0,6547	0,7997	0,5894	0,5644
	10	3,4285	0,6236	0,6531	1,1089	0,633	0,7231	0,879	0,6894	0,6221

Tabla 62
Exactitud vs (solidity 0.2, umbral y nivel) para Naive Bayes

Con solidity 0.2										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,38	95,351	95,726	95,518	94,232	93,559	92,617	92,944	90,776
	2	94,75	95,782	95,654	94,771	94,973	93,82	93,811	92,735	91,289
	3	95,152	95,601	95,658	94,769	95,085	93,446	92,474	94,482	92,274
	4	95,399	95,696	95,997	94,884	93,694	93,061	92,975	92,667	92,428
	5	95,14	93,968	96,079	95,309	94,057	93,426	93,996	93,34	91,708
	6	94,731	94,584	94,715	94,573	93,473	92,841	92,359	92,402	91,885
	7	95,193	94,548	94,369	94,435	94,712	93,625	92,816	91,929	90,798
	8	93,715	93,946	94,983	95,314	93,453	93,589	93,034	93,349	92,561
	9	95,067	95,577	95,75	94,581	94,009	93,583	93,226	93,152	91,889
	10	93,959	95,141	93,732	94,764	94,588	93,692	92,743	93,602	91,607

Tabla 63
Eficiencia vs (solidity 0.2, umbral y nivel) para Naive Bayes

Con solidity 0.2										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	0,5164	0,0608	0,0499	0,117	0,0671	0,064	0,0644	0,0356	0,0765
	2	0,5889	0,1231	0,1004	0,22	0,1172	0,123	0,1212	0,0672	0,1551
	3	0,6394	0,1869	0,1551	0,32	0,1863	0,179	0,1781	0,0973	0,237
	4	0,6832	0,2493	0,1986	0,422	0,2552	0,23	0,2388	0,1299	0,3181
	5	0,7261	0,3008	0,2478	0,516	0,3213	0,28	0,3079	0,1577	0,3921
	6	0,7693	0,351	0,2973	0,595	0,3779	0,331	0,3579	0,1856	0,4666
	7	0,814	0,403	0,3395	0,673	0,4312	0,381	0,4077	0,2133	0,5344
	8	0,8583	0,4534	0,3797	0,75	0,4814	0,43	0,4638	0,2409	0,5962
	9	0,9024	0,5044	0,4267	0,828	0,5261	0,48	0,5169	0,2685	0,6589
	10	0,9458	0,5558	0,4636	0,91	0,5681	0,53	0,5553	0,296	0,7226

Tabla 64
Exactitud vs (solidity 0.3, umbral y nivel) para Naive Bayes

Con solidity 0.3										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,55	95,705	97,88	95,747	94,464	92,814	92,762	92,98	92,563
	2	94,822	96,091	98,315	96,064	94,636	93,843	93,462	93,297	92,289
	3	96,151	94,891	98,297	95,143	93,699	93,928	92,882	92,656	91,76
	4	94,843	95,852	97,879	95,336	95,484	92,622	92,843	92,194	92,038
	5	94,559	96,371	97,917	95,526	94,457	92,997	93,527	93,933	92,317
	6	92,584	93,927	97,011	94,894	92,771	92,949	92,081	92,516	91,728
	7	92,259	93,708	97,047	93,917	94,293	93,935	93,053	92,979	92,224
	8	93,656	95,07	97,88	94,864	92,773	92,892	93,409	92,696	91,11
	9	95,492	95,397	97,011	95,67	93,946	93,639	93,836	93,51	91,56
	10	93,603	93,166	91,522	93,481	92,716	93,034	93,102	93,054	91,589

Tabla 65
Eficiencia vs (solidity 0.3, umbral y nivel) para Naive Bayes

Con solidity 0.3										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	0,4394	0,0784	0,1007	0,0372	0,0712	0,0708	0,1092	0,0912	0,0413
	2	0,5124	0,1469	0,2016	0,0793	0,137	0,1471	0,2379	0,1837	0,0957
	3	0,5669	0,2209	0,2758	0,1192	0,2313	0,2187	0,3437	0,2736	0,1342
	4	0,6149	0,2892	0,3439	0,1606	0,2982	0,2993	0,4389	0,3453	0,1682
	5	0,659	0,3541	0,4105	0,1958	0,3648	0,3769	0,5343	0,4146	0,2021
	6	0,7034	0,411	0,4745	0,2312	0,4311	0,4452	0,6206	0,4794	0,238
	7	0,7478	0,4668	0,5547	0,2667	0,4971	0,5171	0,6995	0,5416	0,2715
	8	0,7925	0,5225	0,6291	0,3024	0,5631	0,5945	0,7808	0,6092	0,3048
	9	0,8364	0,5783	0,7068	0,3374	0,6301	0,661	0,8638	0,681	0,3331
	10	0,8814	0,6339	0,7769	0,3723	0,696	0,7228	0,9416	0,7586	0,3611

Tabla 66
Exactitud vs (solidity 0.4, umbral y nivel) para Naive Bayes

Con solidity 0.4										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,129	95,875	94,618	94,743	93,272	92,787	97,014	95,592	90,518
	2	96,095	96,411	94,649	93,361	93,973	93,239	96,889	94,389	91,157
	3	96,327	96,147	92,771	94,25	92,761	94,466	95,459	96,244	91,465
	4	93,732	96,463	94,468	95,002	94,738	94,13	95,112	94,623	90,637
	5	96,066	96,32	94,642	94,331	94,268	94,429	96,393	95,795	89,886
	6	91,486	95,413	93,511	95,176	93,367	93,581	92,298	92,23	90,863
	7	95,146	92,704	92,539	93,097	95,359	97,578	93,921	94,7	93,805
	8	94,416	92,976	91,214	93,514	95,105	96,51	95,973	97,148	92,922
	9	96,154	94,784	96,43	95,46	94,011	94,116	95,696	97,285	93,487
	10	89,926	91,552	92,78	93,354	96,563	94,719	94,297	92,851	92,487

Tabla 67
Eficiencia vs (solidity 0.4, umbral y nivel) para Naive Bayes

Con solidity 0.4										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	0,5367	0,0997	0,1024	0,112	0,089	0,0372	0,0368	0,109	0,0615
	2	0,6997	0,1973	0,2073	0,2125	0,179	0,0725	0,0696	0,2105	0,1169
	3	0,8249	0,2937	0,3032	0,3336	0,253	0,1037	0,1029	0,3082	0,174
	4	0,9313	0,3837	0,3803	0,4446	0,313	0,1393	0,1313	0,4069	0,2286
	5	1,0299	0,4621	0,4535	0,5263	0,381	0,1699	0,1595	0,4839	0,3008
	6	1,1284	0,5412	0,5226	0,6037	0,448	0,1987	0,1873	0,5597	0,3603
	7	1,2295	0,6188	0,5841	0,6815	0,518	0,2272	0,2152	0,6355	0,431
	8	1,3297	0,6968	0,6679	0,7586	0,575	0,2564	0,2431	0,7115	0,4836
	9	1,4292	0,7742	0,7491	0,8358	0,631	0,2844	0,271	0,7898	0,535
	10	1,5283	0,8516	0,8278	0,9134	0,686	0,3121	0,2989	0,866	0,5873

Tabla 68
Exactitud vs (solidity 0.5, umbral y nivel) para Naive Bayes

Con solidity 0.5										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,317	95,35	94,8	95,797	94,583	92,91	93,668	92,834	90,719
	2	95,304	96,21	94,73	94,549	94,094	93,155	93,274	93,124	92,04
	3	93,94	94,71	95,28	94,689	94,786	93,813	93,626	93,65	92,698
	4	92,713	92,75	92,25	92,706	93,564	93,903	93,371	93,294	89,147
	5	93,1	93,38	94,37	95,226	95,14	94,976	93,675	92,706	90,993
	6	92,392	93,25	91,3	93,464	92,272	91,855	92,448	92,25	91,792
	7	93,584	93,5	92,36	93,149	94,192	93,833	92,865	92,891	93,65
	8	90,159	89,95	91,19	90,755	90,814	90,986	90,54	89,637	88,241
	9	90,217	90,07	90,55	89,923	92,374	92,523	92,166	91,787	91,435
	10	89,001	89,75	89,99	91,269	91,01	94,142	94,06	93,421	93,85

Tabla 69
Eficiencia vs (solidity 0.5, umbral y nivel) para Naive Bayes

Con solidity 0.5										
Efi. (s)		Coefficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	0,5025	0,0944	0,0337	0,093	0,0464	0,0844	0,0645	0,1018	0,0515
	2	0,5868	0,1875	0,0663	0,181	0,0992	0,163	0,1192	0,218	0,104
	3	0,6769	0,2484	0,095	0,27	0,1494	0,2487	0,1771	0,3117	0,1564
	4	0,7368	0,3087	0,1242	0,352	0,2012	0,3311	0,243	0,3969	0,2111
	5	0,7978	0,3799	0,1531	0,417	0,2524	0,4154	0,3144	0,4958	0,2549
	6	0,8479	0,473	0,1818	0,481	0,3023	0,5021	0,3638	0,5929	0,2958
	7	0,8994	0,5615	0,2104	0,546	0,3404	0,5945	0,4097	0,6907	0,338
	8	0,9497	0,6617	0,2411	0,61	0,3731	0,6782	0,4541	0,7786	0,3834
	9	0,9998	0,7589	0,2716	0,674	0,4095	0,7622	0,5046	0,8645	0,4338
	10	1,0472	0,8359	0,3076	0,738	0,4559	0,8518	0,5504	0,935	0,4692

Tabla 70
Exactitud vs (solidity 0.6, umbral y nivel) para Naive Bayes

Con solidity 0.6										
Exac. (%)		Coefficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	95,166	95,318	95,459	95,925	94,734	94,045	93,493	92,371	92,12
	2	95,581	95,522	95,775	94,846	94,752	93,425	93,6	92,531	90,324
	3	93,349	93,564	94,619	95,529	95,411	94,202	93,434	93,086	93,19
	4	92,157	92,546	90,071	92,068	91,568	92,318	92,997	92,498	88,931
	5	91,602	92,201	92,47	92,859	92,716	93,368	92,133	92,889	91,141
	6	89,964	90,687	91,084	91,421	92,595	92,661	93,186	94,282	93,543
	7	91,395	90,786	91,562	92,235	92,95	93,589	93,412	94,621	93,925
	8	89,147	88,676	88,331	86,909	87,182	88,139	85,572	83,224	87,164
	9	88,702	89,169	87,676	88,592	88,877	86,517	86,317	87,054	86,193
	10	90,343	90,442	91,309	91,791	92,946	93,719	92,984	93,418	93,776

Tabla 71
Eficiencia vs (solidity 0.6, umbral y nivel) para Naive Bayes

Con solidity 0.6										
Efi. (s)		Coefficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	0,4375	0,0911	0,057	0,1086	0,045	0,1151	0,0785	0,0851	0,0897
	2	0,5122	0,1691	0,108	0,2117	0,0894	0,1975	0,1561	0,1664	0,1763
	3	0,5623	0,2383	0,164	0,3177	0,135	0,2851	0,2203	0,2486	0,2345
	4	0,5982	0,3101	0,212	0,421	0,1743	0,3602	0,2682	0,333	0,2939
	5	0,6327	0,3732	0,259	0,503	0,219	0,4366	0,3195	0,4069	0,3526
	6	0,6683	0,4371	0,302	0,5958	0,2638	0,5111	0,3887	0,4779	0,4192
	7	0,7032	0,5013	0,337	0,678	0,2968	0,5894	0,4627	0,5444	0,4788
	8	0,7366	0,5646	0,375	0,7579	0,3278	0,6671	0,5143	0,6043	0,5299
	9	0,7712	0,6355	0,412	0,8458	0,365	0,7446	0,5785	0,6643	0,5832
	10	0,8107	0,7066	0,453	0,9251	0,3997	0,8245	0,6475	0,7249	0,6401

Tabla 72
Exactitud vs (solidity 0.7, umbral y nivel) para Naive Bayes

Con solidity 0.7										
Efi. (%)		Coefficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	94,998	95,106	95,704	95,295	94,406	91,862	93,318	92,433	91,538
	2	95,434	95,404	95,676	95,852	94,908	92,621	92,992	93,331	91,059
	3	90,71	92,461	92,094	93,5	94,346	94,221	94,879	93,947	93,736
	4	87,6	88,427	87,962	87,914	88,228	88,151	88,286	87,919	88,31
	5	89,681	89,064	88,173	88,194	87,056	87,559	83,908	85,069	87,471
	6	89,912	90,772	89,276	88,916	90,774	91,955	92,671	93,691	93,343
	7	90,965	89,864	90,647	89,468	90,017	91,907	93,719	94,677	93,885
	8	86,39	86,983	86,076	85,42	84,988	84,288	80,356	78,213	86,276
	9	86,931	86,985	85,024	84,996	83,002	83,215	81,266	82,729	86,566
	10	92,2	91,228	93,369	94,525	94,124	94,867	94,054	95,063	94,475

Tabla 73
Eficiencia vs (solidity 0.7, umbral y nivel) para Naive Bayes

Con solidity 0.7										
Efi. (s)		Coficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	0,4974	0,0615	0,088	0,034	0,0335	0,0959	0,1033	0,0423	0,1103
	2	0,5716	0,1212	0,174	0,0681	0,0646	0,1775	0,2026	0,0923	0,2088
	3	0,6248	0,179	0,262	0,0974	0,0943	0,2555	0,3046	0,1336	0,287
	4	0,6684	0,2416	0,34	0,1294	0,1223	0,3203	0,381	0,1799	0,3702
	5	0,7144	0,296	0,423	0,1589	0,1538	0,3755	0,4667	0,2209	0,4437
	6	0,7639	0,352	0,511	0,1873	0,1849	0,4343	0,5458	0,2545	0,5178
	7	0,8134	0,4083	0,592	0,2166	0,2181	0,4977	0,6271	0,2923	0,594
	8	0,8574	0,4554	0,664	0,2431	0,2516	0,5546	0,7031	0,3242	0,6777
	9	0,899	0,5011	0,736	0,27	0,2792	0,6335	0,7794	0,3571	0,7535
	10	0,9339	0,5507	0,812	0,2989	0,3096	0,6972	0,8598	0,4061	0,8335

Tabla 74
Exactitud vs (solidity 0.8, umbral y nivel) para Naive Bayes

Con solidity 0.8										
Exac. (%)		Coficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	95,461	95,163	95,556	95,45	94,205	92,896	93,812	92,333	91,44
	2	94,569	95,752	94,715	95,32	95,26	93,542	92,775	93,706	90,638
	3	91,142	91,873	92,397	92,4	94,126	94,487	94,201	93,346	92,66
	4	86,037	86,399	86,125	87,42	87,544	86,668	86,819	86,024	87,661
	5	87,73	87,435	86,885	85,93	85,835	83,546	79,524	83,62	87,217
	6	88,378	90,011	89,995	91,48	91,37	92,519	92,561	93,845	94,165
	7	92,01	91,287	90,831	91,55	92,38	92,528	94,853	95,04	95,165
	8	87,505	86,55	86,073	85,51	84,528	81,932	79,851	75,184	87,035
	9	86,019	84,682	84,471	84,57	81,843	83,713	81,736	79,668	85,272
	10	91,139	90,699	92,115	93,86	94,071	92,69	93,11	93,968	93,617

Tabla 75
Eficiencia vs (solidity 0.8, umbral y nivel) para Naive Bayes

Con solidity 0.8										
Efi. (s)		Coficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	0,5144	0,0667	0,0858	0,044	0,1045	0,0617	0,0396	0,063	0,0629
	2	0,6018	0,1255	0,1735	0,163	0,1856	0,1251	0,0836	0,1209	0,1318
	3	0,6609	0,1863	0,2586	0,281	0,2711	0,1917	0,1259	0,1801	0,1913
	4	0,7062	0,2366	0,3402	0,367	0,348	0,2401	0,1574	0,2344	0,2421
	5	0,7518	0,2813	0,4151	0,455	0,4228	0,2873	0,189	0,2829	0,2953
	6	0,8018	0,3299	0,4763	0,549	0,5003	0,3373	0,2223	0,3289	0,3496
	7	0,851	0,3783	0,5336	0,647	0,5802	0,39	0,2583	0,376	0,3972
	8	0,8973	0,4227	0,5999	0,716	0,6526	0,4383	0,2904	0,4192	0,4383
	9	0,943	0,467	0,6557	0,786	0,7184	0,4848	0,3219	0,4622	0,4826
	10	0,9981	0,5154	0,7169	0,864	0,7845	0,5367	0,3559	0,5079	0,5262

Tabla 76
Exactitud vs (solidity 0.9, umbral y nivel) para Naive Bayes

Con solidity 0.9										
Exac. (%)		Coficiente para umbralización								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Niveles	1	92,807	96,196	97,88	98,297	98,279	97,899	92,752	89,004	88,895
	2	93,659	94,909	97,88	97,428	98,297	97,861	93,261	88,146	85,593
	3	71,021	75,815	90,704	91,975	94,928	95,779	93,241	87,731	86,83
	4	72,941	74,226	74,167	74,557	74,728	75,924	75,072	72,029	64,875
	5	54,954	52,5	48,588	46,268	44,507	44,456	44,993	42,355	42,808
	6	77,112	51,74	41,571	59,75	65,236	64,062	61,326	86,863	84,689
	7	51,159	39,453	66,962	67,844	68,649	67,391	65,259	67,391	72,79
	8	44,022	44,127	43,221	42,409	40,688	40,651	41,957	41,504	42,3
	9	65,199	60,185	54,24	49,984	46,091	46,196	46,104	43,17	43,241
	10	83,007	77,917	66,413	56,812	45,743	45,309	46,159	42,391	43,17

Tabla 77
Exactitud vs (solidity 0.9, umbral y nivel) para Naive Bayes

Con solidity 0.9										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	0,4974	0,0615	0,088	0,034	0,0335	0,0959	0,1033	0,0423	0,1103
	2	0,5716	0,1212	0,174	0,0681	0,0646	0,1775	0,2026	0,0923	0,2088
	3	0,6248	0,179	0,262	0,0974	0,0943	0,2555	0,3046	0,1336	0,287
	4	0,6684	0,2416	0,34	0,1294	0,1223	0,3203	0,381	0,1799	0,3702
	5	0,7144	0,296	0,423	0,1589	0,1538	0,3755	0,4667	0,2209	0,4437
	6	0,7639	0,352	0,511	0,1873	0,1849	0,4343	0,5458	0,2545	0,5178
	7	0,8134	0,4083	0,592	0,2166	0,2181	0,4977	0,6271	0,2923	0,594
	8	0,8574	0,4554	0,664	0,2431	0,2516	0,5546	0,7031	0,3242	0,6777
	9	0,899	0,5011	0,736	0,27	0,2792	0,6335	0,7794	0,3571	0,7535
	10	0,9339	0,5507	0,812	0,2989	0,3096	0,6972	0,8598	0,4061	0,8335

Con los valores simulados se encontró que la máxima exactitud es de 98.315% y con una eficiencia 0.2016 segundos, con los siguientes parámetros:

- *Umbral* = 0.3
- *Solidy* = 0.3
- *Nivel* = 2

Para la evaluación de los parámetros en la técnica ANN (patternnet), se recomienda escalar los datos de la variable (y). El escalonamiento permite dividir la información de presencia de cáncer o no, en dos columnas, la primera se encontrará solo los casos donde existan tumores y se establecerá el número indicando su presencia caso contrario será cero, para la segunda columna se observara los casos que no poseen cáncer y se asignará uno cuando esto suceda, caso contrario será cero. Esto se pudo conocer gracias a la base de datos *cancer_dataset* la cual posee la distribución de su información de esta

manera (MathWorks, 2016). En las Tablas 78-98 se puede ver el análisis para la obtención de parámetros que establecerán el sistema a usarse

Tabla 78
Exactitud vs (solidity 0.1, umbral y nivel) para ANN con red Patternnet

Con solidity 0.1										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	96,927	95,926	94,788	94,997	96,707	95,7	95,951	94,291	90,295
	2	96,621	95,825	96,715	97,184	94,254	96,41	96,206	91,864	93,256
	3	97,601	96,45	96,402	96,944	94,999	95,286	93,415	92,796	90,168
	4	97,369	96,164	97,347	94,708	97,117	96,542	93,281	94,303	93,235
	5	96,559	96,113	97,222	96,962	96,847	95,348	95,592	94,825	91,977
	6	97,236	95,899	96,74	97,523	94,318	97,834	95,167	94,103	92,37
	7	96,793	96,059	96,143	96,01	96,959	96,569	96,365	94,783	92,242
	8	96,494	97,969	95,619	96,768	96,721	95,902	96,427	93,644	92,121
	9	96,593	97,231	96,652	96,681	95,684	95,222	95,044	92,827	90,374
	10	96,849	95,447	96,22	96,986	96,659	97,032	94,364	92,022	91,927

Tabla 79
Eficiencia vs (solidity 0.1, umbral y nivel) para ANN con red Patternnet

Con solidity 0.1										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	10,23	6,7766	6,162	5,4286	5,5385	7,055	8,4235	6,7667	8,7057
	2	13,132	9,2728	12,536	7,9796	12,097	9,479	15,839	9,5407	13,402
	3	15,581	17,57	15,229	10,416	15,558	11,85	19,745	12,046	15,786
	4	18,021	25,259	17,656	12,828	17,985	14,22	26,616	14,443	22,901
	5	20,506	28,362	20,118	18,468	20,399	20,73	32,609	21,541	26,032
	6	22,979	30,768	22,597	25,681	22,909	23,68	40,815	28,016	28,472
	7	25,409	35,341	25,063	34,409	25,368	30,51	48,436	30,794	30,862
	8	32,222	44,535	27,513	39,915	27,831	34,81	52,181	33,235	33,232
	9	38,142	49,19	29,91	44,129	30,323	39,67	54,637	35,822	35,677
	10	44,441	51,577	35,124	46,548	32,819	42,99	57,071	38,273	39,973

Tabla 80
Exactitud vs (solidity 0.2, umbral y nivel) para ANN con red Patternnet

Con solidity 0.2										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,569	96,736	96,117	97,212	96,924	98,419	97,24	95,52	91,179
	2	96,259	97,49	97,043	96,91	96,587	97,76	93,464	93,243	91,02
	3	97,056	96,505	95,662	96,093	94,001	92,934	97,724	94,161	92,242
	4	95,345	96,771	94,593	96,523	95,035	96,264	95,284	92,999	92,064
	5	97,135	96,6	97,24	96,901	96,419	95,791	95,896	95,341	92,627
	6	95,66	95,127	95,611	95,471	97,14	93,364	95,822	91,479	91,687
	7	96,631	94,787	95,516	95,786	96,651	95,902	95,488	93,979	90,76
	8	97,091	95,979	96,352	94,92	96,091	97,034	94,772	92,782	91,561
	9	96,277	95,647	95,871	96,064	98,626	96,649	94,36	95,842	92,083
	10	97,688	96,575	96,249	96,042	95,801	95,649	94,748	95,652	92,611

Tabla 81
Eficiencia vs (solidity 0.2, umbral y nivel) para ANN con red Patternnet

Con solidity 0.2										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	6,6161	5,9266	6,0968	7,5825	5,9985	6,8382	8,4993	8,7847	8,4249
	2	9,003	8,561	13,079	11,015	8,7504	9,3576	11,909	12,514	16,269
	3	11,424	11,043	15,855	13,369	11,16	11,764	14,288	14,961	21,861
	4	13,756	13,542	18,296	15,783	13,588	14,132	16,766	17,356	24,632
	5	16,128	15,936	20,656	18,182	15,988	16,522	19,296	19,798	27,081
	6	18,547	18,371	23,005	24,522	18,319	18,967	21,677	22,226	29,472
	7	20,916	20,814	25,382	27,888	20,72	21,423	24,153	24,759	31,868
	8	23,266	25,428	27,779	30,247	23,142	23,883	26,625	27,121	34,272
	9	25,698	30,238	30,205	32,739	25,55	26,317	29,111	29,65	36,708
	10	32,855	35,847	34,517	38,352	30,568	28,753	31,552	32,103	39,169

Tabla 82
Exactitud vs (solidity 0.3, umbral y nivel) para ANN con red Patternnet

Con solidity 0.3										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	97,724	96,516	96,765	96,9	98	95,433	97,312	94,389	93,385
	2	97,583	94,792	96,199	97,993	97,756	96,479	96,581	91,962	93,388
	3	96,713	97,04	96,545	96,808	95,952	97,156	96,722	94,098	92,776
	4	95,922	95,845	95,614	96,145	96,174	92,494	95,332	94,346	91,248
	5	97,07	96,985	97,064	96,622	95,768	96,426	96,866	95,551	92,131
	6	95,568	96,028	93,51	94,69	96,599	95,439	93,261	91,944	92,347
	7	95,373	96,269	97,212	97,618	96,885	97,83	94,46	92,414	92,344
	8	97,212	96,16	93,46	95,481	95,333	94,179	94,369	95,118	92,067
	9	97,892	95,616	95,954	96,06	97,096	97,056	95,842	89,869	92,274
	10	94,515	95,393	94,756	92,942	96,017	95,542	95,039	91,67	92,147

Tabla 83
Eficiencia vs (solidity 0.3, umbral y nivel) para ANN con red Patternnet

Con solidity 0.3										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	18,863	10,787	9,0921	9,6653	7,5603	9,773	9,5119	9,4963	9,9271
	2	29,045	20,735	18,979	19,198	17,278	19,29	19,44	17,711	19,856
	3	39,842	30,452	28,843	28,085	26,743	28,59	29,651	27,562	29,677
	4	51,085	40,735	38,037	37,708	36,378	37,79	40,088	37,194	38,863
	5	60,811	51,161	46,975	47,749	45,908	47,15	51,231	45,908	48,527
	6	71,148	60,796	56,301	57,376	55,002	57,01	61,618	56,148	57,754
	7	81,133	70,802	65,634	66,991	65,273	66,82	71,195	66,015	67,759
	8	91,202	80,509	74,249	76,838	74,979	74,95	79,394	75,603	77,193
	9	99,995	90,267	83,778	85,649	84,646	84,79	89,692	84,649	86,672
	10	110,29	100,15	93,414	95,076	93,872	94,34	98,588	94	95,55

Tabla 84
Exactitud vs (solidity 0.4, umbral y nivel) para ANN con red Patternnet

Con solidity 0.4										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	98,29	96,809	96,875	96,461	96,473	97,992	95,898	94,354	91,414
	2	97,583	96,478	95,148	96,852	95,426	96,501	96,406	91,346	91,974
	3	96,477	96,364	97,114	94,202	96,758	96,462	96,05	92,669	93,519
	4	97,257	96,544	95,265	96,264	97,425	96,232	97,383	94,783	91,02
	5	96,259	97,429	95,524	96,578	95,989	96,387	93,595	93,482	90,416
	6	95,492	95,699	94,268	96,181	94,225	94,55	94,753	95,666	94,115
	7	95,856	94,247	95,297	97,58	95,912	96,285	96,475	92,361	95,506
	8	95,757	96,111	95,385	96,137	95,264	93,457	93,688	94,435	92,76
	9	94,984	94,483	94,761	96,483	94,918	94,891	97,18	93,896	93,531
	10	91,984	95,794	94,705	94,543	96,706	96,401	95,03	91,883	93,659

Tabla 85
Eficiencia vs (solidity 0.4, umbral y nivel) para ANN con red Patternnet

Con solidity 0.4										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	24,991	10,397	9,274	6,1318	9,8796	9,1355	7,877	6,9059	10,476
	2	39,497	16,136	16,96	13,311	17,978	14,281	16,06	13,654	14,979
	3	50,918	21,279	26,9	16,565	24,726	16,834	19,66	18,123	20,51
	4	61,553	24,679	33,07	22,296	33,986	19,337	22,27	20,698	25,849
	5	72,081	27,227	36,85	28,052	41,722	21,956	24,85	27,844	29,222
	6	81,872	29,801	43,11	31,046	44,71	29,513	27,37	32,396	31,661
	7	92,98	32,218	51,28	33,444	47,526	38,291	29,94	40,268	38,634
	8	101,07	34,71	59,39	39,815	56,402	46,683	32,52	45,277	41,795
	9	110,99	38,247	68,46	43,784	61,249	53,733	35,05	52,889	44,276
	10	121,58	43,316	73,33	48,91	63,896	59,067	37,63	58,079	46,897

Tabla 86
Exactitud vs (solidity 0.5, umbral y nivel) para ANN con red Patternnet

Con solidity 0.5										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	96,869	96,587	96,16	96,623	95,487	96,586	96,39	95,06	92,621
	2	96,427	96,691	96,473	95,838	97,689	96,251	93,868	92,639	92,036
	3	96,477	96,007	96,372	95,143	96,269	97,578	94,99	93,685	92,63
	4	96,262	97,182	95,033	95,413	95,612	95,525	95,595	92,462	92,355
	5	96,04	96,594	96,88	95,713	96,973	97,625	97,829	95,027	90,707
	6	94,646	94,133	93,912	93,514	92,173	94,761	91,051	93,614	90,026
	7	96,363	96,285	96,836	95,807	94,797	95,336	93,288	91,365	92,561
	8	91,879	90,468	92,007	91,185	92,387	91,507	92,926	90,394	88,462
	9	92,229	93,033	93,496	93,246	95,694	94,958	95,193	92,512	92,715
	10	91,081	92,702	92,202	92,599	93,859	95,195	94,219	94,46	95,081

Tabla 87
Eficiencia vs (solidity 0.5, umbral y nivel) para ANN con red Patternnet

Con solidity 0.5										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	8,6506	6,3056	6,3227	7,4709	9,4194	8,7771	6,1565	6,136	5,994
	2	15,154	9,1225	9,0867	10,651	13,057	12,488	8,8071	8,7846	8,6144
	3	21,62	11,652	15,475	13,127	15,635	16,791	11,268	11,468	11,121
	4	25,889	14,232	22,23	15,504	18,094	21,126	13,775	13,987	13,769
	5	28,502	16,708	27,099	17,94	20,503	23,632	16,399	16,612	17,773
	6	31,095	19,177	29,544	20,442	22,98	26,145	18,984	19,16	22,051
	7	33,733	21,689	35,962	22,807	25,378	33,004	21,401	21,678	27,229
	8	36,304	24,211	40,698	25,149	27,813	38,742	23,81	24,209	30,356
	9	43,507	26,713	43,164	27,549	30,306	41,407	26,33	26,753	32,757
	10	46,811	29,326	45,693	30,044	32,713	43,945	28,922	29,324	35,331

Tabla 88
Exactitud vs (solidity 0.6, umbral y nivel) para ANN con red Patternnet

Con solidity 0.6										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	96,586	95,782	96,018	96,754	95,832	95,504	94,631	94,874	93,973
	2	97,989	96,507	96,525	95,768	97,877	97,196	96,208	94,8	90,873
	3	96,45	96,62	96,847	97,294	97,573	96,991	97,426	96,634	94,065
	4	95,243	92,432	94,758	93,928	94,446	95,329	93,84	91,016	90,265
	5	94,894	94,039	96,663	96,086	95,038	93,488	95,494	90,592	91,655
	6	89,079	90,745	91,059	91,537	94,103	94,916	91,604	90,47	92,802
	7	92,99	92,32	93,286	93,242	92,684	96,66	92,281	93,336	92,729
	8	86,072	89,194	90,244	90,588	89,452	90,033	89,253	84,874	88,392
	9	86,642	87,171	87,074	89,648	90,84	91,052	92,165	88,571	88,042
	10	93,828	93,833	91,99	91,02	93,529	94,398	95,187	94,623	93,618

Tabla 89
Eficiencia vs (solidity 0.6, umbral y nivel) para ANN con red Patternnet

Con solidity 0.6										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	9,4672	5,9115	7,7341	5,9034	6,035	7,0712	9,2433	6,704	5,948
	2	12,999	8,5881	11,085	8,5124	8,8047	10,045	18,048	9,386	8,616
	3	15,431	11,051	13,522	10,94	11,307	12,491	23,042	11,94	12,88
	4	17,806	13,417	16	17,333	13,778	14,855	25,681	14,37	17,42
	5	20,201	15,76	18,407	20,095	16,155	17,318	28,784	19,78	22,15
	6	22,678	18,133	20,803	22,546	18,598	19,836	38,617	22,41	24,61
	7	25,094	20,48	23,198	24,954	20,973	22,319	49,045	24,94	27,03
	8	27,558	22,844	25,565	27,4	23,397	24,784	54,053	27,25	29,39
	9	29,988	25,247	27,925	29,755	25,785	27,235	56,499	29,62	31,72
	10	32,438	27,727	30,383	32,141	28,356	30,856	58,972	32,05	34,18

Tabla 90
Exactitud vs (solidity 0.7, umbral y nivel) para ANN con red Patternnet

Con solidity 0.7										
Exac. (%)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	95,493	96,585	97,208	96,684	98,636	96,271	95,101	95,239	90,493
	2	95,131	98,008	97,055	93,721	96,599	92,544	97,578	93,539	90,926
	3	92,394	93,389	96,604	95,002	95,992	97,88	93,899	94,244	94,201
	4	85,444	87,145	88,415	90,082	90,152	89,525	87,598	87,743	88,052
	5	86,048	87,279	91,226	90,321	92,062	90,023	88,579	88,322	90,275
	6	89,167	89,858	89,2	91,389	92,348	90,463	93,291	91,333	95,028
	7	90,173	91,515	91,92	88,944	91,578	94,946	94,02	94,769	97,152
	8	88,27	88,21	87,498	88,949	86,683	86,211	88,389	87,459	84,554
	9	86,127	86,823	88,207	88,164	88,338	83,932	87,547	87,281	87,796
	10	92,14	90,653	90,929	91,25	94,566	93,445	95,716	94,269	93,94

Tabla 91
Eficiencia vs (solidity 0.7, umbral y nivel) para ANN con red Patternnet

Con solidity 0.7										
Efi. (s)	Coeficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	9,0893	5,5987	5,966	6,114	6,1214	7,7032	5,5711	5,9806	6,0383
	2	11,99	13,386	8,574	13,23	8,823	15,524	9,0886	8,5923	8,5315
	3	14,455	16,975	11	16,52	11,29	19,405	13,941	11,063	11,006
	4	16,832	19,322	13,39	18,93	13,681	21,798	16,225	13,544	13,355
	5	19,274	21,793	15,83	21,33	16,137	24,343	18,631	16,004	15,737
	6	21,658	24,146	18,26	23,84	18,636	29,747	21,106	18,413	21,282
	7	24,056	26,577	23,76	26,29	21,003	32,22	23,759	21,117	24,092
	8	26,51	28,982	26,27	28,67	23,48	34,613	26,161	23,567	26,453
	9	28,896	31,348	28,66	31,1	25,916	37,04	28,526	25,932	28,932
	10	31,383	33,813	31,09	33,62	28,5	39,525	30,984	28,417	31,506

Tabla 92
Exactitud vs (solidity 0.8, umbral y nivel) para ANN con red Patternnet

Con solidity 0.8										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	96,168	95,222	95,993	97,135	98,261	94,265	96,196	95,741	90,394
	2	96,296	96,892	96,643	96,933	95,786	96,056	95,948	94,343	90,076
	3	94,387	94,545	94,706	96,9	96,736	97,818	96,746	95,508	93,335
	4	87,382	87,079	88,298	90,029	89,144	85,299	86,081	87,308	86,199
	5	89,191	88,265	84,804	85,391	86,837	85,722	87,624	86,403	87,563
	6	90,658	91,228	91,096	91,47	93,118	93,863	93,573	93,437	96,335
	7	91,551	90,632	92,193	90,161	94,066	90,832	94,748	93,401	91,583
	8	86,823	88,033	89,067	86,358	88,678	85,412	85,591	86,771	86,035
	9	88,695	86,395	84,99	86,714	87,166	86,968	86,105	86,167	84,638
	10	90,588	92,725	91,833	92,676	90,586	95,399	96,469	96,227	93,16

Tabla 93
Eficiencia vs (solidity 0.8, umbral y nivel) para ANN con red Patternnet

Con solidity 0.8										
Efi. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	20,952	8,7996	7,7082	8,5308	6,5576	6,6897	6,3175	6,4676	6,0813
	2	30,833	12,305	12,459	12,315	12,795	9,5223	8,9385	9,6087	8,648
	3	41,795	14,951	14,98	14,919	15,569	16,705	11,491	17,189	15,022
	4	51,484	17,455	19,829	22,619	17,981	20,294	13,991	19,971	17,774
	5	61,653	19,904	23,728	28,18	20,464	22,625	16,499	22,502	20,259
	6	71,994	22,454	26,228	30,866	23,053	27,719	19,08	25,093	22,909
	7	82,336	24,965	28,853	33,36	25,613	30,98	21,726	27,751	25,334
	8	86,138	27,461	31,3	35,869	28,01	33,433	24,16	30,107	29,827
	9	88,7	30,028	33,82	41,237	30,382	39,248	26,62	32,562	35,597
	10	91,239	34,992	36,374	48,021	32,986	42,199	29,231	35,178	38,424

Tabla 94
Exactitud vs (solidity 0.9, umbral y nivel) para ANN con red Patternnet

Con solidity 0.9										
Exac. (%)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	97,033	96,859	95,898	95,627	95,323	95,57	96,678	92,26	90,625
	2	97,003	97,429	96,209	96,141	95,325	97,597	95,962	92,809	92,803
	3	87,612	94,52	95,517	97,158	97,421	96,783	96,632	95,385	96,823
	4	87,716	88,923	84,585	87,731	88,145	87,241	88,802	86,086	86,832
	5	87,22	88,145	87,796	86,621	87,231	88,526	86,542	87,698	88,211
	6	90,247	91,427	91,987	91,493	95,944	95,792	96,693	93,358	95,079
	7	94,208	89,407	88,693	90,902	86,304	94,659	89,848	94,288	94,366
	8	87,085	88,971	86,386	88,62	87,444	88,464	88,374	87,435	86,332
	9	86,943	87,823	86,946	88,659	84,847	87,513	87,803	86,503	88,14
	10	92,088	86,246	95,346	92,135	93,391	95,034	95,763	97,296	95,4

Tabla 95
Eficiencia vs (solidity 0.9, umbral y nivel) para ANN con red Patternnet

Con solidity 0.9										
Efi. (s)	Coefficiente para umbralización									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Niveles	1	6,7982	6,0898	7,2464	6,197	8,6295	5,3933	6,7066	6,6894	5,792
	2	9,4386	8,9449	12,595	8,8287	12,206	8,009	9,58	9,4937	8,4525
	3	12,037	11,469	15,283	11,392	14,966	10,55	12,151	12,045	15,454
	4	14,498	13,93	20,566	13,844	17,548	12,975	14,583	14,507	18,967
	5	16,961	16,288	23,678	16,235	19,985	15,421	16,98	19,531	21,291
	6	19,601	18,772	26,221	18,771	22,507	22,516	19,522	22,841	23,788
	7	22,186	21,325	28,831	21,311	28,815	25,896	22,011	27,499	29,179
	8	24,649	23,859	31,333	23,742	31,709	30,866	24,522	31,005	37,029
	9	27,136	26,347	33,829	26,257	34,158	33,89	26,953	33,429	44,661
	10	29,628	28,826	36,382	32,068	36,678	36,347	29,569	36,003	48,563

Con los valores simulados se encontró que la máxima exactitud es de 98.636% y con una eficiencia 6.1214 segundos, con los siguientes parámetros:

- *Umbral* = 0.5
- *Solidy* = 0.7
- *Nivel* = 1

Como se mencionó anteriormente los parámetros de la etapa de segmentación son muy importantes ya que encuentran las mejores condiciones para poder definir la región más propensa en la existencia de cáncer. Para poder definir la técnica de aprendizaje de máquina que se va implementar se recomienda simular todas las opciones que se originan en los parámetros para encontrar la mejor eficiencia y exactitud del sistema, en la Tabla 49 podremos ver los parámetros obtenidos para analizarlos en las siguientes secciones

Tabla 96
Eficiencia en algoritmos de aprendizaje de maquina

Machine Learning		Parámetros	Exactitud (%)	Eficiencia (s)
Red Neuronal Artificial	Fitnet	Umbral: 0.4 Solidy: 0.6 Nivel: 1	99.8	5.35
	Patternnet	Umbral: 0.5 Solidy: 0.7 Nivel: 1	98.636	6.1214
Vecino más Cercano	Fitcknn	Umbral: 0.3 Solidy: 0.1 Nivel: 1	99.8	0.3619
Árbol de decisiones	Fitctree	Umbral: 0.2 Solidy: 0.1 Nivel: 1	99.8	0.6464
Naive Bayes	fitNaiveBayes	Umbral: 0.3 Solidy: 0.3 Nivel: 2	98.315	0.2016

Con los resultados obtenidos anteriormente es necesario corroborar los mismos para ello se implementaran herramientas de validación, como primer mecanismo se utilizara la matriz de confusión

A continuación se mostrará las matrices de confusión, medidas estadísticas y curvas ROC para cada uno de los sistemas. Primero se mostrara la matriz de confusión, la cual será la sumatoria de cada elemento de la matriz para todos los posibles casos que se generan en las 10 iteraciones de k-fold preestablecidas, en la Tabla 97. Se podrá ver las respectivas matrices para cada técnica propuesta

Tabla 97
Partes de una matriz de confusión

Técnica	Matriz de Confusión	
ANN (Fitnet)	94	2
	2	138
ANN (Patternnet)	139	3
	2	92
Vecino más cercano	93	1
	2	140
Naive Bayes	93	1
	2	140
Árbol de decisiones	92	1
	3	140

Con las matrices de confusión aplicamos las formulas respectivas para las proporciones estadísticas, en la Tabla 98 podemos ver estos resultados.

Tabla 98
Valores estadísticos aplicados en varios sistemas

Técnica	Sensibilidad	Precisión	Tasa de falso descubrimiento
ANN (Fitnet)	0.985	0.979	0.021
ANN (Patternnet)	0.981	0.986	0.014
Vecino más cercano	0.989	0.979	0.021
Naive Bayes	0.989	0.979	0.021
Árbol de decisiones	0.989	0.968	0.021

Como se observa los valores en los diferentes sistemas generan una gran confiabilidad en los resultados debido a que los sistemas propuestos tienen una tasa de confiabilidad, permitiendo una identificación correcta y precisa en las regiones donde se cree existe cáncer.

Con los valores establecidos anteriormente se realizaron las curvas ROC para cada uno de las técnicas establecidas en el sistema en las Figuras 41-45

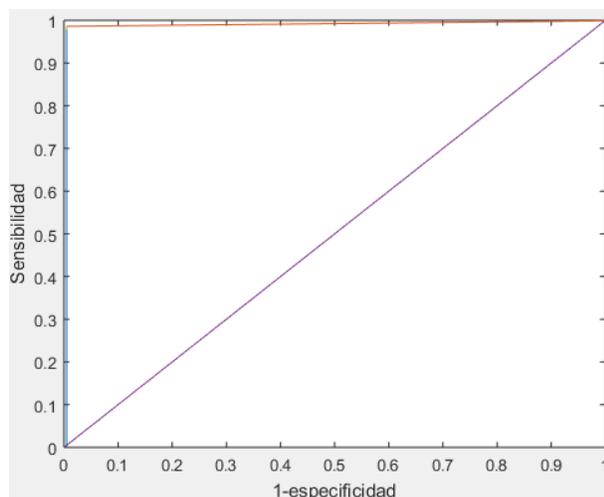


Figura 41 Curva ROC para ANN (Fitnet)

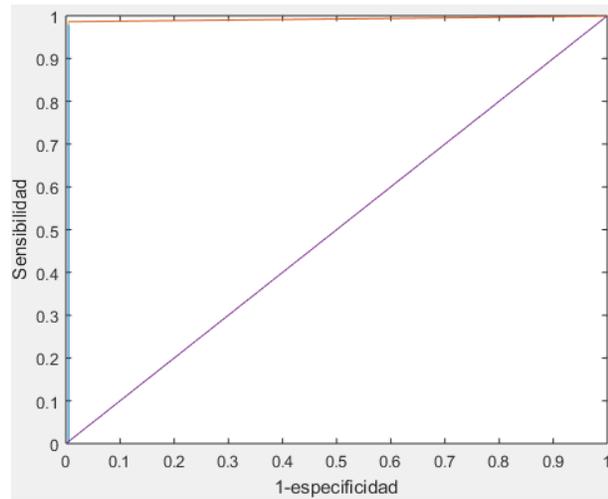


Figura 42 Curva ROC para ANN (Patternnet)

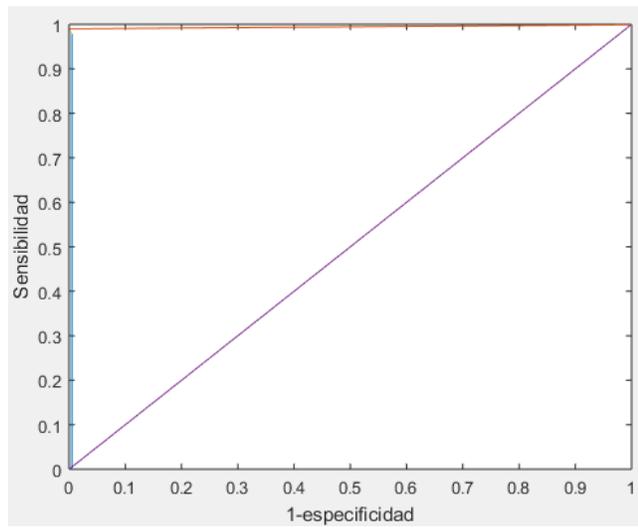


Figura 43 Curva ROC para vecino más cercano

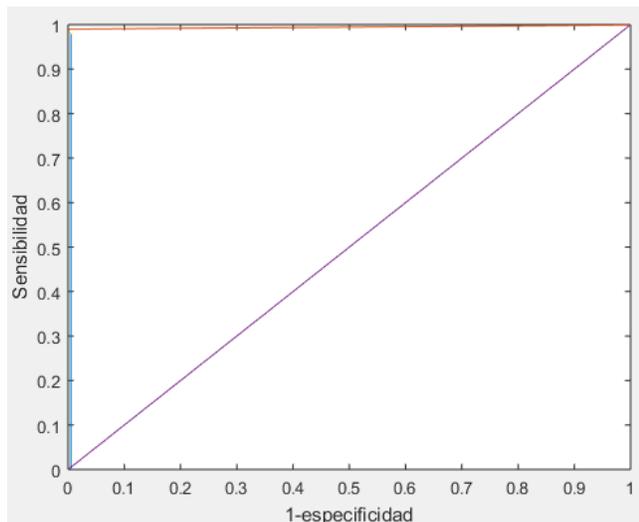


Figura 44 Curva ROC para árbol de decisiones

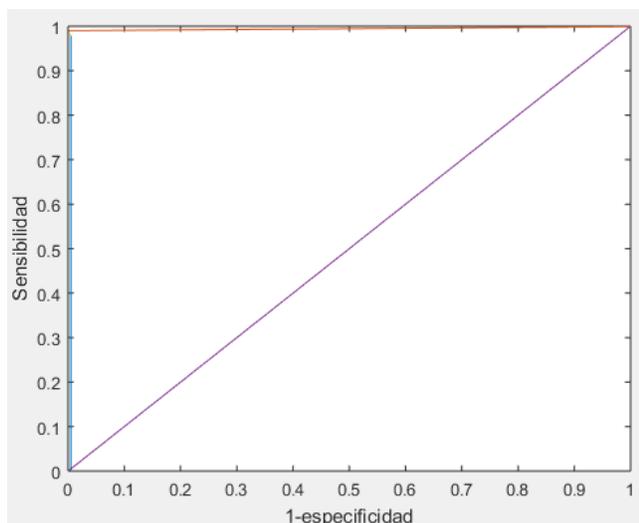


Figura 45 Curva ROC para Naive Bayes

Al aplicar las curvas ROC en los diferentes algoritmos se encuentra un alto rendimiento a todos los sistemas propuestos, debido a que las curvas se encuentran en la zona superior izquierda, la cual proporciona un alto índice de sensibilidad y especificidad, al momento de generar una identificación de células cancerígenas.

4.3. Discusión

Los resultados obtenidos en la matriz de confusión específicamente en la tasa de falso negativo que posee un valor de 2 que puede ser considerado como un valor muy pequeño, pero este a su vez es muy importante ya que indica la cantidad de casos de prueba que el sistema dicta de la no presencia de cáncer en la mamografía aunque esta si la posea, generando así un gran problema al momento de tomar como absoluto los resultados emitidos por el sistema ya que interfieran en un tratamiento efectivo.

Al analizar más profundamente estos casos pudimos encontrar que el error se presenta, se debe a que el sistema no encuentra adecuadamente en la etapa de erosión un valor correcto de umbralización en la eliminación de zonas de no interés. Por ende se realizó un ajuste en el contraste de la imagen al momento de ingresar a las etapas de segmentación, para que pueda identificar las posibles células cancerígenas. Para ello se realizó la ejecución de 0.1 a 0.9 en el comando `imadjust` con respecto a los resultados emitidos por los médicos en caso de 0 cuando no lo es y 1 cuando exista dicha enfermedad, esto se observa en la figura 46

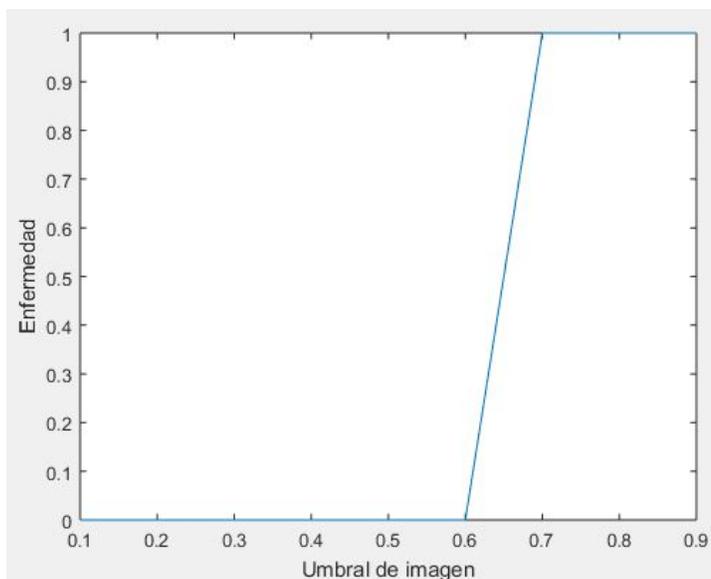


Figura 46 Variación de umbral para datos que poseen error en la detección

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

Con la investigación propuesta se pudo abordar de manera práctica un tema muy sensible como es el cáncer de mama brindando un sistema sencillo, práctico y preciso en la identificación de melanomas en la imagen. Esta herramienta permitirá un mayor grado de seguridad en los resultados obtenidos ya que será un soporte en los criterios que brindará el galeno al emitir sus resultados,

El conocimiento previo en el análisis de las diferentes técnicas de segmentación permitió encontrar de mejor manera la eliminación de regiones de no interés ya sea por medio de la umbralización de regiones con células cancerígenas o por la forma geométrica que dichas células poseen.

Como mencionamos anteriormente los parámetros que se usan en la etapa de segmentación influyen gravitadamente en la eficiencia del sistema, por ende se debe simular todas las condiciones probables que puedan existir hasta llegar la máxima eficiencia que permitirá seleccionar los mejores valores para cada parámetro en todas las técnicas de aprendizajes propuestas.

Las sensibilidades que generan los sistemas tienen un valor cercano al 99%, haciendo que las mamografías con resultados positivos de cáncer sean identificados correctamente con respecto a todos los casos que se presentan en la base de datos, esta seguridad en los resultados genera una gran reducción en el riesgo de emitir un diagnóstico errado, ya que si se emitiera a una paciente con cáncer un resultado errado conllevaría hasta la muerte de la misma sin el tratamiento oportuno.

Las curvas ROC de los algoritmos planteados brindan una gran confiabilidad debido a que sus curvas están muy cerca de la región superior izquierda justificando los valores de eficiencia de cada uno de ellos. Como conocemos la curva relaciona costo/beneficio, haciendo que los resultados obtenidos generen decisiones abruptas por esta relación, en nuestro caso el sistema posee un

equilibrio al momento de optar por un resultado indeciso sin perder las ventajas del sistema por un caso difícil de identificación.

Con los resultados obtenidos en la Tabla 50 pudimos ver que la mejor técnica de aprendizaje es vecino más cercano, con una exactitud de 99.8% y con una eficiencia del 0.3619 segundos, esto se debe a que este algoritmo internamente permite reconocer patrones a partir de las características obtenidas en las mamografías, gracias a ello el sistema permite clasificar más fácilmente los datos a cada clase a la que pertenece. Además este sistema brinda una evaluación previa en la red generada haciendo que el sistema tenga una mayor fiabilidad en la red generada.

Las eficiencias de las otras técnicas propuestas no se diferencian mucho con respecto a vecino más cercano, haciendo que el sistema propuesto en la Figura 19 justifique la utilización de cada etapa propuesta en la generación de características de cada imagen mamográfica sin importar la técnica de aprendizaje de maquina implementado.

BIBLIOGRAFÍA

- Allende, H. (2015). *El clasificador de mínimo error (Bayes)*. Obtenido de El clasificador de mínimo error (Bayes): <http://slideplayer.es/slide/3779051/>
- American Cancer Society. (2016). *Factores del cancer de Seno*. Obtenido de <http://www.cancer.org/espanol/cancer/cancerdeseno/guiadetallada/cancer-de-seno-causas-factores-de-riesgo>
- Ander-Egg, E. (1993). *Tecnica de Investigacion Social*. Buenos Aires: Magisterio del Rio de la Plata.
- Andrew, G. (2010). *The Lab Book Pages*. Obtenido de <http://www.labbookpages.co.uk/software/imgProc/otsuThreshold.html>
- Aparici, R., Garcia, A., Fernandez, J., & Osuna, S. (2009). *La imagen Analisis y representacion de la realidad*. Barcelona: Gedisa.
- Argote, L., Toledo, G., Delgado, R., Dominguez, D., Cano, P., & Noa, A. (2010). Factores de riesgo del cáncer de mama en pacientes diagnosticadas en el hospital Julio Trigo. *Revista Cubana de Salud y Trabajo*, 3-6.
- Arquero, B., Berzosa, A., Garcia, N., & Monje, M. (2009). *Investigacion experimental Metodos de la investigacion educativa*. Madrid: Universidad Autonoma de Madrid.
- Barón López, F. (2016). *bioestadistica*. Obtenido de bioestadistica: <https://www.bioestadistica.uma.es/baron/apuntes/>
- Ben-David, S. (2014). *Understanding Machine Learning : From Theory to Algorithms*. doi:9781107057135
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Singapur: Springer.
- Bonilla, A., & Ojeda, M. (2006). *Implemenatcion de mineria de datos basados en redes bayesianas para la torre de decisiones en los registros academicos*. Guayaquil: Escuela Politecnica del Litoral.
- Chaves Hernandez, D. (2 de Noviembre de 2011). *Que es una imagen Mapa de bits y una imagen de vector*. Obtenido de Que es una imagen Mapa de bits y una imagen de vector: https://www.youtube.com/watch?v=aM6AVNTCn_w
- Columbia Electrical Engineering. (2015). *Nearest Neighbor Classifiers*. Obtenido de Nearest Neighbor Classifiers: <http://www.ee.columbia.edu/~vittorio/lecture8.pdf>
- Correa, J. C. (2010). *Diagnosticos de Regresion Usando la FDR (Tasa de Descubrimientos Falsos)*. Obtenido de Diagnosticos de Regresion Usando la FDR (Tasa de Descubrimientos Falsos): <http://revistas.usta.edu.co/index.php/estadistica/article/download/22/21>
- Cunningham, P., & Cord, M. (2008). *Supervised Learning*. Dublin: Springer.

- Departamento de Electronica y Automatico, S. (2011). *Procesamiento Dlgital de Imagenes*. Obtenido de Procesamiento Dlgital de Imagenes: <http://dea.unsj.edu.ar/imagenes/recursos/Capitulo1.pdf>
- Ericson, G. (19 de Agosto de 2006). *Microsoft Azure*. Obtenido de Evaluación del rendimiento de un modelo en Aprendizaje automático de Azure: <https://azure.microsoft.com/es-es/documentation/articles/machine-learning-evaluate-model-performance/>
- Felgaer, P. (2005). *OPTIMIZACIÓN DE REDES BAYESIANAS BASADO*. Buenos Aires: Universidd de Buenos Aires.
- Fisher, R., Perkins, S., & Wolfart, W. (2003). *Grayscale Images*. Obtenido de Grayscale Images: <http://homepages.inf.ed.ac.uk/rbf/HIPR2/gryimage.htm>
- Flach, P. (2012). *Machine Learning The art and Science of algorithms that make sense of data*. New York: Cambridge.
- Gamez, C. (2010). *Using the Receiver Operating Characteristic (ROC) curve to analyze a classification model*. Obtenido de Using the Receiver Operating Characteristic (ROC) curve to analyze a classification model: <http://www.math.utah.edu/~gamez/files/ROC-Curves.pdf>
- Gazziro, M. (Enero de 2013). *Quadro Segmenta_c~ao de Imagens*. Obtenido de Quadro Segmenta_c~ao de Imagens: http://wiki.icmc.usp.br/images/b/bb/Otsu_e_derivadas.pdf
- Gershenson, C. (2010). *Artificial Neural Networks for Beginners*. Obtenido de Artificial Neural Networks for Beginners: <https://datajobs.com/data-science-repo/Neural-Net-%5BCarlos-Gershenson%5D.pdf>
- Gonzales, R., & Woods, R. (2006). *Digital Image Processing Using Matlab*. Tennessee: Prentice Hall.
- ILCE, B. D. (2011). *Bibliotecadigital Ilce*. Obtenido de Bibliotecadigital Ilce: http://bibliotecadigital.ilce.edu.mx/sites/ciencia/volumen2/ciencia3/084/htm/sec_9.html
- Jepson, A., & Fleet, D. (2007). *Image Segmentation*. Obtenido de Image Segmentation: <http://www.cs.toronto.edu/~jepson/csc2503/segmentation.pdf>
- Laya D., Urra E., Rodriguez J., Linares P., Saldivia P., & Reigosa A. (2008). *Diseño de Software para la Estimación de la Relación de Volumen Tumor/Mama a Partir de Mamografía*. Carabobo: Facultad de Ciencias de la Salud.
- Martin, M. (21 de Mayo de 2002). *Técnicas Clasicas de Segmentacion de Imagen*. Obtenido de Técnicas Clasicas de Segmentacion de Imagen: <http://poseidon.tel.uva.es/~carlos/ltif10001/segmenclasica.pdf>
- MathWorks. (2016). *Funciones en MatLab*. Obtenido de <http://www.mathworks.com/help/matlab.html>

- Ministerio de Salud Publica. (2015). *El Ministerio de Salud comprometido en la lucha contra el cancer de mama*. Quito.
- Mitchell, T. M. (2011). *Machine Learning*. Obtenido de http://personal.disco.unimib.it/Vanneschi/McGrawHill_-_Machine_Learning_-_Tom_Mitchell.pdf
- Moreno Escobar, J. A. (2007). *Clasificación de mamografías usando la función de base radial de rango tipo M*. Mexico: Instituto Politecnico Nacional.
- Moujahid, A. (26 de Junio de 2016). *A Practical Introduction to Deep Learning with Caffe and Python*. Obtenido de A Practical Introduction to Deep Learning with Caffe and Python: <http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe/>
- Muñoz, J. (2009). *Segmentación de imágenes*. Obtenido de Segmentación de imágenes: http://www.lcc.uma.es/~munozp/documentos/procesamiento_de_imagenes/temas/pi_cap6.pdf
- Olivas, R. (2007). *Decision Trees*. UC Berkeley: Microsoft Corporation. Obtenido de Decision Trees.
- OPENLAB at City Tech. (2014, Diciembre). *ADVANCED RADIOLOGIC CERTIFICATIONS*. Retrieved from *ADVANCED RADIOLOGIC CERTIFICATIONS*: https://openlab.citytech.cuny.edu/advancedrtdraft/files/2014/12/Blausen_0628_Mammogram.png
- Ordoñez, C. A. (10 de Mayo de 2005). *FORMATOS DE IMAGEN DIGITAL*. Obtenido de *FORMATOS DE IMAGEN DIGITAL*: http://www.revista.unam.mx/vol.6/num5/art50/may_art50.pdf
- Organizaion World Health. (february de 2016). *Cancer Breast cancer : prevention and control*. Obtenido de Cancer Breast cancer : prevention and control: <http://www.who.int/cancer/detection/breastcancer/en/>
- Ortiz, F. (2002). *Procesamiento morfológico de imágenes en color. Aplicación a la reconstrucción geodésica*. Alicante: Universidad de Alicante .
- Para pacientes Radiologyinfo.org. (2015, Septiembre 30). *Mamografía*. Retrieved from Mamografía: <http://www.radiologyinfo.org/sp/pdf/mammo.pdf>
- Quilmes, U. N. (Octubre de 2005). *Segmentación por Umbralización*. Obtenido de Segmentación por Umbralización: <http://iaci.unq.edu.ar/materias/vision/archivos/apuntes/Segmentaci%C3%B3n%20por%20umbralizaci%C3%B3n%20-%20M%C3%A9todo%20de%20Otsu.pdf>
- Radiological Society of North America. (30 de 9 de 2015). *Mamografía*. Obtenido de Mamografía: <http://www.radiologyinfo.org/sp/info.cfm?pg=mammo>

- Refaeilzadeh, P., Tang, L., & Liu, H. (2008 de Noviembre de 2008). *Cross-Validation*. Arizona: Arizona State University. Obtenido de Cross-Validation: <http://leitang.net/papers/ency-cross-validation.pdf>
- Roche Farma, S. A. (2011). *Hablemos de el cancer de mama*. Madrid: acv ediciones.
- Roy, S. (Junio de 2010). *Busto de su propio graphcut segmentación de imágenes basado en OpenCV*. Obtenido de Busto de su propio graphcut segmentación de imágenes basado en OpenCV: <http://www.morethantechnical.com/wp-content/uploads/2010/05/GMM-GC-segmentation.png>
- Ruiz, A., Garcia, J., & Mesa, J. (2 de Febrero de 2012). *ERROR, INCERTIDUMBRE, PRECISIÓN Y EXACTITUD*. Obtenido de ERROR, INCERTIDUMBRE, PRECISIÓN Y EXACTITUD,: http://coello.ujaen.es/congresos/cicum/ponencias/Cicum2010.2.02_Ruiz_y_otros_Error_incertidumbre_precision.pdf
- Sánchez, C. T. (2011). *La glándula mamaria*. Obtenido de http://www.sitios.itesm.mx/webtools/Zs2Ps/libros/lagl_ndula.pdf
- School of management. (2003). *k-Nearest Neighbor Algorithms for Classification and Prediction*. Obtenido de k-Nearest Neighbor Algorithms for Classification and Prediction: <https://ocw.mit.edu/courses/sloan-school-of-management/15-062-data-mining-spring-2003/lecture-notes/knn3.pdf>
- Shai, B. D., & Shai, S. S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press. Obtenido de Understanding Machine Learning: From Theory to Algorithms.
- Sheets, K. (2013). *ImageJ*. Obtenido de ImageJ: <https://sites.google.com/site/learnimagej/image-processing/what-is-a-digital-image>
- SIVP Toolbox. (Abril de 2006). *rgb2gray*. Obtenido de rgb2gray: <http://sivp.sourceforge.net/func-list-0.5.0/rgb2gray.htm>
- Smola, A. (2010). *Introduction to machine learning*. Obtenido de <http://alex.smola.org/drafts/thebook.pdf>
- Smola, A., & Vishwanathan, S. (2008). *Introduction to Machine Learning*. The Edinburgh Building: Cambridge.
- Susan G, k. (2013). *Datos para la vida Mamografía*. Retrieved from Datos para la vida Mamografía: https://ww5.komen.org/uploadedFiles/Content_Binaries/806-375-SP.pdf
- Sutton, O. (2012). *Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction*. Obtenido de Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction: http://me.seekingqed.com/files/intro_KNN.pdf

- UNAM, R. (2010). *Antecedentes de los métodos de Segmentacion*. Obtenido de Antecedentes de los métodos de Segmentacion: <http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/171/A4.pdf?sequence=4>
- Universidad Politecnica de Madrid, D. (2010). *Procesamiento morfológico*. Obtenido de Procesamiento morfológico: <http://www.elai.upm.es/webantigua/spain/Asignaturas/Robotica/ApuntesVA/cap6VAProcMorf.pdf>
- Universidad del Sur de Florida. (2006). *Base de datos digital para las mamografías exploratorias: DDSM*. Recuperado el 10 de 4 de 2016, de Base de datos digital para las mamografías exploratorias: DDSM: <http://marathon.csee.usf.edu/Mammography/Database.html>
- Universidad Metodista del Sur. (2008). *Confusion Matrix*. Obtenido de Confusion Matrix: http://faculty.smu.edu/tfomby/eco5385_eco6380/lecture/Confusion%20Matrix.pdf
- Vierada, S. (2012). *DIAGNOSIS OF BREAST REGIONS THROUGH THE USE OF RIPLEY'S KFUNCTION AND SVM*. Proceedings of the IASTED International Conference.
- Williams, B. (2010). *Photo of 1877 lithograph of funerary papyrus*. Obtenido de Photo of 1877 lithograph of funerary papyrus: <http://static2.bigstockphoto.com/thumbs/8/3/7/large1500/7383651.jpg>
- Yanfeng Li, H. C. (2016). Health & Medical Informatics A Survey of Computer-aided Detection of Breast Cancer with Mammography. *Health & Medical Informatics A Survey of Computer-aided Detection of Breast Cancer with Mammography*, 7. doi:10.4172/2157-7420.1000238